

A Corpus-Based Analysis of Frequency Effects in English Pluralization

Basima Othman Mahmood Al Hussein

*Department of English/ College of Education/ University of Salahaddin
Erbil/ Kurdistan Region/ Iraq*
basima.mahmood@su.edu.krd

Rashwan Ramadan Salih

*Department of English/ College of Languages/ University of Salahaddin
Erbil/ Kurdistan Region/ Iraq*
rashwan.salih@su.edu.krd

Submission date: 1/9/2023

Acceptance date: 28 /9/2023

Publication date: 23 /1/2024

Abstract:

When it comes to how the human mind processes linguistic information, frequency effects have been considered as supporting evidence for the dual mechanism approach. This approach proposes that only irregulars are stored in the associative memory, while regulars rely on rule-based processing. Accordingly, followers of this approach predict that irregular forms may exhibit responsiveness to frequency due to their storage nature, whereas regular forms may show a lack of sensitivity to frequency. The study aims to explore the influence of frequency on regulars and irregulars in the English language, with a specific focus on the noun domain. We investigate whether irregular nouns are more frequent than regular nouns and whether frequency effects differ between the two frequency-based groups. The study depends on data from the Corpus of Contemporary American English. The top regular and irregular nouns, as well as their low-frequency counterparts, were selected for analysis. The study then compares the word frequency distributions of these nouns, both in singular and plural forms, and employs descriptive and statistical analyses to assess the significance of variation in frequency effects between regulars and irregulars. Contrary to the predictions of the dual mechanism approach, the findings indicate that regular nouns generally occur more frequently than irregular nouns in Contemporary English, both in singular and plural forms, challenging the dual mechanism's expectations. These findings have significant implications and consequences for our understanding of language processing mechanisms and indicate the requirement for more elaborate models.

Keywords: word frequency, associative memory, the dual mechanism approach, Contemporary American English

تحليل قائم على المتون لتأثير التكرار على صيغ الجمع باللغة الإبخيلية

باسمة عثمان محمود الحسين

قسم اللغة الإنجليزية/ كلية التربية/ جامعة صلاح الدين أربيل/ إقليم كردستان/ العراق

رشوان رمضان صالح

قسم اللغة الإنجليزية/ كلية اللغات/ جامعة صلاح الدين أربيل/ إقليم كردستان/ العراق

المستخلص:

عندما يتعلق الأمر بكيفية معالجة العقل البشري للمعلومات اللغوية، تعد آثار التكرار دليلاً داعماً للنهج لنظرية الآلية المزدوجة. تقترح هذه النظرية أن الكلمات الشاذة تُخزن فقط في الذاكرة الرباطية، بينما يعتمد الكلمات العادية على المعالجة المبنية على القواعد. وبناءً على ذلك، يتوقع اتباع هذه النظرية أن تظهر الأشكال الشاذة استجابةً للتكرار بسبب طبيعة تخزينها، في حين قد تظهر الأشكال العادية عدم حساسية للتكرار. تهدف هذه الدراسة إلى استكشاف تأثير التكرار على الكلمات العادية والشاذة في اللغة الإنجليزية، مع التركيز الخاص على مجال الأسماء. نحن نبحث فيما إذا كانت الأسماء الشاذة أكثر تكراراً من الأسماء العادية، وما إذا كانت آثار التكرار تختلف بين المجموعتين المعتمدة على التكرار. تعتمد الدراسة على بيانات من مستودع اللغة الإنجليزية الأمريكية المعاصرة. للتحليل تم اختيار أعلى الأسماء العادية والشاذة، وكذلك نظرائها ذات التكرار المنخفض. تقوم الدراسة بمقارنة توزيعات تكرار الكلمات هذه، سواء في الأشكال المفردة أو الجمع، وتستخدم تحليلات وصفية وإحصائية لتقييم مدى أهمية تباين آثار التكرار بين العاديات والشواذ. على عكس توقعات نظرية الآلية المزدوجة، تشير النتائج إلى أن الأسماء العادية عموماً تحدث بتكرار أكبر من الأسماء الشاذة في الإنجليزية الأمريكية المعاصرة، سواء في الأشكال المفردة أو الجمع، مما يشكل تحدياً لتوقعات نظرية الآلية المزدوجة. تلك النتائج لها آثار وعواقب مهمة على فهمنا لآليات معالجة اللغة وتشير إلى ضرورة وجود نماذج أكثر تفصيلاً.

الكلمات الدالة: تكرار الكلمات، الذاكرة الرباطية، نظرية الآلية المزدوجة، الإنجليزية الأمريكية المعاصرة

1. Introduction

With a focus on understanding the inner mechanisms of language processing, generative grammar proposes that the human language faculty consists of a finite list of lexical items and a computational component. This component combines these lexical items using combinatorial rules to generate an unlimited range of complicated phrases and sentences [1],[2],[3],[4],[5],[6],[7]. For instance, take the English past tense, a regular verb is generated by a rule that adds the suffix "-ed" to a verb stem, as in "play-played." Within this rule-based framework, all linguistic expressions emerge via the application of rules. As a result, these expressions are not anticipated to be influenced by their frequency.

After an extended period dominated by generative grammar, which serves as a prototype for rule-based models of language processing, connectionism has emerged to offer a different understanding of this processing. These connectionist approaches propose certain assumptions, asserting that all linguistic knowledge is acquired and stored in an associative memory. In these approaches, the frequency of occurrence plays a

crucial role in establishing connections between words [8],[9],[10]. This distinct viewpoint has ignited ongoing debates concerning the acquisition of morphology in particular, and more broadly, how language is mentally represented [6],[8],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19],[20],[21],[22]. Followers of the dual mechanism approach integrate the fundamental characteristics of both generative grammar and connectionism. They utilize rules while also integrating the associative component for language processing. In accordance with this approach, regular patterns are created using rules, whereas irregular ones are stored within the associative memory [15],[19],[23],[24],[25],[26]. Consequently, the followers of the dual mechanism approach predict that there is a correlation between word frequency and (ir)regularity. According to this perspective, irregulars are expected to exhibit frequency effects as they are fully stored in the associative memory and hence, they should be frequency-sensitive. In contrast, regular words are supposed to be generated by a rule that is insensitive to frequency; thus, they are expected to be rule-derived. Therefore, words that are more frequent are more likely to be irregular in this dual mechanism framework.

[26:p223] is one of the followers of the dual mechanism approach who emphasizes on regulars and irregulars as a means to display that those words and rules are the "ingredients" of language. He affirms that are processed by different mechanisms: regulars are formed by rules, while irregular verb forms are stored in the mental lexicon. Likewise, [28:p456] emphasize in a similar manner that:

The regular-irregular distinction is an epiphenomenon of the design of the human language faculty, in particular, the distinction between lexicon and grammar made in most traditional theories of language.

They establish the lexicon as a segment of memory encompassing all arbitrary combinations of sounds and meanings that represent the morphemes and simple words within a language. Meanwhile, grammar constitutes a generative mechanism responsible for organizing complex words, phrases, and sentences from these morphemes and basic words. They also propose that irregular forms, much like other words, are acquired and stored, but with a grammatical feature like [PAST] merged into their lexical entries. In contrast, regular forms have a productive nature, allowing them to be produced using a rule, much like phrases and sentences. The presence of a memorized inflected form of an irregular form may block the application of the rule to that specific form. In cases where no stored irregular inflected form exists for a particular stem, the default affix of a regular form is attached to that stem. This default affix is applicable to any unspecialized item, as illustrated in the diagram below.

regulars. Several empirical findings lend support to these predictions. The top ten frequent verbs in English, namely *be, have, do, go, say, can, will, see, take, and get* - are all irregular [25],[27]. Likewise, [28] exhibit that the most frequently used English verbs are irregular, while less frequently used verbs adhere to regular patterns. They assert that among the top 30 most frequent English verbs in the past tense, 22 are irregular. Further empirical evidence arises from diachronic changes in language: irregular verbs that become less frequent over time, for various reasons, often undergo regular inflections. For instance, verbs like "*chide*," "*glide*," and "*gripe*" have transitioned to regular forms throughout history [25:p69].

In this present research, we will examine the predictions of the dual mechanism within Contemporary American English. The main purpose is to explore the possible connection between (ir)regularity and word frequency within the noun system. With this goal in mind, we plan to conduct a corpus analysis utilizing internet-sourced data. The objective of this investigation is to assess the extent to which contemporary usage of regular and irregular nouns can offer support for the dual mechanism approach. Specifically, the study will address the following research questions:

- Do irregular nouns currently tend to appear more frequently than regular ones?
- Do irregular nouns exhibit higher frequencies than regular ones in both their singular and plural forms within Contemporary American English?

The next section of this research will provide a detailed overview of the methodology deployed to carry out this investigation. By using robust research methods, our intention is to provide valuable perspectives into the dual mechanism theory's applicability in explaining noun processing within contemporary American English. This research may enhance our understanding of how language evolves and is processed in the era of digital communication.

2. Method

The aim of this research is to detect and compare the influence of frequency on regulars and irregulars, specifically on the noun domain. Frequency effects are taken as a piece of evidence for supporting the notion of storage. If frequency effects are noticed only for irregulars and not for regulars, this may propose that irregulars and regulars are processed by different mechanisms, with irregulars depending on storage mechanisms and regulars depending on rules. Yet, if frequency effects are observed (or not) for both irregulars and regulars, it may suggest that both types of verbs are processed using similar mechanisms, either storage or rules. To explore these expectations, we will compare nouns with the highest word frequency and those with the lowest word frequency. The data of this study was sourced from the Corpus of Contemporary American English (COCA), containing billion words (www.wordfrequency.info). COCA is unique in its modernity, and genre diversity, containing a wide range of different genres like fiction, popular magazines, newspapers, academic texts, TV and Movies subtitles. This study was accomplished by adhering to the four sketched steps:

1. By using the word frequency data available on the COCA corpus website, we extracted a subset of singular nouns from a wordlist containing the top 5,000 most commonly used words. The aim was to cover all the top possible nouns used in the corpus. This step

involved using the part-of-speech tag [POS] in COCA. As a result, a compilation of 2,500 nouns was obtained. Table 1 below displayed the initial 30 words from the 5,000 list while Table 2 exhibited the top 30 nouns with their word frequencies from the 2,500 list (See Appendix 1 for the complete list).

Table 1: The top 25 words with their word frequencies from a wordlist containing the top 5,000 most commonly used words in COCA

rank	words	PoS	rank	words	PoS	rank	words	PoS	rank	words	PoS
1	the	a	26	say	v	51	just	r	76	our	a
2	be	v	27	at	i	52	up	r	77	into	i
3	and	c	28	what	d	53	think	v	78	here	r
4	a	a	29	his	a	54	time	n	79	then	r
5	of	i	30	from	i	55	there	e	80	than	c
6	to	t	31	go	v	56	see	v	81	look	v
7	in	i	32	or	c	57	her	a	82	way	n
8	i	p	33	by	i	58	as	i	83	more	r
9	you	p	34	get	v	59	out	r	84	these	d
10	it	p	35	she	p	60	one	m	85	no	a
11	have	v	36	my	a	61	come	v	86	thing	n
12	to	i	37	can	v	62	people	n	87	well	r
13	that	c	38	as	c	63	take	v	88	because	c
14	for	i	39	know	v	64	year	n	89	also	r
15	do	v	40	if	c	65	him	p	90	two	m
16	he	p	41	me	p	66	them	p	91	use	v
17	with	i	42	your	a	67	some	d	92	tell	v
18	on	i	43	all	d	68	want	v	93	good	j
19	this	d	44	who	p	69	how	r	94	first	m
20	n't	x	45	about	i	70	when	c	95	man	n
21	we	p	46	their	a	71	which	d	96	day	n
22	that	d	47	will	v	72	now	r	97	find	v
23	not	x	48	so	r	73	like	i	98	give	v
24	but	c	49	would	v	74	other	j	99	more	d
25	they	p	50	make	v	75	could	v	100	new	j

Table 2: The top 30 nouns in the singular form with their word frequencies from a wordlist containing the top 2,500 most commonly used nouns in COCA

N.	noun	word freq.	noun	word freq.	noun	word freq.	noun	word freq.	noun	word freq.
1	time	2018725	day	1068902	state	638012	place	508803	group	470971
2	people	1800205	life	852257	family	544520	problem	504175	guy	464082
3	year	1729962	woman	759817	President	538319	country	499369	case	458383
4	way	1260011	child	732511	house	537037	week	484834	question	457301
5	thing	1202004	world	685426	student	530196	point	484094	work	456169
6	man	1091176	school	660191	part	517693	hand	481332	night	452094

- Afterward, we chose the top ten regulars and the top ten irregulars versus the bottom ten regulars and the bottom ten irregulars from the list of 2,500 nouns (See Table 3 for the selected regulars and irregulars based on frequency factor and for further information, see Appendix 1). This selection was made based on the comparison between the top nouns with the highest word frequencies and the nouns with the lowest word frequencies in our sample. To deepen the analysis of the noun system,

for each of the chosen nouns, word frequencies in the plural forms were gathered and compared from the COCA corpus.

Table 3: The selected regulars and irregulars based on a frequency factor from a wordlist containing the top 2,500 most commonly used nouns in COCA

N.	freq. level	regular nouns	word freq.	irregular nouns	word freq.
1-	high	time	2018725	man	1091176
2-		people	1800205	life	852257
3-		year	1729962	woman	759817
4-		way	1260011	child	732511
5-		thing	1202004	foot	209334
6-		day	1068902	wife	197306
7-		world	685426	analysis	128029
8-		school	660191	crisis	73691
9-		state	638012	half	55207
10-		family	544520	tooth	46405
11-	low	reform	70010	self	46225
12-		gold	69903	leaf	43674
13-		planet	69822	knife	35408
14-		location	69669	wolf	29042
15-		bird	69662	mouse	28329
16-		principle	69403	phenomenon	28258
17-		search	69392	shelf	21854
18-		possibility	69316	hypothesis	20698
19-		sky	69285	sheep	14430
20-		horse	68883	thief	14374
Totals			12303303		4428025

3. Next, word frequencies and average frequencies of the chosen regular and irregular nouns split by frequency (high versus low) are computed and displayed in tables and various types of graphs to facilitate comparison and description. Subsequently, statistical models will be employed to assess the significance of the variation in frequency effects between regulars versus irregulars. By comparing the obtained results with the predictions of dual mechanism models, the most appropriate model that aligns well with the data will be recognized, ultimately leading to a conclusion.

2.2. Results and Discussion

As mentioned earlier, the dual mechanism approach predicts that irregulars and word frequency are related because they require additional memory effort for successful retrieval while regulars are generated through a rule and do not need to be fully retrieved from associative memory. To investigate this forecast in Present-day English, our aim is

to explore the possible link between (ir)regularity and word frequency within the noun system by conducting a corpus study. Accordingly, this investigation raises the following inquiries:

- Do irregular nouns generally occur more frequently than regular ones in Contemporary American English?
- In singular and plural forms, are irregular nouns more frequent than regular ones in Contemporary American English?

The first phase of this research involves conducting an initial inquiry with the purpose of investigating the claims put forth by [25]. These claims propose that irregulars tend to have a high frequency and conversely, high-frequency words are more likely to be irregular. It is also suggested that high-frequent irregulars are retrieved faster compared to their low-frequency counterparts. However, no significant difference in retrieval rates between the high and low-frequency regulars. Empirical evidence supporting these predictions includes the fact that the top ten frequent verbs in English, *be, have, do, go, say, can, will, see, take, and get*, are all irregular [25]. Similarly, [28] demonstrate that the most frequent verbs of English are all irregular, whereas verbs with lower frequencies tend to be all regular. Specifically, of the top 30 most frequent English verbs in the past form, 22 of them are irregular. To explore this inquiry, let's begin by examining the top ten singular verbs with the highest frequency found in our corpus. Remarkably, all the top ten verbs in the corpus display irregular forms, namely *be, have, do, say, go, get, can, know, will, and make*. This discovery concerning the verb system aligns with the predictions proposed by [25] and [28] concerning the top ten most common verbs in English. Table 4 and Figure 2 display the word frequency distributions of these verbs.

Table 4: The top ten singular verbs with the highest frequency found in the corpus of the study

verb	irregular	Word freq.	%
1	be	32394756	44.6
2	have	10514314	14.5
3	do	8186412	11.3
4	say	4096416	5.6
5	go	3546732	4.9
6	get	3347615	4.6
7	can	3091046	4.3
8	know	2761628	3.8
9	will	2372215	3.3
10	make	2290830	3.2
Total		72601964	100

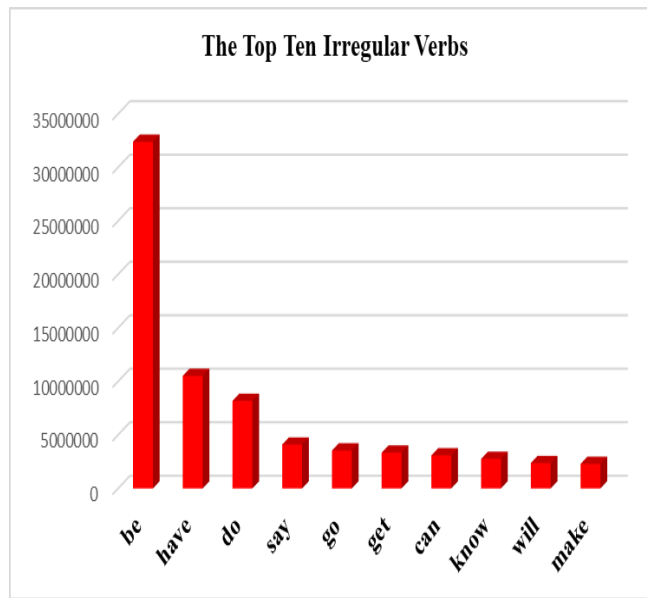


Figure 2: The top ten singular verbs with the highest frequency found in the corpus of the study

Shifting to the noun system and checking the same investigation, we notice that the top ten singular nouns with the highest frequency found in the corpus are *time*, *people*, *year*, *way*, *thing*, *man*, *day*, *life*, *woman*, and *child*. Among these top ten nouns, six are regulars, accounting for approximately 72.5% of the total word frequencies (See Table 5 and Figure 3). However, only four are irregulars, making up only about 27.3 % of the total word frequencies. Notably, the top five nouns are all regulars (See Figure 4). Clearly, these outcomes from our sample do not support the observations made by [25] and [28] who suggest that irregulars represent the most commonly used forms in the English language. From the perspective of the dual mechanism approach, irregulars reveal sensitivity to frequency to confirm successful storage and retrieval from associative memory. We may reasonably infer that this outcome can be logical, grounded on the fact that the English language naturally comprises a greater number of nouns compared to verbs. During 2017, the predictable count of nouns in the Oxford English Corpus was around 600,000, while the number of verbs was approximately 15,000 (The Oxford English Corpus. Sketch Engine: <https://www.sketchengine.eu/oxford-english-corpus/>). It is essential to remember that the English language is in a continual state of evolution, with constantly new words being added over time. Therefore, the real count of nouns and verbs in the English language could potentially be higher.

Table 5: Frequency distributions of the top ten nouns from the selected sample

	nouns	word frequency	average
1	time	2,018,725	16.1
2	people	1,800,205	14.4
3	year	1,729,962	13.8
4	way	1,260,011	10.1
5	thing	1,202,004	9.6
6	man	1,091,176	8.5
7	day	1,068,902	8.5
8	life	852,257	6.8
9	woman	759,817	6.1
10	child	732,511	5.9
Total	irregular		27.3
	regular		72.5
	both	12515570	100

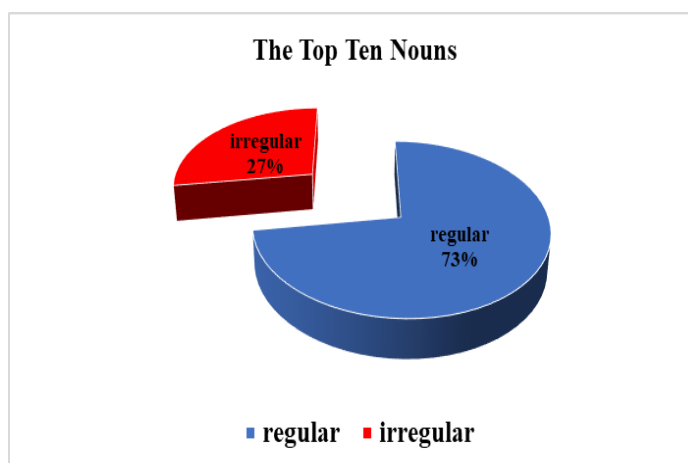


Figure 3: Frequency distributions of the top ten nouns from the selected sample

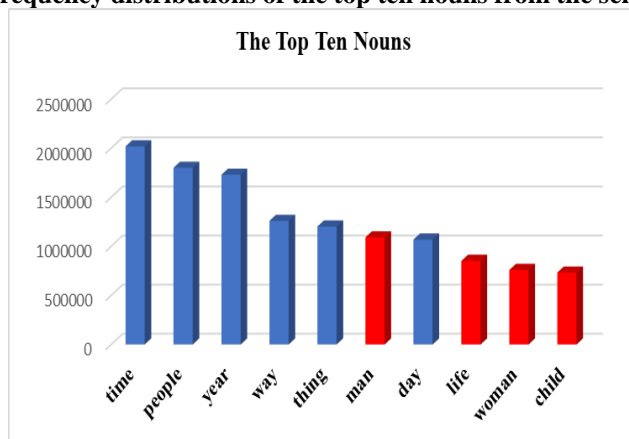


Figure 4: The top ten nouns from the selected sample

To further delve into the investigation of the noun system, my next step will involve comparing the ten nouns with the highest word frequencies to the ten nouns with the lowest word frequencies in our sample. This comparison will be directed at both the singular and plural forms of the nouns. By expanding the scope to include more data points with two different types (singular and plural), the aim is to investigate whether irregulars display significantly higher frequencies than regulars and whether irregular nouns in both singular and plural forms are currently more frequent in English compared to regular nouns. To accomplish this aim, we show an overview of word frequency distributions for the chosen nouns, categorized by their type (regular versus irregular), form (singular versus plural), and frequency (high versus low) (See Table 5 below).

Table 6: Word frequencies of the verbs by type, form, and frequency in the selected sample

	freq. level	regular	singular word freq.	plural word freq.	irregular	singular word freq.	irregular plural forms	plural word freq.
1.		time	2018725	354135	man	1091176	men	348687
2.		people	1800205	1783263	life	852257	lives	171538
3.		year	1729962	1032451	woman	759817	women	443323
4.		way	1260011	1032451	child	732511	children	442355
5.		thing	1202004	1032451	foot	209334	feet	150072
6.	high	day	1068902	1032451	wife	197306	wives	14851
7.		world	685426	17735	analysis	128029	analyses	20511
8.		school	660191	134420	crisis	73691	crises	6474
9.		state	638012	396711	half	55207	halves	3114
10.		family	544520	104385	tooth	46405	teeth	38072
11.		nest	12951	2679	self	46225	selves	3583
12.		goat	12944	4600	leaf	43674	leaves	65174
13.		equivalent	12941	1239	knife	35408	knives	6179
14.		banker	12931	7599	wolf	29042	wolves	9496
15.	low	interval	12928	6755	mouse	28329	mice	12750
16.		gasoline	12925	39	phenomenon	28258	phenomena	8342
17.		bucket	12920	3107	shelf	21854	shelves	9713
18.		theft	12907	1079	hypothesis	20698	hypotheses	5202
19.		laundry	12903	269	sheep	14430	sheep	14375
20.		constraint	12902	10272	thief	14374	thieves	6003
Totals	high		11607958	6920453	high	4145733	high	1638997
	low		129252	37638	low	282292	low	140817
	both		11737210	6958091	both	4428025	both	1779814

As summarized in Table 7, out of the total word frequency of the selected nouns (24,903,140), the word frequency of regular nouns (18,695,301) exceeds that of irregular

nouns (6,207,839). For the frequent group, regular nouns have higher word frequencies in both singular (11,737,210) and plural forms (6,958,091) compared to irregular nouns (4,428,025 for singular and 1,779,814 for plural). Yet, in the infrequent group, the situation is reversed. Irregular nouns have higher word frequencies (282,292 for singular and 140,817 for plural) than regular nouns (129,252 for singular and 37,638 for plural) in the singular and plural forms. Charts 5 and 6 visually illustrate these obtained findings. The increased word frequencies of regular nouns in the frequent domain and their decreased word frequencies in the infrequent domain contradict the dual mechanism approach's prediction that irregulars with high frequencies typically show higher frequencies than regulars.

Table 7: Word frequencies of the nouns by type, form, and frequency in the selected sample

	word frequencies		Total
	high	low	
regulars/ singular	11607958	129252	11737210
irregulars/ singular	4145733	282292	4428025
regulars/plural	6920453	37638	6958091
irregulars/plural	1638997	140817	1779814
Total	regulars		18695301
	irregulars		6207839
	both		24903140

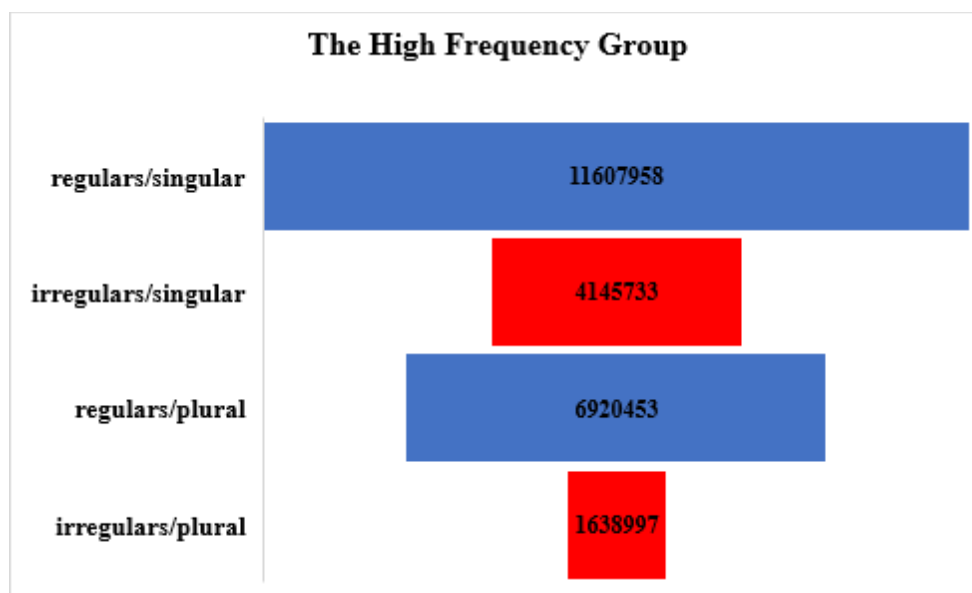


Figure 5: The high-frequency group of nouns in the selected sample

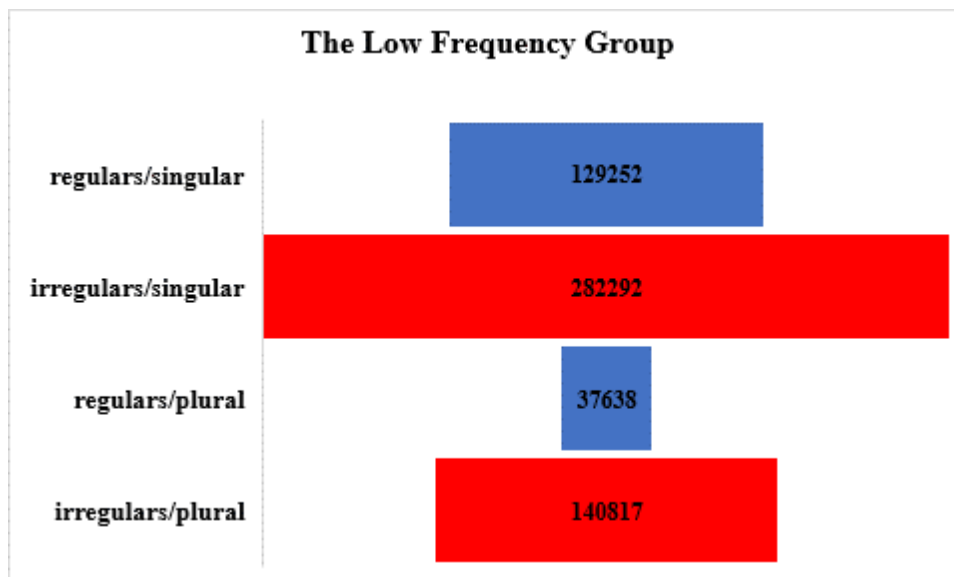


Figure 6: The low-frequency group of nouns in the selected sample

We conducted a statistical model to explore how type, form, and frequency affect the word frequency of nouns in the sample of this study. A linear model was adopted, considering word frequency as a dependent variable and treating type (regular and irregular), form (singular and plural), and frequency (high and low) as fixed factors. The outcomes of the linear model display that the main effects of type ($\beta = 0.54$, $t = 2.38$, $p = 0.07$), form ($\beta = -0.75$, $t = -2.63$, $p = 0.0004$), and frequency ($\beta = -4.23$, $t = -21.95$, $p < 2e-15$) are all statistically significant. Moreover, the effects of the interactions between type and frequency ($\beta = -2.04$, $t = -8.44$, $p < 2e-13$), form and frequency ($\beta = 1.04$, $t = 3.23$, $p = 0.002$), and type, form, and frequency ($\beta = 1.55$, $t = 2.44$, $p = 0.0005$) are also found to be significant. Nevertheless, the interaction between type and form ($\beta = 0.07$, $t = 0.25$, $p = 0.688$) is not significant. These findings provide strong evidence that the differences in frequency distributions of nouns categorized by type, form, and frequency are statistically significant. However, the differences in frequency distributions between regular and irregular nouns in their singular and plural forms are not significant. This supports the idea that regular nouns tend to occur more frequently than irregular nouns in Contemporary English. In other words, regular nouns are generally more frequent than irregular nouns in Contemporary English. The increased word frequencies of regular nouns in the high-frequency area and their decreased word frequencies in the low-frequency area compared to irregular ones challenge the prediction of the dual mechanism approach, which proposes that irregulars with high frequencies typically show higher frequencies than regulars.

Conclusion:

This corpus-based study focused on understanding how the human mind processes linguistic information, specifically in relation to the frequency effects of regulars and irregulars in Contemporary American English. The study is framed within the context of the dual mechanism approach, which suggests that irregulars are stored in

memory while regulars are generated by rules, indicating potential distinct processing mechanisms. The focus was on the noun category to explore whether irregular nouns are more prevalent than regular ones in both singular and plural forms using data from the Corpus of Contemporary American English.

The outcomes of this investigation challenge the predictions of the dual mechanism approach, which proposes that irregular words, being stored in memory, should manifest higher word frequencies compared to regulars. Contrary to this expectation, the results reveal that among the ten most frequently occurring singular nouns, a majority of six are regulars, dominating total word frequencies. This pattern extends to a comparison of the ten nouns with the highest and lowest frequencies, revealing that regular nouns show higher frequencies in common usage and lower frequencies in less common usage, contrary to the dual mechanism's assumptions.

This conclusion disrupts existing notions about the link between (ir)regularity, word frequency, and processing mechanisms, suggesting a more intricate relationship than previously assumed. The discrepancies observed between the predictions and the actual frequency patterns highlight the need for a deeper investigation into the cognitive processes involved in the storage and retrieval of irregulars and regulars. This could potentially pave the way for more refined language processing models.

CONFLICT OF INTERESTS

There are no conflicts of interest

References

- [1] Lapointe, S. *The representation of inflectional morphology within the lexicon*. In North East Linguistic Conference (Vol. 11, pp. 190–204). (1981).
- [2] Kiparsky, P. *Word-formation and the lexicon*. Mid-America Linguistics Conference. Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/20827/MALC_1982_3-29 Kiparsky.pdf?sequence=1. (1982).
- [3] Di Sciullo, A.-M., & Williams, E. *On the definition of the word* (Vol. 14). Springer. (1987).
- [4] Lieber, R. *Deconstructing morphology : word formation in syntactic theory*. Chicago: University of Chicago Press. (1992).
- [5] Chomsky. *The Minimalist Program*. Cambridge, MA: The MIT Press. (1995).
- [6] Yang, C. *Knowledge and learning in natural language*. Oxford University Press on Demand. (2002).
- [7] Kremers, J. *The noun phrase in Arabic: a minimalist approach*. Doctoral Dissertation, University of Nijmegen. (2003).
- [8] Rumelhart, D. E., & McClelland, J. L. *On learning the past tenses of English verbs*. Explorations in the Microstructure of Cognition, (1986).
- [9] Plunkett, K., & Marchman, V. *U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition*. Cognition, 38(1), 43–102. [https://doi.org/10.1016/0010-0277\(91\)90022-V](https://doi.org/10.1016/0010-0277(91)90022-V). (1991).
- [10] Bybee, J. Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5), 425–455. (1995).

- [11] Berko, J. *The child's learning of English morphology*. *Word*, 14(2-3), 150-177. (1958).
- [12] Chomsky, N., & Halle, M. *The sound pattern of English*. (1968).
- [13] Halle, M., & Mohanan, K. P. *Segmental phonology of modern English*. *Linguistic Inquiry*, 16(1), 57-116. (1985).
- [14] Pinker, S., & Mehler, J. *Connections and symbols* (Vol. 28). MIT Press. (1988).
- [15] Pinker, S., & Prince, A. *On language and connectionism: Analysis of a parallel distributed processing model of language acquisition*. *Cognition*, 28(1), 73-193. [https://doi.org/10.1016/0010-0277\(88\)90032-7](https://doi.org/10.1016/0010-0277(88)90032-7). (1988).
- [16] Smolensky, P. *On the Comprehension/Production Dilemma in Child Language*. *Linguistic Inquiry*, 27(4), 720-731. <https://doi.org/10.2307/4178959>. (1996).
- [17] Seidenberg, M. S., & Gonnerman, L. M. *Explaining derivational morphology as the convergence of codes*. *Trends in Cognitive Sciences*, 4(9), 353-361. [https://doi.org/10.1016/S1364-6613\(00\)01515-1](https://doi.org/10.1016/S1364-6613(00)01515-1). (2000).
- [18] Chomsky, N. *Minimalist inquiries: The framework*. Step by step: Essays on minimalist syntax in honor of Howard Lasnik, eds. R. Martin, D. Michaels & J. Uriagereka, 89-155. Cambridge, MA: MIT Press. (2000).
- [19] Pinker, S., & Ullman, M. T. *The past and future of the past tense*. *Trends in Cognitive Sciences*, 6(11), 456-463. (2002).
- [20] McClelland, J. L., & Patterson, K. *Rules or connections in past-tense inflections: What does the evidence rule out?* *Trends in Cognitive Sciences*, 6(11), 465-472. [https://doi.org/10.1016/S1364-6613\(02\)01993-9](https://doi.org/10.1016/S1364-6613(02)01993-9). (2002).
- [21] Peters, P. *Comparative studies in Australian and New Zealand English: grammar and beyond*. Macquarie University. (2009).
- [22] Fertig, D. *Analogy and morphological change*. Edinburgh University Press. Retrieved from http://hosting03.snu.ac.kr/~korean/old/data/morphology/Fertig_Analogy_and_Morphological_Change.pdf. (2013).
- [23] Prasada, S., & Pinker, S. *Generalisation of regular and irregular morphological patterns*. *Language and Cognitive Processes*, 8(1), 1-56. <https://doi.org/10.1080/01690969308406948>. (1993).
- [24] Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. *Overregularization in Language Acquisition*. *Monographs of the Society for Research in Child Development*, 57(4), i-178. <https://doi.org/10.2307/1166115>. (1992).
- [25] Pinker, S. *Words and rules*. The ingredients in language. London: Weidenfeld & Nicolson. (1999).
- [26] Pinker, S. *Words and rules*. *Lingua* 106 (1):219-42. (1998).
- [27] Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. *Quantifying the evolutionary dynamics of language.*, 449(7163), 713-6. Retrieved from <http://www.jbe-platform.com/content/journals/10.1075/jhl.2.2.01car>. (2007).
- [28] Kučera, H., & Francis, W. N. *Computational analysis of present-day American English*. Dartmouth Publishing Group. (1967).

Appendix 1: The top 2,500 most commonly used nouns in COCA

N.	noun	word freq.	noun	word freq.	noun	word freq.	noun	word freq.
1	time	2018725	Christmas	68542	origin	31342	lemon	19153
2	people	1800205	strength	68537	teen	31336	bankruptcy	19151
3	year	1729962	move	68522	coat	31266	invasion	19112
4	way	1260011	Monday	68488	tower	31252	replacement	19102
5	thing	1202004	mean	68411	license	31242	consensus	19057
6	man	1091176	beach	68303	twitter	31238	kiss	19047
7	day	1068902	survey	68301	curriculum	31171	french	19041
8	life	852257	writing	68200	mail	31141	hardware	19041
9	woman	759817	master	68182	estimate	31122	rail	19039
10	child	732511	scale	68139	insight	31094	cold	19037
11	world	685426	resident	68096	investigator	31082	mate	19036
12	school	660191	football	68073	virus	31075	cabin	19019
13	state	638012	failure	67885	hurricane	31072	dining	19017
14	family	544520	reporter	67831	provision	31059	liberal	19015
15	president	538319	fight	67767	strike	31051	snake	18993
16	house	537037	one	67755	cross	31025	tobacco	18992
17	student	530196	vision	67714	jacket	30995	orientation	18984
18	part	517693	function	67710	green	30968	input	18921
19	place	508803	human	67605	pay	30935	competitor	18893
20	problem	504175	will	67581	implication	30921	freshman	18878
21	country	499369	connection	67550	fiction	30883	donation	18869
22	week	484834	camp	67510	protest	30878	donor	18858
23	point	484094	stone	67456	mama	30857	toilet	18844
24	hand	481332	hundred	67455	twin	30810	ideology	18802
25	group	470971	key	67404	pant	30810	glory	18788
26	guy	464082	truck	67403	shock	30746	organ	18750
27	case	458383	afternoon	67378	criteria	30727	starting	18738
28	question	457301	secretary	67325	toy	30678	brick	18730
29	work	456169	conflict	67061	assault	30609	gut	18712
30	night	452094	learning	66879	peak	30595	reservation	18704
31	game	445149	share	66642	anger	30587	rebel	18704
32	number	444412	newspaper	66598	boot	30566	oak	18691
33	money	437583	foundation	66578	peer	30537	instructor	18675
34	lot	437545	variety	66555	math	30465	racism	18651
35	book	435387	perspective	66540	favor	30405	pension	18644
36	system	435303	document	66467	dust	30376	diabetes	18635
37	government	434971	presence	66412	aunt	30364	cluster	18617
38	city	433843	lesson	66252	prospect	30358	marijuana	18607
39	company	432469	crowd	65965	mood	30344	combat	18595
40	story	432184	apartment	65884	settlement	30250	costume	18568
41	job	430995	guest	65818	depth	30176	inspiration	18543
42	friend	423755	soul	65749	juice	30156	orange	18542
43	word	402216	protection	65690	gallery	30109	cooking	18528
44	fact	399574	intelligence	65687	tension	30082	conspiracy	18508
45	right	389906	border	65422	throat	30071	trait	18502
46	month	381833	reading	65417	draft	30062	van	18497
47	program	375708	leadership	65396	reputation	29988	garlic	18473
48	business	373744	chief	65311	index	29987	drinking	18467
49	home	370758	attitude	65288	mess	29961	crystal	18456
50	kind	367584	start	65218	joy	29949	stretch	18453
51	study	364915	website	64875	steel	29886	pro	18444
52	issue	362137	distance	64404	motor	29863	associate	18443
53	name	361916	climate	64223	enterprise	29857	helicopter	18430
54	idea	358006	relation	64199	salary	29856	counsel	18429
55	room	357682	flight	64148	cousin	29840	equation	18417
56	percent	357515	boat	64026	dozen	29771	timing	18399
57	law	354958	aspect	63967	evolution	29760	pope	18398
58	power	351483	january	63714	fishing	29682	opera	18377
59	kid	351451	volume	63626	anxiety	29659	mount	18361
60	war	350575	theater	63549	professional	29639	indication	18351
61	head	348664	speed	63502	carbon	29639	motive	18345
62	mother	348470	identity	63440	essay	29637	porch	18345

63	team	348063	demand	63402	drama	29588	gaze	18339
64	eye	347153	guard	63290	stranger	29572	lap	18314
65	side	342841	tuesday	63179	belt	29548	starter	18292
66	water	342574	facility	63117	formula	29504	colonel	18262
67	service	332313	farm	62942	potato	29496	beef	18253
68	area	331866	fun	62839	monster	29491	thanksgiving	18234
69	person	330987	august	62821	smoke	29483	intensity	18229
70	end	328699	link	62629	telephone	29482	cave	18219
71	hour	323999	shoe	62523	rape	29477	basement	18210
72	line	319492	institute	62451	palm	29472	illustration	18179
73	girl	319120	living	62399	jet	29468	essence	18172
74	father	318173	quarter	62138	navy	29450	container	18171
75	information	314341	forest	61942	plot	29444	gym	18149
76	car	313437	poll	61677	angle	29422	bat	18132
77	minute	306541	background	61431	prisoner	29394	promotion	18122
78	party	304325	winter	61394	discipline	29386	uncertainty	18119
79	back	302390	focus	61392	negotiation	29385	ownership	18114
80	health	300262	shop	61038	butter	29372	tag	18076
81	reason	295577	division	60892	desert	29353	documentary	18071
82	member	292596	advice	60861	prize	29331	stem	18061
83	community	290489	reaction	60811	bullet	29277	flood	18059
84	news	289175	injury	60769	awareness	29265	guilt	18057
85	body	287817	it	60763	sequence	29257	inside	18050
86	level	279770	ticket	60721	drop	29201	alarm	18039
87	boy	279283	grade	60710	pack	29197	turkey	18036
88	university	279132	birth	60673	provider	29196	conduct	18031
89	change	274132	painting	60548	activist	29161	fitness	18001
90	center	268821	outcome	60498	poem	29138	gulf	17985
91	face	266458	enemy	60473	vacation	29123	mommy	17978
92	food	262697	damage	60384	gang	29116	sacrifice	17936
93	history	261045	being	60354	privacy	29115	spectrum	17916
94	result	259961	storm	60280	clock	29112	dragon	17915
95	morning	259129	shape	60252	penalty	29096	bacteria	17913
96	parent	258181	bowl	60234	stomach	29089	shore	17908
97	office	258157	commission	60218	concert	29086	pastor	17896
98	research	255123	captain	60187	statistics	29070	cliff	17891
99	door	252623	ear	60160	bureau	29044	addition	17884
100	court	250891	troop	60133	wolf	29042	occupation	17878
101	moment	246362	wood	60115	recommendation	28998	slice	17862
102	street	242552	lead	59773	maker	28989	military	17849
103	policy	240181	minister	59752	impression	28975	stimulus	17842
104	table	238887	neighbor	59746	battery	28962	patent	17836
105	care	237259	software	59696	narrative	28928	powder	17830
106	process	235304	finding	59632	sake	28905	chaos	17809
107	teacher	234642	lord	59563	delivery	28903	kit	17795
108	data	234516	drive	59524	rep	28848	piano	17786
109	death	233153	temperature	59493	diversity	28848	psychologist	17738
110	experience	232376	cut	59434	string	28844	servant	17732
111	plan	231629	influence	59350	lawsuit	28834	elevator	17718
112	education	231036	procedure	59288	deputy	28806	transmission	17702
113	age	228610	respect	59226	buddy	28763	butt	17681
114	sense	226539	wave	59169	evil	28696	cowboy	17658
115	show	225884	tradition	59099	lane	28689	intent	17648
116	college	224634	actor	58856	bean	28639	counseling	17615
117	music	222767	cat	58763	sauce	28639	rhythm	17605
118	mind	220441	bus	58707	testing	28626	complexity	17603
119	class	219160	clothes	58660	trick	28621	effectiveness	17583
120	police	219041	affair	58553	fantasy	28620	strain	17571
121	use	218006	contact	58518	absence	28618	garbage	17560
122	effect	217999	category	58470	symbol	28609	venture	17532
123	season	217695	topic	58410	recognition	28585	lock	17526
124	tax	217131	victory	58342	tablespoon	28554	aide	17521
125	heart	216345	map	58218	arrest	28504	rod	17510
126	son	216216	egg	58198	approval	28499	t-shirt	17501
127	art	215981	general	58129	aids	28489	civilian	17499
128	market	210848	expression	58120	independence	28463	basket	17488
129	air	210307	session	58108	rose	28398	loser	17479
130	force	210010	competition	58102	ghost	28365	franchise	17479
131	foot	209334	technique	58028	involvement	28357	saint	17465
132	baby	207577	average	57985	wire	28337	aim	17460
133	love	206700	wine	57614	mouse	28329	prosecution	17457
134	republican	204793	youth	57132	airline	28327	bite	17445
135	interest	202642	weather	56939	founder	28315	lyrics	17443
136	security	201542	medicine	56814	objective	28315	compound	17435
137	control	200879	novel	56709	phenomenon	28258	architecture	17424
138	rate	200828	thursday	56611	silver	28244	reach	17415
139	report	200188	exchange	56534	cent	28231	destination	17414
140	nation	200021	bay	56420	teenager	28219	province	17402
141	action	198530	conclusion	56296	viewer	28181	sum	17402
142	wife	197306	progress	56162	chocolate	28147	lecture	17399
143	decision	196426	surprise	56116	grandmother	28091	trading	17380
144	value	195639	assessment	55956	missile	28086	acceptance	17373
145	phone	194941	smile	55936	adventure	28045	revelation	17367
146	thanks	194839	cash	55856	manufacturer	27966	indicator	17361
147	event	194748	pound	55670	crop	27946	collaboration	17360
148	site	193251	pair	55530	survival	27914	rhetoric	17356
149	church	191872	loan	55348	dimension	27889	monkey	17334
150	model	191448	village	55217	roll	27831	protocol	17316

151	relationship	191344	half	55207	darkness	27827	finish	17287
152	movie	190773	suit	55200	educator	27782	jeans	17273
153	field	190489	meaning	55014	personnel	27773	companion	17269
154	player	189280	supply	54988	terrorist	27765	recipient	17240
155	couple	189272	honey	54912	elite	27712	array	17233
156	record	187057	bone	54911	emphasis	27699	ranch	17219
157	difference	185310	consequence	54883	destruction	27666	scout	17206
158	light	185287	regulation	54726	chart	27628	medal	17200
159	development	185238	award	54712	pace	27613	translation	17186
160	role	184483	bottom	54701	finance	27594	ski	17185
161	view	182417	classroom	54542	champion	27593	conservation	17184
162	price	181918	reference	54390	suspect	27585	habitat	17173
163	effort	181751	lunch	54300	barrier	27530	contractor	17156
164	voice	181328	ad	54265	advocate	27479	trailer	17146
165	department	181181	desire	54265	label	27479	pitcher	17123
166	leader	180573	instruction	54207	burden	27443	towel	17122
167	photo	177976	emergency	54204	violation	27435	bonus	17099
168	space	177126	thinking	54200	idiot	27432	genre	17072
169	project	177076	tour	54157	lifetime	27432	caller	17070
170	position	176578	moon	54069	routine	27266	exit	17065
171	million	175586	address	54018	presentation	27265	hook	17057
172	film	175442	december	53957	gear	27243	pit	17033
173	need	175440	chicken	53916	stadium	27233	cholesterol	17011
174	type	173975	fuel	53882	mortgage	27227	advertisement	16993
175	town	173905	abuse	53777	sheriff	27193	bombing	16988
176	article	173819	construction	53745	clinic	27190	encounter	16987
177	road	173413	wednesday	53706	coalition	27166	expertise	16978
178	form	173161	intervention	53351	mix	27117	creator	16976
179	chance	172820	visit	53262	menu	27113	tablet	16958
180	drug	172756	coast	53114	advertising	27069	blow	16948
181	situation	171026	cover	53057	muslim	26994	ruling	16930
182	practice	170654	cop	52982	arrangement	26976	launch	16908
183	science	170488	baseball	52864	storage	26941	warming	16906
184	brother	167417	email	52826	framework	26926	equity	16905
185	matter	167351	accident	52784	utility	26839	classic	16902
186	image	167053	front	52764	cookie	26809	pine	16894
187	star	167036	duty	52673	terror	26792	surgeon	16865
188	cost	166775	revenue	52579	worth	26778	tennis	16863
189	post	166265	trend	52452	forum	26724	suffering	16837
190	society	165769	ring	52405	alliance	26718	accuracy	16834
191	picture	165703	breath	52280	empire	26692	literacy	16828
192	piece	165538	inch	52275	corn	26609	treasury	16827
193	paper	165070	neck	52214	testimony	26542	crown	16806
194	energy	164622	core	52195	transfer	26541	sergeant	16794
195	building	164119	billion	52087	stake	26539	thumb	16776
196	doctor	163129	miss	51928	tissue	26475	colony	16776
197	activity	162943	drink	51766	concentration	26470	walking	16767
198	american	162614	literature	51523	developer	26466	integration	16746
199	media	162177	mayor	51481	ballot	26460	spouse	16732
200	evidence	162050	importance	51419	chamber	26416	excitement	16728
201	product	161999	danger	51331	stick	26373	residence	16727
202	arm	161185	knee	51298	tiger	26334	adolescent	16726
203	technology	160819	ass	51238	ratio	26305	doll	16714
204	comment	160451	traffic	51138	lover	26282	oxygen	16714
205	look	160145	train	50975	expansion	26265	gravity	16674
206	term	160014	plate	50934	unemployment	26248	palace	16669
207	color	160011	equipment	50894	pet	26240	cotton	16664
208	choice	159277	studio	50758	laboratory	26198	rescue	16660
209	source	158588	secret	50730	administrator	26195	lawmaker	16643
210	mom	158511	engine	50724	quarterback	26181	determination	16633
211	director	158028	luck	50589	rocket	26167	dynamics	16623
212	rule	157468	panel	50544	preparation	26152	shell	16622
213	campaign	157373	hero	50520	relative	26146	hint	16621
214	ground	156706	circle	50510	quote	26087	rush	16589
215	election	156532	critic	50462	publisher	26079	christianity	16585
216	page	154863	episode	50354	innovation	26039	ban	16560
217	test	154189	check	50298	nut	26017	statute	16524
218	patient	154125	requirement	50272	fighter	26014	biology	16518
219	video	153563	politician	50171	rank	26007	follower	16508
220	support	152769	rain	50093	electricity	26003	trap	16420
221	rest	151864	colleague	50084	instance	26001	consent	16419
222	step	151394	beer	50024	fortune	25994	pump	16418
223	opportunity	151226	exercise	49985	variation	25982	down	16415
224	official	150423	democracy	49951	laughter	25929	hate	16397
225	oil	150410	setting	49873	wisdom	25925	trunk	16376
226	call	149896	honor	49871	correspondent	25907	belly	16361
227	organization	149261	works	49867	mixture	25895	banking	16356
228	character	148825	theme	49821	assistant	25884	asshole	16356
229	county	148165	male	49791	tomato	25826	journalism	16352
230	future	147923	pool	49670	indian	25825	flash	16348
231	dad	147520	pocket	49462	ingredient	25809	obstacle	16342
232	industry	146746	sugar	49230	galaxy	25775	ridge	16341
233	second	146634	teaching	49167	scheme	25772	bastard	16320
234	list	146495	chairman	49123	gop	25770	apology	16314
235	stuff	145992	error	49042	shame	25764	tumor	16309
236	figure	145812	bridge	49030	contest	25761	architect	16305
237	attention	145669	flower	48912	introduction	25722	wrist	16293
238	risk	145334	universe	48854	boundary	25694	harbor	16291
239	fire	144750	limit	48738	tube	25647	bullshit	16278
240	dog	144648	coverage	48693	chef	25646	realm	16277
241	hair	143875	crew	48673	load	25620	bet	16276
242	condition	143732	balance	48537	tongue	25612	inspector	16272

243	wall	143389	lip	48507	creek	25595	surveillance	16262
244	daughter	142747	zone	48403	hip	25581	trauma	16257
245	deal	142589	wedding	48365	decline	25563	romance	16244
246	author	142352	copy	48365	terrorism	25562	deadline	16240
247	truth	142017	joke	48299	respondent	25555	compensation	16221
248	husband	141289	meal	48158	narrator	25554	insect	16198
249	period	141173	minority	48052	voting	25547	output	16190
250	series	140231	sight	48022	refugee	25533	parliament	16176
251	order	140025	sleep	48020	assembly	25533	suite	16164
252	officer	139938	dress	47994	fraud	25518	separation	16144
253	land	139579	release	47967	limitation	25511	demon	16143
254	computer	139082	profit	47886	partnership	25493	eating	16142
255	thought	139016	gender	47825	crash	25472	equality	16126
256	economy	138752	tape	47805	representation	25448	probability	16109
257	goal	138693	ocean	47773	hold	25448	acquisition	16072
258	bank	138673	host	47739	ministry	25434	custody	16070
259	behavior	138400	circumstance	47687	excuse	25386	compromise	16070
260	act	137668	boss	47643	comedy	25303	trash	16056
261	north	137666	tea	47503	purchase	25302	earthquake	16050
262	blood	137579	manner	47282	infrastructure	25291	wake	16026
263	culture	137140	percentage	47208	islam	25282	necessity	16013
264	top	136658	cloud	47140	port	25211	horn	16012
265	language	136124	winner	47124	fate	25194	parameter	16010
266	window	136043	sentence	46932	dialogue	25173	grip	16005
267	response	135942	request	46865	plus	25166	mathematics	15994
268	population	135733	appearance	46855	frequency	25158	hallway	15988
269	tree	135666	salt	46825	tribe	25147	liability	15952
270	park	135440	beauty	46819	appointment	25144	fluid	15938
271	worker	135244	bottle	46801	establishment	25090	hometown	15934
272	earth	134202	component	46800	shopping	25052	exam	15933
273	cause	133998	fee	46768	celebrity	24989	anchor	15923
274	race	133182	dance	46601	congressman	24982	vitamin	15887
275	letter	132828	shirt	46579	taxpayer	24954	hostage	15872
276	gun	132572	tip	46537	pride	24923	mill	15859
277	course	131911	plastic	46455	clothing	24923	processing	15825
278	hell	131265	tooth	46405	domain	24914	sentiment	15815
279	answer	130534	meat	46401	ceiling	24902	offering	15814
280	nature	130378	stress	46396	shelter	24892	photography	15803
281	administration	130297	february	46378	obligation	24879	coin	15797
282	message	129799	constitution	46355	shower	24876	laptop	15796
283	song	129648	definition	46347	musician	24829	goodness	15790
284	congress	128923	uncle	46291	dirt	24819	affiliation	15784
285	attack	128893	metal	46260	pc	24814	bee	15760
286	amount	128075	album	46247	bell	24813	blessing	15758
287	analysis	128029	self	46225	smell	24783	landing	15743
288	store	127984	investor	46202	ceremony	24779	repair	15742
289	defense	127924	fruit	46189	clue	24755	worry	15739
290	bill	127919	desk	46133	guideline	24752	ritual	15737
291	cell	127818	valley	46050	comfort	24745	bath	15725
292	performance	127318	abortion	45934	pregnancy	24687	mud	15699
293	hospital	127317	chapter	45911	tourist	24638	reminder	15692
294	bed	127232	commitment	45809	incentive	24612	slavery	15684
295	board	127202	proposal	45546	custom	24610	supervisor	15672
296	century	126668	capacity	45515	cheek	24607	quantity	15672
297	summer	126418	nose	45433	tournament	24576	olympics	15664
298	material	126254	variable	45399	satellite	24573	slope	15651
299	individual	125578	convention	45389	medication	24549	skirt	15648
300	example	125442	jury	45342	script	24547	outlet	15645
301	animal	124835	root	45327	psychology	24479	curtain	15640
302	factor	123915	incident	45300	logic	24470	declaration	15635
303	sir	123629	hearing	45235	economics	24463	switch	15630
304	agency	123524	payment	45132	update	24444	calendar	15617
305	help	123017	bear	45125	devil	24433	paragraph	15612
306	ability	122943	surgery	45042	beat	24421	quest	15593
307	mile	122899	shadow	45022	charity	24405	flat	15583
308	statement	122852	witness	44979	fiber	24394	entrepreneur	15581
309	democrat	122325	chest	44918	friendship	24359	clay	15557
310	floor	122291	amendment	44879	net	24355	noon	15551
311	career	122177	extent	44804	motivation	24342	elbow	15543
312	dollar	121984	pleasure	44800	reserve	24319	unity	15519
313	sex	121352	holiday	44775	observer	24293	rent	15515
314	south	120894	talent	44687	humanity	24254	airplane	15512
315	subject	120773	expectation	44646	survivor	24240	portfolio	15509
316	decade	120539	ride	44508	fence	24223	northwest	15505
317	bit	120467	wing	44427	humor	24213	sweat	15505
318	sister	120346	corporation	44407	major	24191	treasure	15484
319	quality	120236	sector	44401	funeral	24184	flour	15472
320	press	119485	fat	44341	spokesman	24143	rally	15471
321	sound	119028	experiment	44341	extension	24142	default	15454
322	plant	118323	farmer	44255	historian	24128	plain	15447
323	movement	118118	english	44219	chemical	24096	exploration	15421
324	scene	118003	confidence	44154	singer	24095	limb	15391

325	section	117170	bunch	43989	onion	24049	charter	15390
326	treatment	116893	speaker	43781	specialist	24031	southwest	15385
327	benefit	116806	breast	43780	white	24014	escape	15376
328	west	116683	contribution	43738	pan	24011	arena	15361
329	candidate	116347	leaf	43674	deck	23996	witch	15359
330	approach	116173	interaction	43639	reflection	23994	courtesy	15350
331	resource	116078	hat	43629	database	23984	statue	15343
332	sort	115486	doubt	43620	permission	23927	fist	15331
333	size	115388	promise	43606	lifestyle	23917	castle	15330
334	knowledge	115356	nurse	43524	lion	23871	squad	15328
335	sport	114894	gene	43463	consumption	23864	cruise	15327
336	tv	114859	package	43448	capability	23836	embassy	15303
337	loss	114630	difficulty	43423	emission	23770	patience	15298
338	note	114251	daddy	43376	sidebar	23763	bush	15262
339	meeting	114209	era	43331	crap	23756	outfit	15213
340	skill	113912	aid	43250	dealer	23742	cage	15211
341	card	113472	notion	43225	measurement	23736	shark	15207
342	feeling	113414	explanation	43197	adviser	23663	lobby	15206
343	degree	113046	airport	43177	imagination	23622	sidewalk	15203
344	crime	112978	chain	43156	tragedy	23619	runner	15188
345	sign	112603	channel	43041	consultant	23591	ankle	15182
346	vote	112405	tear	43024	lieutenant	23579	attraction	15180
347	king	112088	estate	43023	acre	23558	mercy	15139
348	box	112035	offer	42944	drawing	23549	tune	15126
349	disease	111433	guide	42871	scandal	23525	dancer	15125
350	lady	111384	bond	42742	ambassador	23522	candle	15125
351	design	111026	birthday	42727	bench	23481	needle	15115
352	concern	110976	travel	42709	counter	23475	chronicle	15106
353	ball	110770	improvement	42657	odds	23462	suburb	15106
354	east	110589	trail	42439	rat	23442	sensor	15084
355	network	110210	fashion	42405	horror	23430	debut	15081
356	success	110030	length	42379	prevention	23383	magnitude	15074
357	district	109941	sheet	42366	segment	23372	suspicion	15070
358	cup	109915	funding	42356	tail	23344	icon	15065
359	growth	109567	fault	42283	implementation	23294	grandma	15063
360	standard	107818	motion	42232	strip	23266	info	15051
361	fan	107631	combination	42138	preference	23230	jurisdiction	15049
362	theory	107486	description	42091	feedback	23220	senior	15032
363	staff	107396	snow	42027	health-care	23216	parade	15028
364	september	106932	marketing	41781	criminal	23214	seal	15019
365	set	106920	bomb	41679	can	23201	archive	15007
366	strategy	106236	muscle	41650	buyer	23157	rage	14996
367	property	105851	rise	41609	cooperation	23135	ending	14992
368	authority	105586	creation	41523	profession	23135	loop	14987
369	method	105544	spending	41483	killing	23084	chase	14979
370	region	105382	instrument	41473	photographer	23062	reception	14976
371	impact	105330	mass	41453	engagement	23041	local	14975
372	committee	105133	noise	41144	paint	23029	premise	14964
373	dream	104797	e-mail	41092	pitch	23028	whale	14947
374	training	104563	cycle	41088	salad	23010	credibility	14944
375	shit	104498	signal	41085	stair	23001	loyalty	14935
376	option	104245	app	41082	notice	22997	tide	14929
377	opinion	103920	perception	41010	grace	22995	traveler	14928
378	interview	103566	visitor	41006	divorce	22992	grief	14918
379	play	103352	contrast	40982	vessel	22977	metaphor	14916
380	union	102805	judgment	40961	pig	22965	skull	14911
381	professor	102741	representative	40894	assignment	22955	pursuit	14909
382	pressure	102669	pass	40865	distinction	22938	therapist	14907
383	purpose	102642	regime	40863	circuit	22927	backup	14904
384	stage	102611	producer	40843	acid	22925	workplace	14897
385	sun	102043	route	40816	efficiency	22889	instinct	14894
386	pain	101918	lie	40794	memorial	22884	export	14889
387	artist	101861	analyst	40785	blue	22857	broadcast	14871
388	employee	101818	female	40661	entity	22849	execution	14857
389	account	101683	disability	40559	iphone	22844	pal	14857
390	fund	101503	comparison	40525	punishment	22839	density	14854
391	environment	101479	rating	40458	pill	22802	correction	14850
392	version	101091	campus	40453	myth	22770	jump	14819
393	shot	101016	branch	40300	economist	22756	teammate	14789
394	reality	100700	gentleman	40277	craft	22724	brush	14780
395	club	100411	opposition	40268	poetry	22722	corridor	14779
396	justice	100360	hit	40121	thread	22683	russian	14778
397	river	100270	athlete	40094	privilege	22671	enthusiasm	14777
398	brain	100235	opening	40053	coal	22670	panic	14761
399	memory	100235	prayer	40037	fool	22667	pad	14749
400	rock	100028	basketball	40007	cow	22660	bid	14746
401	talk	99986	revolution	39959	pie	22654	productivity	14732
402	camera	99855	emotion	39949	resort	22647	guess	14729
403	detail	99477	fuck	39924	legacy	22631	defeat	14716
404	challenge	99455	platform	39915	scope	22460	railroad	14713
405	argument	99403	frame	39817	formation	22441	inspection	14686
406	weapon	98828	appeal	39798	grandfather	22432	cab	14672
407	station	98720	potential	39779	wish	22423	nonsense	14670
408	island	98596	struggle	39777	margin	22417	rider	14637
409	discussion	98359	brand	39754	wound	22414	theology	14636
410	trouble	97439	legislation	39690	exhibition	22400	accent	14628
411	conversation	97193	lab	39668	legislature	22388	invitation	14626

412	date	97094	row	39635	portrait	22377	reply	14613
413	public	97031	immigration	39570	catholic	22367	israeli	14606
414	army	97025	asset	39563	uniform	22332	liar	14604
415	charge	96872	observation	39551	miracle	22275	registration	14577
416	seat	96853	taste	39481	harm	22267	handle	14576
417	writer	96279	household	39430	tactic	22260	momentum	14567
418	credit	95981	existence	39315	mask	22256	clerk	14547
419	marriage	95885	mirror	39284	inflation	22242	chin	14538
420	operation	95878	pilot	39272	hunting	22232	hockey	14536
421	agent	95371	stand	39230	final	22206	laser	14523
422	access	95371	relief	39223	flesh	22204	proposition	14515
423	bar	95305	milk	39204	fellow	22194	beam	14505
424	debate	95244	warning	39194	nerve	22174	ancestor	14501
425	leg	95185	heaven	39105	lung	22170	creativity	14474
426	glass	94803	flow	39098	headline	22141	verse	14470
427	adult	94547	calorie	39066	defendant	22117	objection	14466
428	trade	94509	seed	39062	pole	22115	carbohydrate	14454
429	review	94364	bike	38920	entrance	22111	steam	14454
430	association	94100	employer	38895	aircraft	22111	glance	14445
431	measure	93952	disaster	38801	cabinet	22078	forehead	14437
432	stock	93809	sale	38751	happiness	22053	slide	14433
433	gas	93763	bathroom	38712	admission	22052	sheep	14430
434	lawyer	93526	agenda	38615	correlation	22034	carpet	14417
435	production	93507	enforcement	38612	proportion	22031	cloth	14408
436	management	93289	diet	38595	dispute	22027	interior	14401
437	victim	93207	mark	38593	candy	21987	running	14396
438	cancer	93157	silence	38552	reward	21978	questionnaire	14392
439	speech	92837	journalist	38495	counselor	21975	departure	14389
440	trial	92762	bible	38492	recording	21968	graph	14381
441	item	92545	queen	38445	pile	21960	thief	14374
442	weight	92516	expense	38330	explosion	21954	herb	14373
443	citizen	92182	cream	38234	couch	21942	subsidy	14371
444	trip	91830	exposure	38231	furniture	21915	cast	14354
445	executive	91696	priority	38168	significance	21902	fossil	14345
446	politics	91621	soil	38159	magic	21888	patrol	14342
447	customer	91563	angel	38155	commissioner	21855	pulse	14341
448	manager	91527	journey	38098	shelf	21854	mechanic	14340
449	conference	91185	trust	38062	warrior	21822	cattle	14338
450	base	90915	tank	38032	physics	21798	screening	14328
451	direction	90826	cheese	38025	garage	21790	dignity	14312
452	sunday	90771	schedule	38022	flavor	21772	prophet	14308
453	past	90674	bedroom	37979	oven	21721	commentary	14305
454	majority	90548	tone	37956	satisfaction	21718	spread	14295
455	peace	90518	selection	37940	discrimination	21715	servings	14291
456	dinner	90516	wheel	37921	recession	21714	homework	14277
457	partner	90455	gap	37884	allegation	21709	android	14261
458	user	90342	veteran	37880	boom	21690	alpha	14253
459	bag	90234	characteristic	37848	restriction	21665	mortality	14244
460	owner	89612	protein	37846	diamond	21652	elephant	14238
461	voter	89245	resolution	37841	conviction	21599	continent	14230
462	tool	89236	whole	37840	heel	21595	attribute	14222
463	june	89233	engineer	37679	fame	21544	ecosystem	14216
464	may	88994	walk	37624	swing	21541	leave	14202
465	mountain	88984	dish	37580	playoff	21528	syndrome	14192
466	range	88892	waste	37529	actress	21526	minimum	14187
467	coach	88826	print	37504	format	21521	catch	14181
468	fear	88724	depression	37492	controversy	21521	accusation	14179
469	friday	88632	dude	37467	auto	21510	sock	14170
470	attorney	88616	present	37433	grant	21506	pickup	14166
471	expert	88134	ceo	37374	grocery	21497	shuttle	14163
472	structure	88114	initiative	37346	headquarters	21481	calculation	14158
473	budget	88084	gate	37317	shade	21457	jaw	14133
474	insurance	88036	entry	37187	meter	21425	nutrition	14118
475	text	87811	assistance	37154	pipe	21413	mentor	14114
476	freedom	87806	height	37131	celebration	21353	stance	14103
477	reader	87422	phase	37013	handful	21346	rabbit	14099
478	style	87244	conservative	36998	copyright	21330	pause	14099
479	mark	87215	criticism	36941	signature	21285	dot	14080
480	machine	87156	faculty	36906	bishop	21282	contributor	14076
481	november	87108	achievement	36879	soup	21273	disk	14075
482	generation	87071	alcohol	36823	carrier	21245	likelihood	14049
483	income	86902	therapy	36823	anniversary	21242	sibling	14047
484	sea	86544	offense	36790	pizza	21238	southeast	14046
485	mouth	86406	touch	36766	ethics	21228	painter	14037
486	web	85868	killer	36754	legend	21223	counterpart	14033
487	threat	85754	personality	36750	eagle	21220	believer	14032
488	solution	85661	landscape	36719	scholarship	21183	lamp	14028
489	scientist	85428	transition	36569	crack	21181	inmate	14023
490	investigation	84783	column	36526	membership	21172	trace	14018
491	senator	84544	button	36451	standing	21145	filter	14005
492	unit	84518	opponent	36412	possession	21141	pillow	14004
493	photograph	84476	immigrant	36379	treaty	21138	orbit	14003
494	july	84439	distribution	36338	consciousness	21114	purse	13998
495	television	84433	golf	36327	manufacturing	21096	feed	13963
496	radio	84218	alternative	36296	announcement	21081	clause	13942
497	senate	83931	favorite	36231	makeup	21035	left	13935
498	violence	83882	stop	36214	pop	21033	pond	13934
499	feature	83853	portion	36150	prediction	21032	shield	13931
500	audience	83493	suicide	36059	stability	21031	popularity	13929
501	evening	83312	deficit	36029	norm	21014	cartoon	13928
502	hall	82960	symptom	36021	genius	20987	exhibit	13890
503	task	82839	complaint	35954	operator	20947	sink	13890
504	score	82738	supporter	35863	mall	20944	graphic	13890
505	skin	82598	philosophy	35824	rumor	20921	darling	13887

506	spring	82557	childhood	35815	poet	20918	vendor	13880
507	safety	82535	well	35789	tendency	20906	gathering	13824
508	weekend	82429	dark	35705	alien	20885	fever	13809
509	while	82196	wealth	35662	maintenance	20861	arrow	13796
510	title	82165	welfare	35639	courage	20859	capitalism	13777
511	heat	82112	poverty	35629	principal	20834	kick	13771
512	hope	81896	prosecutor	35613	vaccine	20826	fork	13767
513	yard	81852	shift	35516	identification	20823	meantime	13763
514	finger	81817	back	35423	sandwich	20804	stay	13749
515	mission	81524	display	35418	bull	20803	illusion	13749
516	participant	81371	knife	35408	lens	20785	removal	13735
517	hotel	81364	round	35407	presidency	20723	arab	13720
518	judge	81312	tech	35394	integrity	20719	organism	13713
519	pattern	81097	detective	35355	bias	20699	sculpture	13689
520	break	81057	disorder	35345	hypothesis	20698	marine	13677
521	institution	80998	creature	35342	nomination	20678	harassment	13671
522	faith	80912	housing	35224	delay	20668	drum	13670
523	folk	80843	watch	35203	adjustment	20658	helmet	13668
524	surface	80819	chip	35203	sanction	20651	certIFICATE	13657
525	fall	80585	tie	35058	examination	20620	cart	13625
526	client	80584	command	35010	currency	20592	spy	13625
527	edge	80577	shooting	34970	flame	20550	sunlight	13624
528	council	80469	layer	34911	guitar	20550	rookie	13617
529	device	80398	bread	34840	doctrine	20549	hunger	13609
530	firm	80364	exception	34835	palestinian	20540	practitioner	13606
531	responsibility	80294	passion	34825	commerce	20539	performer	13603
532	chair	80186	highway	34789	robot	20517	jar	13598
533	internet	80144	commander	34731	particle	20474	programming	13597
534	october	80099	publication	34656	glove	20469	dawn	13590
535	investment	79833	win	34613	jazz	20420	salmon	13579
536	ship	79715	mystery	34553	infant	20410	census	13578
537	content	79496	championship	34425	summary	20407	pick	13577
538	consumer	79489	tale	34400	server	20393	accomplishment	13575
539	element	79440	liberty	34393	leather	20393	conscience	13568
540	spirit	79123	passenger	34335	radiation	20383	molecule	13548
541	track	78828	physician	34292	composition	20367	threshold	13540
542	run	78645	graduate	34285	discourse	20342	inventory	13538
543	wind	78623	substance	34262	wildlife	20304	transport	13531
544	lack	78587	atmosphere	34247	soccer	20295	ranking	13487
545	journal	78364	passage	34197	complex	20289	gram	13479
546	ice	78316	pepper	34195	mandate	20285	puzzle	13476
547	collection	78281	cable	34164	monitor	20280	envelope	13474
548	soldier	78276	square	34163	nightmare	20245	cure	13442
549	governor	78194	recipe	34098	barrel	20229	discount	13415
550	fish	78114	roof	34005	globe	20226	addiction	13411
551	shoulder	78062	vegetable	34000	feel	20197	marker	13409
552	saturday	77290	habit	33933	gesture	20188	juror	13406
553	editor	77280	discovery	33930	tent	20170	mention	13404
554	claim	77050	total	33836	receiver	20165	businessman	13380
555	agreement	76793	recovery	33830	horizon	20141	premium	13379
556	researcher	76737	dna	33819	diagnosis	20070	retailer	13374
557	concept	76679	gain	33808	gospel	20013	liver	13374
558	band	76423	territory	33787	fighting	20007	pirate	13368
559	restaurant	76197	girlfriend	33779	stroke	20004	protester	13352
560	league	76178	proof	33755	duck	19968	outlook	13350
561	capital	76018	parking	33735	grain	19956	elder	13336
562	april	75906	sin	33703	beast	19956	gallon	13333
563	magazine	75615	rice	33676	remark	19951	ignorance	13323
564	status	75577	engineering	33630	fabric	19948	chemistry	13323
565	hill	75365	advance	33626	civilization	19929	weed	13310
566	kitchen	75167	cake	33519	corruption	19925	babe	13309
567	screen	74705	flag	33516	collapse	19925	fraction	13308
568	mistake	74643	good	33457	ma'am	19916	cook	13308
569	battle	74296	jail	33453	workshop	19894	conversion	13304
570	spot	74296	match	33368	inquiry	19877	merit	13286
571	corner	74047	interpretation	33342	cd	19876	citizenship	13283
572	target	73925	wage	33339	rifle	19857	coordinator	13280
573	driver	73902	breakfast	33309	closet	19855	validity	13278
574	beginning	73853	profile	33259	reporting	19850	european	13273
575	religion	73744	saving	33236	curve	19846	lightning	13272
576	crisis	73691	reduction	33167	patch	19841	turtle	13267
577	museum	73559	bitch	33157	touchdown	19832	ambition	13267
578	communication	73413	entertainment	33057	earnings	19824	comic	13245
579	murder	73388	academy	33028	hunter	19821	soap	13244
580	object	73304	teaspoon	33027	fly	19815	shortage	13236
581	blog	73163	ally	32994	tunnel	19814	booth	13234
582	return	73109	count	32971	corps	19808	diary	13233
583	belief	73067	avenue	32968	german	19765	transit	13220
584	block	73065	priest	32965	weakness	19718	publishing	13216
585	debt	73015	employment	32957	doc	19715	curiosity	13208
586	labor	72759	grass	32880	medium	19713	grid	13205
587	understanding	72651	scholar	32846	virtue	19711	bow	13189
588	neighborhood	72625	edition	32842	ounce	19680	disclosure	13185
589	contract	72568	infection	32785	confusion	19634	rival	13176
590	middle	72563	suggestion	32775	midnight	19599	denial	13171
591	species	72542	resistance	32768	deer	19597	speculation	13167
592	sample	72489	prince	32712	heritage	19592	sympathy	13166
593	path	72335	illness	32704	summit	19581	sodium	13145
594	apple	72273	republic	32646	sword	19580	ipad	13111
595	god	72022	volunteer	32631	telescope	19572	sack	13104
596	library	71921	evaluation	32514	blade	19562	barn	13093

597	prison	71871	mine	32473	toe	19559	reliability	13090
598	hole	71837	black	32406	agriculture	19558	hormone	13088
599	attempt	71833	iron	32359	dose	19458	footage	13085
600	code	71659	phrase	32319	prescription	19428	alley	13070
601	sales	71601	assumption	32317	adoption	19416	ease	13065
602	gift	71372	sand	32307	grave	19407	cafe	13055
603	increase	71248	designer	32303	rope	19405	vampire	13030
604	garden	71243	planning	32282	transaction	19400	dessert	13024
605	lake	70940	mode	32229	lawn	19393	well-being	13022
606	turn	70831	occasion	32220	bubble	19365	northeast	13021
607	plane	70679	festival	32159	princess	19338	fleet	13005
608	vehicle	70662	cap	32115	log	19318	availability	13004
609	application	70441	mechanism	32060	blanket	19318	compliance	12998
610	thousand	70404	intention	32022	laugh	19315	pin	12991
611	coffee	70366	scenario	32007	kingdom	19315	pork	12986
612	file	70059	transportation	31888	bug	19312	astronomer	12976
613	reform	70010	pot	31864	primary	19308	like	12976
614	gold	69903	boyfriend	31817	nominee	19303	installation	12961
615	planet	69822	consideration	31756	transformation	19286	nest	12951
616	location	69669	retirement	31749	temple	19285	goat	12944
617	bird	69662	cigarette	31674	arrival	19265	equivalent	12941
618	principle	69403	factory	31626	frustration	19262	willingness	12933
619	search	69392	clip	31612	demonstration	19226	banker	12931
620	possibility	69316	giant	31557	pollution	19218	interval	12928
621	sky	69285	slave	31539	poster	19209	gasoline	12925
622	horse	68883	christian	31496	nail	19206	bucket	12920
623	advantage	68866	wonder	31463	cry	19190	theft	12907
624	basis	68758	participation	31395	guidance	19172	laundry	12903
625	context	68668	stream	31393	pen	19160	constraint	12902