PNAS
www.pnas.org

Supporting Information for

Dominant words rise to the top by positive frequency-dependent selection

Mark Pagel, Mark Beaumont, Andrew Meade, Annemarie Verkerk and Andreea Calude

Author for correspondence: Mark Pagel
Email: m.pagel@reading.ac.uk

This PDF file includes:

Supporting Information Text

Data

LAMSAS items. We obtained LAMSAS interview response data from materials made available by the Applied Linguistics department at the University of Groningen (http://www.let.rug.nl/~kleiweg/lamsas/overview/lex.txt) as well as from the Linguistic Atlas Project website (http://old.lap.uga.edu/, http://old.lap.uga.edu/cgi-bin/lapsite.fcgi/lamsas/). One hundred meanings were listed in the LAMSAS Handbook(1) as 'lexical' or 'grammatical'. From this set, we excluded seven food-related concepts as they seemed to collect responses to a range of concepts rather than a single one, for instance BAKERS BREAD. This left n=93 LAMSAS concepts for analysis (Table S1).

LAGS items. We collected LAGS concepts from the Linguistic Atlas Project website (http://old.lap.uga.edu/cgi-bin/lapsite.fcgi/lags/). There were some discrepancies between the LAGS Technical Index and the files retrieved from the Linguistic Atlas Project website, and where this occurred we gave priority to the files from the Linguistic Atlas Project website. We obtained 345 concepts from LAGS identified as 'lexical' or 'grammatical'. From this set, we excluded twenty concepts that seemed poorly defined by virtue of failing to elicit consistent responses, for instance the item POOR WHITES (WHITES' TERMS). This left 325 LAGS items (Table S2).

Filtering criteria. All n=418 retained LAMSAS and LAGS data were lemmatized (e.g., plurals, such as aprons, were combined with their singular, apron) and inappropriate responses were removed. These included responses not belonging to the intended word class (for instance, a verb 'clabbered' for the noun CLABBER), responses where the interviewee was talking about something else (for instance, 'spider webs' for BAKER'S BREAD), when a response was noted down for the wrong concept (for instance, 'pieces' for RIND).

We merged some responses and deleted others following the criteria in Table S4, always with a view towards identifying the actual word alternatives that could be thought to be competing within a population of speakers. The single most common action we took was to delete irrelevant context around the key response, such as for 'apron' (LAGS item 0011). Responses included red apron, bib apron and cook's apron, all of which we took to be instantiations of the concept 'apron'. Our criteria always reduce rather than create word alternatives. Our merging and exclusion operations affected roughly 1% of the 417,000 responses.

Number of respondents per meaning. Interviews did not always include all items and respondents sometimes failed to give a reply for some meanings or said 'don't know'. The average number of respondents per meaning was 729±220 (range 50 – 1152). For the 914 LAGS respondents it was 669±147 (range = 63-905, median = 698), and for the 1162 LAMSAS respondents it was 963±292 (range = 83 – 1152, median = 1057).

The number of respondents per meaning does not correlate with model outcomes (all p-values > 0.36) and the distribution of model outcomes is the same when we restrict the

analyses to meanings with > 300 respondents (n=401): *Drift(LAGS, LAMSAS)* :8.7 (8.9,8.6) *Frequency-dependent selection(LAGS, LAMSAS)*: 18.5 (18.2, 16.1) *Directional selection(LAGS, LAMSAS)*: 72.8(72.9,75.3).

Multiple responses. Most interviewees provided a single word for each meaning but sometimes reported two or more. The interview records do not make it possible to identify the first word a respondent gave for a particular meaning, so we used all responses. This could introduce a degree of non-independence in the data, but this degree is low. For the combined data set of n=418 meanings both the median and modal number of responses is 1. The number of responses per respondent ranges across meanings from 1 - 2.36, with an average of 1.37±0.27. The 1162 individuals interviewed for the LAMSAS project provided an average of 1.39±0.24 responses per individual (range 1 – 1.93) for the 93 items. The 914 individuals interviewed for LAGS provided an average of 1.36±0.28 responses per individual (range 1 – 2.36). LAMSAS item CLOUDING UP drew the highest number of responses per individual (average = 1.93) and for LAGS it was TROUSERS (2.36).

Among parts of speech, only adjectives drew slightly more responses per individual (mean = 1.53±0.26), and there was no relationship of responses per individual either to 'concreteness' (see *Word Attributes*, below) or to COCA frequencies (see *Frequency of use in Corpus of Contemporary American English,* below).

Paired LAMSAS-LAGS items. Sixty-six meanings are found in both LAGS and LAMSAS. The correlation between the number of alternative word forms per meaning for these meanings is 0.87 ($p<0.0001$), and in n=58 (66%) cases they share the same highest frequency word. Other correlations on the summary statistics (see *Summary statistics*, below for definitions) are as follows: (heterozygosity, *H*, r=0.66, $p<0.0001$; frequency of most common word, r=0.42, $p<0.0004$; ratio of frequencies of 2nd to 1st word, r=0.21, p=0.09). The correlation between the paired posterior model probabilities was: *FDS*, r=0.37 ($p<0.001$), *D*, r = 0.18 (p=0.075) and for *DS*, r = 0.15 (p=0.11). The correlations are modest because the paired summary statistics themselves have only modest correlations (especially the *2/1 ratio*) and the summary statistics are good predictors of a model's posterior probability: Drift: R^2 = 0.90; Directional selection: R^2 = 0.82; Frequency-dependent selection: R^2 = 0.88. These results reinforce the stochastic nature of word-use evolution relatively divorced from characteristics of the words themselves (Figure 4, main text).

Word and Meaning characteristics

Frequency of appearance in Corpus of Contemporary American English (COCA). We identified the most commonly reported word given for each of the meanings in our sample. We then consulted the Corpus of Contemporary American English (COCA(2)), and recorded that word's frequency-of-appearance (written and spoken use), noting its rank-order position in the list. The LAP meanings are drawn from parts of the lexicon that are frequently used in everyday speech and writing compared to the average frequency of use in the COCA (Figure S1): geometric mean of COCA = 1.49±6.83 appearances per million, geometric mean for n=368 LAMSAS and LAGS meanings that have a match in the COCA = 19.16±16.42. COCA frequency does not correlate with any of the models' posterior probabilities.

Pronunciation. We scored each of the alternative words in a representative sample of 232 of the meanings, comprising 252,506 responses, on four characteristics: *complexity*, *length*, *obstruents* and *sonorants*. The n=232 meanings do not differ statistically from the remaining 185 meanings on any of the summary statistics or on their distribution among the parts of speech. *Complexity* is the number of words in the reply (some replies took the form of short expressions such as *a little way* or *help yourself*); *length* refers to the number of distinct sounds or phones in the reply; *obstruents* (the number of consonants that obstruct airflow) and *sonorants* (non-obstructive consonants).

Concreteness scores. We obtained 'concreteness' rankings for 40,000 commonly used English words and two-word expressions(3), where concreteness was defined as the extent to which the meaning refers to something that can be experienced directly through the senses (1-5 scale where 5 is concrete and 1 is abstract). We found matches or near-matches in this list to the highest frequency word for n=292 of the meanings in our sample of n=418. The concreteness scores correlate r=0.94 with concreteness ratings obtained from an earlier study of 4291 words(4). Meanings with higher concreteness scores had slightly more respondents: (r^2=0.02, p=0.0015).

Age of Acquisition. We recorded the mean age of acquisition(5) for each of our meanings. We found, as above, matches or near matches to n=312 of our meanings.

Methods

Approximate Bayesian Computation. ABC (6, 7) provides a way to sample from the posterior distribution of a model's parameters most likely to have given rise to an observed set of data, *y* (main text, **Materials and Methods**).

Model estimation. The set of *n* responses for each meaning describes a frequency distribution *f(y)* in which the *k* different words used to describe that meaning are arranged in rank-order from the word with the largest number of speakers to the fewest (Figure 1, main text). Assume *f(y)* arises from a genealogical process according to parameters $\Theta=(\theta,Z)$, where θ is the mutation rate and Z denotes those parameters related to selection (equation 1, main text).

Given a frequency distribution of alternative words for a given meaning, we use ABC to discover the parameters of Θ that are sampled approximately from $p(\Theta_i|y)$. For our purposes, θ is an unknown nuisance parameter that the ABC procedure integrates out. The remaining parameters of Θ are those related to selection in the frequency-dependent and the directional models. In the former case, this is the posterior distribution on *s*, and in the latter the posterior estimates of the parameters describing the distribution of fitness values of alternative words w_i, (note that the w_i are not themselves distinct parameters and are not estimated. Rather they are assigned to simulated word-tokens (a label within the simulation, not an actual word) at random based on draws from the prior, see below for descriptions of the priors).

Description of priors. We chose generous priors for the parameters of the models, based on variation we observed in a preliminary inspection of approximately 10% (n=48) of the LAMSAS and LAGS items.

Mutation rate θ (all models): ~*Gamma(5.5, 2.2)*

Directional selection parameters for the distribution of w_i (model DS): W_δ (*pdel upper)* ~ U(0.75,1), W_α *(pdel lower*) ~ U(0,0.25). These two parameters describe a mixture distribution centred at 1, falling away to the left (deleterious) and right (advantageous) of 1 in a manner roughly corresponding to exponential decline (8). The mixture distribution provides a rate multiplier by which the current value of w_i for word *i* is changed to a new value according to $r_i w_i$. Thus, most multipliers leave the value of w_i relatively unchanged but some improve it and some make it worse.

Frequency dependent selection parameter s (Model FDS):
log-normal ~ lnN(-5.44, 0.63)

Model selection. We ran a large number of simulations for each model following the sequential ABC procedure outlined in Beaumont(6, 9) assessing simulated data by a function $\rho(S(x_i),S(y))$ measuring the distance of the simulated data (x_i) to the observed data. Our inferences are based on 50,000 simulated points from the final round of simulation. Distance is measured on a set of summary statistics *S(.)*,with data rejected if $\rho(S(x_i),S(y)) > \varepsilon$. We chose ε such that only the top 5% of runs were retained.

We use four summary statistics to define the distance measure: number of alternative words, highest word frequency, the ratio of second highest to highest word frequency and heterozygosity (*H*), where *H* is defined as

$$H = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} f_i - 1}\left(1 - \frac{\sum_{i=1}^{k} f_i^2}{\left(\sum_{i=1}^{k} f_i\right)^2}\right)$$

and f_i is the frequency of word *i*, and *k* is the number of alternative words. For large *k* and $f_i \approx 1/k$, $H \rightarrow 1$.

Posterior probability of models. We compete the three models (drift, directional selection and frequency-dependent selection) by calculating a model's posterior probability in the usual way as

$P(M_i|D) = \frac{P(D|M_i)p(M_i)}{\sum_i P(D|M_i)p(M_i)}$ where $P(M_i|D)$ is the probability of the data under model *i*, and $p(M_i)$ is the prior probability of model *i*. $P(D|M_i)$ is calculated as the proportion of simulations in which model *i* best describes the summary statistics. In an ABC setting, a model's posterior probability is proportional to the number of simulations (out of a large number) for which the model best matched the *S(y)*.

Across meanings, the summary statistics predicted the models' posterior probabilities well: *D* (R^2 = 0.90), *DS* (R^2= 0.82), *FDS* (R^2=0.88) showing that the summary statistics do a good job of describing the observed frequency distributions.

Validation of ABC models

We simulated 100 datasets each under drift, directional selection and positive frequency-dependent selection. We simulated data using the median values of the parameters for θ, selection and the prior distribution of weighting terms (DS model) as calculated from each of the models' winners in the real data. We then competed the models in their own and in the others models' datasets.

The percentage results for the three models are as follows, with the percentages in bold down the main diagonal identifying correct outcomes.

	Analysis Model			
Simulation model	Drift, n=100	Frequency-dependent selection, n=100	Directional selection, n=100	Total
Drift	**91**	6	3	100
Frequency-dependent selection	0	**85**	15	100
Directional selection	3	7	**90**	100

In 88.67% of cases the best fitting model corresponds to the model used to simulate the data. Datasets simulated under drift are the most distinctive, differing from the two selection models. Directional selection and frequency-dependent selection can sometimes masquerade as the other process. In particular frequency-dependent selection can sometimes produce results that look like directional selection.

Figure S1

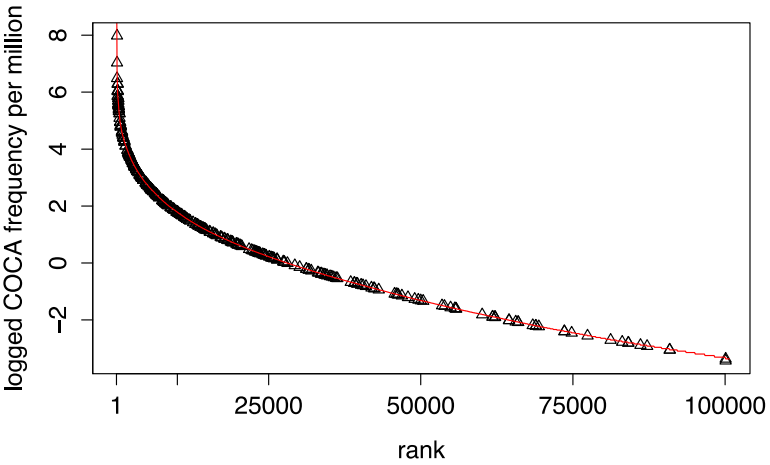

Figure S1: Position of LAGS and LAMSAS items on the rank order curve of the first 100,000 words in COCA, sorted by frequency. The red curve plots the log-transformed frequencies (per million) of use against rank-order position of every 100th word of the first 100,000 most frequent words in COCA (www.wordfrequency.info). The black triangles record the rank-order positions in COCA for the n=368 LAMSAS and LAGS meanings whose highest frequency word in the LAP data had a match in the COCA (http://corpus.byu.edu/coca/). The meanings (Tables S1 and S2) come from a cross-section of the American Corpus, tending towards the more frequently-used end (median rank-order position in the list – first 100,000 words – is 50,000, median rank-order position of n=368 words is 13,000).

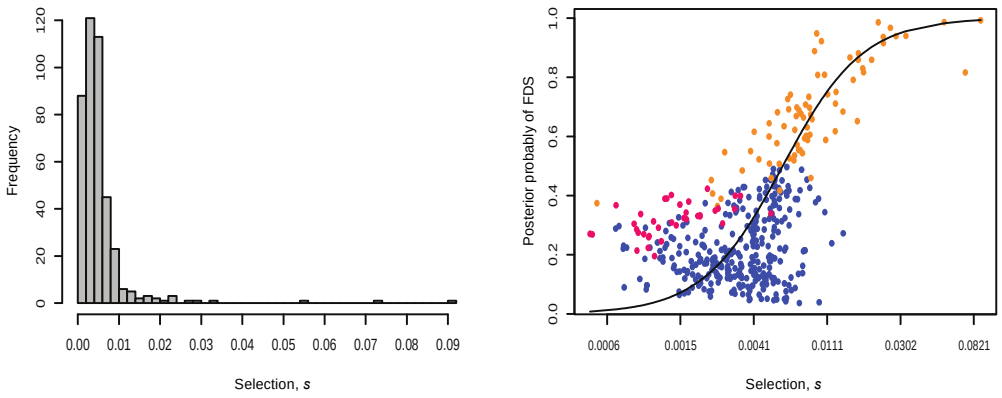

Figure S2: Left panel: frequency distribution of selection coefficients, s: mean±sd = 0.005±0.007 (median = 0.004). Right panel: Posterior probability of FDS model against the size of the selection coefficient, *s*, showing curvilinear relationship (note: x-axis on log-scale). Points are all FDS posterior probabilities but colour-coded to indicate the model that had the highest posterior probability for that meaning: blue = *DS*, magenta = *D*, mustard = *FDS*. The value of *s* for which FDS' posterior probability > 0.5 corresponds to $s \approx 0.006$.

Table S1 LAMSAS Meanings

meaning/concept	most frequent word	2nd most frequent word	number of alternative words	P(D)	P(FDS)	P(DS)	winning model
a little ways	a little ways	a little piece	45	0	0.054	0.946	DS
afternoon	afternoon	evening	9	0.137	0.133	0.729	DS
andiron	andiron(s)	firedog(s)	41	0.119	0.227	0.654	DS
attic	attic	garret	27	0.03	0.089	0.881	DS
backlog	backlog	log	18	0.166	0.229	0.605	DS
biscuits	biscuit	rolls	54	0	0.985	0.015	FDS
boiled eggs	boiled eggs	eggs	13	0	0.543	0.456	FDS
bottom	bottom	bottom land	66	0.084	0.277	0.639	DS
brush	mop	file	11	0.062	0.171	0.767	DS
bureau	bureau	dresser	40	0.014	0.18	0.806	DS
calming down	lay	calm down	117	0.011	0.344	0.645	DS
canal	canal	ditch	17	0.009	0.146	0.845	DS
chimney	chimney	smokestack	7	0.019	0.266	0.714	DS
clabber	clabber	lobbered milk	49	0.043	0.55	0.408	FDS
cleans up	clean up	clean	13	0.014	0.306	0.681	DS
clearing up	clear up	clear off	48	0.019	0.134	0.847	DS
closet	closet	clothespress	18	0.004	0.179	0.816	DS
cloudburst	cloudburst	downpour	135	0.042	0.273	0.685	DS
clouding up	clouding up	threatening	240	0.003	0.816	0.181	FDS
cobbler	cobbler	pie	45	0.638	0.324	0.038	D
cornbread	corn bread	corn pone	55	0.038	0.288	0.674	DS
cottage cheese	cottage cheese	smearcase	47	0.311	0.337	0.352	DS
dragonfly	snake doctor	mosquito hawk	61	0.358	0.339	0.302	D
dry spell	dry spell	drouth	13	0.112	0.096	0.792	DS

food	food	vittle	15	0.098	0.167	0.735	DS
fortnight	fortnight	two weeks	8	0.135	0.221	0.644	DS
from the south	from the south	to the southward	23	0.003	0.322	0.675	DS
frost	frost	freeze	23	0.017	0.169	0.814	DS
froze over	froze over	froze	36	0	0.043	0.957	DS
furniture	furniture	plunder	16	0	0.75	0.25	FDS
garden	garden	patch	7	0	0.859	0.141	FDS
going on	going on	out	7	0.014	0.243	0.743	DS
good day	good day	goodbye	10	0.019	0.132	0.849	DS
gully	gully	gutter	42	0.003	0.406	0.591	DS
half past seven	half past	half after	16	0.013	0.274	0.713	DS
he died with	died with	died of	8	0.125	0.12	0.755	DS
help yourself	help yourself	have some	12	0	0.664	0.336	FDS
hog pen	hog pen	pig pen	55	0.003	0.188	0.809	DS
hominy	hominy	hulled corn	13	0.017	0.225	0.758	DS
i dont care for any	don't care for (any)	no thank you	38	0.004	0.398	0.599	DS
kitchen	kitchen	cook room	11	0.017	0.314	0.67	DS
lightwood	lightwood	lighterd	42	0.378	0.306	0.317	D
make some coffee	make (coffee)	boil (coffee)	23	0.029	0.202	0.769	DS
mantel	mantel	mantelpiece	19	0.228	0.307	0.465	DS
marsh	marsh	meadow	26	0	0.881	0.119	FDS
meadow	meadow	swale	51	0	0.519	0.481	FDS
molasses	molasses	sorghum	17	0	0.791	0.209	FDS
mop	mop	scrub	8	0.095	0.083	0.823	DS
mush	mush	pudding	28	0	0.507	0.492	FDS
my wife	my wife	my lady	24	0.001	0.451	0.548	DS
nice day	pretty day	nice day	57	0.181	0.285	0.534	DS
northeast	northeast	northeaster	7	0.004	0.411	0.585	DS

northwest	northwest	northern	11	0.091	0.186	0.723	DS
over there	over yonder	over there/here	102	0.007	0.162	0.831	DS
pallet	pallet	bunk	21	0.02	0.216	0.763	DS
pancakes	pancake	battercake	58	0.437	0.4	0.163	D
parlor	parlor	living room	20	0.556	0.245	0.199	D
poached	poached (egg)	coddled (egg)	19	0	0.859	0.141	FDS
porch	porch	piazza	23	0.134	0.233	0.633	DS
quarter to eleven	quarter to (eleven)	quarter of (eleven)	16	0.174	0.204	0.621	DS
quilt	quilt	comfort	19	0.128	0.149	0.723	DS
rancid	rancid	strong	37	0.085	0.161	0.754	DS
rind	skin	rind	9	0.075	0.073	0.852	DS
rising	rising	raising	75	0.173	0.616	0.211	FDS
sauce	sauce	dip	8	0.14	0.164	0.697	DS
shades	shades	curtains	12	0.286	0.188	0.525	DS
shut the door	shut	close	17	0.003	0.312	0.685	DS
sit down	sit down	set down	28	0.061	0.128	0.811	DS
snack	snack	lunch	37	0.073	0.121	0.806	DS
soda pop	soft drink	pop	10	0.295	0.23	0.474	DS
sofa	sofa	lounge	25	0.126	0.374	0.5	DS
southeast wind	southeast wind	southeast	16	0.057	0.165	0.778	DS
southwest wind	southwest wind	southwest	15	0.05	0.241	0.709	DS
spelled me	spelled me	spelled (me) off	4	0.022	0.064	0.914	DS
spoiled	spoilt	tainted	20	0.002	0.425	0.573	DS
stairs	stairs	stairsteps	14	0.499	0.214	0.287	D
steady drizzle	steady drizzle	shower	62	0.013	0.356	0.631	DS
sunday before last	sunday before last	sunday week	40	0.109	0.429	0.462	DS
sunday week	Sunday week	Sunday after next	38	0.016	0.113	0.871	DS
sundown	sundown	sunset	12	0.044	0.106	0.85	DS

sunup	sunup	sunrise	12	0.084	0.144	0.772	DS
swamp	swamp	bog	37	0.002	0.421	0.577	DS
syrup	syrup	molasses	9	0.004	0.223	0.773	DS
thunderstorm	thunderstorm	thundershower	39	0.349	0.333	0.318	D
umbrella	umbrella	parasol	10	0.01	0.256	0.734	DS
wardrobe	wardrobe	clothespress	28	0	0.622	0.378	FDS
warmed up	warmed over	warmed up	12	0.117	0.155	0.728	DS
weatherboarding	weatherboarding	clapboard	30	0.002	0.137	0.862	DS
what	what	how's that	61	0.002	0.485	0.512	DS
what time is it	what time is it	what time of day is it	47	0	0.808	0.192	FDS
wheat bread	bread	light bread	22	0.074	0.121	0.805	DS
while	while	spell	4	0.007	0.282	0.711	DS
yolk	yolk	yelk	6	0.391	0.271	0.338	D

Table S2 LAGS Meanings

meaning/concept	most frequent word	2nd most frequent word	number of alternative words	P(D)	P(FDS)	P(DS)	winning model
a little way	little piece (away/off/over)	little ways (away/down/off/out/over)	43	0.263	0.389	0.348	FDS
actress	actress	actor [female]	10	0.007	0.306	0.686	DS
afraid	scared	afraid	9	0.221	0.297	0.482	DS
afternoon	afternoon	evening	7	0.075	0.089	0.836	DS
andirons	andirons	dog horns	44	0.063	0.191	0.746	DS
angry	mad	angry	42	0.003	0.179	0.818	DS
appendicitis	appendicitis	appendix [sic]	34	0	0.808	0.192	FDS
apple core	core	seeds [sic]	14	0	0.602	0.398	FDS
apron	apron	smock	9	0.001	0.617	0.382	FDS
attic	attic	loft	32	0	0.103	0.897	DS
axle	axle	tongue [sic]	5	0	0.992	0.008	FDS
baby carriage	carriage	baby carriage	20	0.292	0.209	0.499	DS
backlog	log	backlog	27	0.058	0.121	0.821	DS
barnyard	lot	barnyard	18	0.063	0.133	0.803	DS
barrow	barrow	bar	23	0.024	0.125	0.85	DS
bawl calf	bleat	bawl	26	0.247	0.257	0.496	DS
bedroom	bedroom	room	12	0	0.711	0.289	FDS
bedspread	bedspread	spread	11	0.384	0.224	0.392	DS
belly buster	belly buster	belly bust	33	0	0.634	0.366	FDS
best man	best man	groom maid	15	0	0.658	0.342	FDS
biscuits	biscuits	biscuit bread	7	0	0.651	0.349	FDS

black coffee	black (coffee)	straight (coffee)	29	0.088	0.213	0.699	DS
blood pudding	blood pudding	blood sausage	19	0.116	0.231	0.653	DS
boil noun	boil	rising	22	0.06	0.149	0.792	DS
bottom	bottomland	bottom	84	0	0.212	0.788	DS
boyfriend	boyfriend	beau	42	0.037	0.27	0.693	DS
bridesmaid	bridesmaid	maid of honor	30	0.032	0.133	0.835	DS
bucket wooden	bucket	wooden bucket	6	0.231	0.224	0.545	DS
bull	bull	male	49	0.002	0.369	0.628	DS
bureau	dresser	chest of drawers	32	0.121	0.19	0.689	DS
butter bean	butter bean	lima	34	0.006	0.065	0.929	DS
calming down	dying down	calming down	110	0.102	0.459	0.439	FDS
canal	ditch	canal	26	0.017	0.168	0.815	DS
candle fly	moth	candle fly	33	0.011	0.115	0.874	DS
cantaloupe	cantaloupe	mushmelon	8	0.158	0.224	0.618	DS
car	car	automobile	23	0.005	0.083	0.912	DS
castrate	castrate	cut	65	0.009	0.383	0.608	DS
cattle	stock(s)	cattle	6	0.303	0.374	0.322	FDS
caught a cold	caught (a/the/that) cold	took (a/this) cold	6	0.407	0.269	0.325	D
cemetery	cemetery	graveyard	38	0.001	0.049	0.95	DS
cherry seed	seed	pit	18	0.027	0.078	0.895	DS
chicken coop	coop	chicken house	68	0.212	0.333	0.455	DS
children	children	kids	19	0.066	0.103	0.832	DS
chilly	chilly	cool	31	0.068	0.223	0.709	DS
chimney	chimney	chimley	6	0.106	0.118	0.776	DS
church music	music	singing	16	0	0.684	0.316	FDS
civil war	Civil War	War between the State(s)	56	0	0.138	0.862	DS
clabber	clabber	buttermilk	11	0.203	0.254	0.543	DS

clean up	clean up	clean	53	0.123	0.296	0.58	DS
clear up	clear up	clear off	75	0.071	0.645	0.284	FDS
cleared the land	cleared	cleaned up	62	0	0.536	0.464	FDS
cliff	cliff	bluff	36	0.001	0.432	0.567	DS
cling peach	cling (peach)	clingstone	27	0.429	0.32	0.251	D
cloth sack	sack	bag	14	0.229	0.299	0.472	DS
clothes basket	basket	clothes basket	25	0.016	0.129	0.855	DS
cloudburst	downpour	rain	124	0.05	0.39	0.56	DS
clouding up	changing	threatening	182	0.045	0.743	0.213	FDS
clumsy	clumsy	awkward	94	0	0.377	0.623	DS
coal bucket	coal bucket	scuttle	22	0.371	0.3	0.329	D
cobbler	cobbler	pie	45	0.007	0.128	0.865	DS
coffin	casket	coffin	9	0.102	0.075	0.823	DS
cork	cork	stopper	12	0.048	0.059	0.894	DS
corn crib	crib	corncrib	20	0.027	0.068	0.905	DS
corn silk	silk	corn silk	16	0.009	0.339	0.652	DS
cottage cheese	cottage [sic]	cheese	24	0.016	0.395	0.589	DS
county seat	county seat	seat	21	0	0.497	0.503	DS
courting	courting	dating	62	0.025	0.453	0.522	DS
cow barn	barn	shed	28	0.402	0.37	0.228	D
cow pen	cow pen	pen	28	0.151	0.375	0.474	DS
crawfish	crawfish	crawdad	16	0.034	0.182	0.784	DS
crawl	crawls	creeps	3	0.002	0.454	0.544	DS
creek tidal	bayou	creek	40	0.36	0.355	0.285	D
crowd whole	bunch	crowd	58	0.098	0.422	0.48	DS
dance	dance	party	47	0	0.741	0.259	FDS
deaf	deaf	hard of hearing	2	0.006	0.178	0.816	DS
devil	devil	Satan	50	0.012	0.129	0.86	DS

died crude	kicked the bucket	croaked	64	0	0.922	0.078	FDS
died neutral	died	passed into the great beyond	53	0.01	0.1	0.889	DS
dish towel	dish towel	drying cloth	25	0.466	0.292	0.243	D
dishrag	dishrag	dishcloth	19	0.057	0.187	0.756	DS
dock	dock	wharf	35	0.11	0.359	0.531	DS
dont remember	don't remember	can't remember	20	0.249	0.311	0.44	DS
double tree	doubletree	double singletree	18	0	0.554	0.445	FDS
dry spell	drought	dry spell	11	0.073	0.116	0.811	DS
education	education	learning	4	0.02	0.166	0.814	DS
employer	boss	mister	36	0.368	0.349	0.283	D
ewe	ewe	doe	20	0	0.669	0.331	FDS
father	father	daddy	14	0.412	0.286	0.303	D
faucet	faucet	spicket	23	0.196	0.283	0.521	DS
feeding time	feeding time	time to feed	29	0.05	0.352	0.598	DS
fell out of bed	fell out of bed	fell off the bed	19	0.081	0.129	0.79	DS
fetch	go get	bring	16	0.379	0.338	0.283	D
field patch	field	patch	25	0.005	0.051	0.944	DS
first grade	first grade	primer	34	0.057	0.419	0.525	DS
flambeau	torch	flambeau	54	0.077	0.299	0.624	DS
food	food	victuals	12	0.058	0.183	0.759	DS
fool	fool	idiot	68	0	0.588	0.412	FDS
fowl	fowl(s)	chicken(s)	12	0.226	0.241	0.532	DS
fox squirrel	fox squirrel	red squirrel	21	0.013	0.131	0.856	DS
freestone peach	freestone (peach)	clear seed (peach)	42	0.004	0.097	0.899	DS
frost	frost	freeze	3	0.146	0.09	0.764	DS
froze over	froze over	froze	16	0.046	0.076	0.879	DS
funeral	funeral	burial	24	0	0.698	0.302	FDS
funnel	funnel	faucet	11	0	0.83	0.17	FDS

furniture	furniture	suite	14	0	0.554	0.446	FDS
gap	gap	pass	17	0.123	0.159	0.718	DS
garden	garden	patch	8	0.001	0.426	0.573	DS
get rid of	get rid of	get shut of	5	0.038	0.121	0.841	DS
ghosts	ghosts	haunts	35	0.04	0.194	0.765	DS
girlfriend	girl friend	sweetheart	54	0	0.215	0.785	DS
gloomy day	cloudy (day/weather)	gloomy (day)	75	0.205	0.6	0.196	FDS
go	go to	get to	6	0.052	0.097	0.851	DS
goal	base	home base	17	0.115	0.264	0.62	DS
good day	good-bye	good day	29	0.078	0.158	0.764	DS
good natured	good-natured	pleasant	102	0.372	0.417	0.211	FDS
good night	good night	good evening	14	0	0.606	0.394	FDS
got married	married	hitched	30	0	0.192	0.808	DS
got sick	got sick	took sick	11	0.099	0.144	0.757	DS
grandfather	grandfather	grandpa	47	0.029	0.151	0.82	DS
grandmother	grandmother	grandma	52	0.003	0.12	0.876	DS
grasshopper	grasshopper	hoppergrass	18	0.003	0.178	0.819	DS
gray squirrel	squirrel	gray squirrel	26	0.014	0.091	0.894	DS
green onions	green onions	shallots	40	0.28	0.364	0.356	FDS
grindstone	grind stone	grind rock	39	0.004	0.119	0.876	DS
grits	grits	hominy	6	0	0.94	0.06	FDS
grove	grove	orchard	26	0.012	0.154	0.834	DS
gully	gully	ditch	41	0.006	0.229	0.765	DS
gutters	gutters	troughs	13	0.008	0.331	0.662	DS
half past seven	(seven, etc.)-thirty	half past (seven, etc.)	16	0.029	0.081	0.89	DS
happy new year	Happy New Year	New Year's Gift	9	0.004	0.358	0.638	DS
harmonica	harmonica	harp	14	0.429	0.223	0.348	D
harness	harness (the horse)	hitch (the horse)	13	0.043	0.104	0.854	DS

harrow	harrow	disc (harrow/harrower)	37	0.012	0.15	0.838	DS
hauling	hauling	carrying	14	0.002	0.451	0.547	DS
haunches	haunches	hunkers	37	0.176	0.253	0.572	DS
haunted house	haunted house	ghost house	17	0.001	0.417	0.581	DS
have a calf	have a calf/cow	drop a baby/calf	34	0.559	0.39	0.052	D
hay shed	hayrack	hay shed	23	0.527	0.402	0.071	D
haycock	windrow	shock	32	0.225	0.407	0.368	FDS
haystack	haystack	stack	27	0.004	0.136	0.86	DS
help yourself	help yourself	serve yourself	22	0.007	0.398	0.595	DS
hill	hill	knob	41	0.006	0.409	0.585	DS
hog pen	hogpen	pen	41	0.258	0.296	0.446	DS
hominy	hominy	homily	3	0.003	0.168	0.829	DS
hooky	play hooky	cut/cutting	54	0.003	0.682	0.315	FDS
hoot owl	hoot owl	owl	58	0.003	0.389	0.608	DS
horseshoes ringtoss	horseshoe(s)	pitching (the/those) horseshoes	34	0.007	0.227	0.766	DS
hull	hull	shell	21	0.008	0.281	0.71	DS
hunker down	squat	squat down	67	0.231	0.522	0.247	FDS
intending to	fixing to	planning to	19	0.469	0.313	0.218	D
into ran	ran into	ran accross	23	0.001	0.367	0.632	DS
Jews harp	Jew's harp	juice harp	15	0.032	0.072	0.896	DS
joggling board	jump board	springboard	18	0.082	0.17	0.748	DS
junk	junk	trash	34	0.007	0.453	0.54	DS
junk room	junk room	storage room	76	0.081	0.508	0.411	FDS
kerosene	kerosene	coal oil	19	0.019	0.067	0.914	DS
kissing	kissing	smooching	52	0.02	0.192	0.788	DS
kitchen	kitchen	stove room	21	0	0.867	0.133	FDS
legal guardian	guardian	foster parent	12	0.008	0.316	0.676	DS

lets out school	is/was/will be out	turn/turns/turning/turned out	25	0.351	0.342	0.307	D
light bulb	bulb	light bulb	8	0.071	0.069	0.86	DS
lightning bug	lightning bug	firefly	20	0.002	0.221	0.777	DS
lightwood	kindling	lighterd	51	0.218	0.547	0.235	FDS
like as not	probably	like as not	16	0.579	0.305	0.116	D
liver pudding	liver pudding	liverwurst	25	0.554	0.39	0.056	D
loft	loft	hayloft	27	0.004	0.124	0.872	DS
low animal sound	moo	low	19	0.069	0.083	0.848	DS
magnolia	magnolia (tree)	cucumber (tree)	8	0.001	0.353	0.646	DS
make coffee	make	boil	12	0.374	0.259	0.367	D
male hog	boar	male hog	27	0.04	0.356	0.604	DS
mantel	mantel	mantelpiece	25	0.032	0.136	0.831	DS
maple	maple	maple tree	11	0.048	0.162	0.79	DS
marsh	marsh	marshland	45	0	0.215	0.785	DS
meadow	meadow	pasture	59	0.044	0.43	0.527	DS
merry christmas	Merry Christmas	Christmas Gift	31	0.003	0.121	0.876	DS
merry go round	merry-go-round	flying jenny	53	0.003	0.039	0.958	DS
midwife	midwife	granny	32	0	0.135	0.865	DS
minnow	minnow	shiner	43	0	0.674	0.326	FDS
molasses	molasses	sorghum	14	0.023	0.211	0.767	DS
mongrel	feist (dog)	cur (dog)	55	0.503	0.423	0.074	D
mother	mother	mamma	14	0.171	0.189	0.64	DS
mountain laurel	mountain laurel	laurel	19	0.015	0.078	0.907	DS
mourning	mourning	carrying on	38	0	0.888	0.112	FDS
mush	mush	cush	19	0.009	0.348	0.643	DS
mushroom	mushroom	toadstool	11	0	0.817	0.183	FDS

nice day	clear (day/weather)	beautiful (day/morning/weather)	41	0.54	0.38	0.081	D
nickname	nickname	pet name	13	0.005	0.114	0.881	DS
no kin	no kin	no relation(s)	18	0.13	0.177	0.693	DS
obligated dont want to be	obligated	beholden (to no one/nobody)	27	0.041	0.384	0.575	DS
orphan	orphan	motherless/fatherless	9	0	0.588	0.411	FDS
over there	over there/here	over yonder	27	0.392	0.309	0.299	D
pancakes	pancakes	Flapjacks	42	0.146	0.366	0.488	DS
pantry	pantry	Storeroom	34	0	0.726	0.274	FDS
paper bag	paper bag	Bag	17	0.498	0.263	0.24	D
pasture	pasture	Range	19	0	0.689	0.311	FDS
peach seed	seed	Kernel	19	0.283	0.327	0.391	DS
peaked	peaked	Pale	86	0.081	0.388	0.531	DS
pick flowers	pick	Cut	12	0.261	0.212	0.527	DS
picket fence	picket fence	paling fence	44	0.113	0.485	0.403	FDS
poached egg	poached (egg)	coddled (egg)	15	0	0.936	0.064	FDS
poison ivy	poison ivy	poison oak	44	0.004	0.076	0.919	DS
potatoes	potatoes	Irish (potatoes)	26	0.026	0.096	0.878	DS
pregnant	pregnant	Expecting	59	0.021	0.321	0.658	DS
pretty well	fine	pretty good	99	0.024	0.337	0.638	DS
proud flesh	proud flesh	plowed flesh	17	0.003	0.46	0.538	DS
public square	park	Square	44	0.368	0.356	0.276	D
pulley bone	pulley bone	V bone	22	0.016	0.05	0.934	DS
purse	purse	pocketbook	30	0.036	0.174	0.79	DS
pus	pus	Corruption	21	0.021	0.154	0.825	DS
push the baby	push	Stroll	17	0.565	0.269	0.165	D
quarter till	quarter till	quarter to	29	0.114	0.184	0.701	DS

quilt	quilt	Comfort	13	0.03	0.204	0.765	DS
rail fence	rail fence	Rail	73	0	0.677	0.323	FDS
railroad station	depot	railroad station	30	0.12	0.217	0.663	DS
ram	ram	Buck	23	0.002	0.197	0.801	DS
rancid	rancid	Rank	43	0.396	0.452	0.151	FDS
rather	kind of	Pretty	22	0.625	0.195	0.179	D
ravine	gully	Ravine	53	0.323	0.331	0.345	DS
reared	raise/raised	rear/reared	13	0.037	0.211	0.752	DS
relatives	relatives	kinfolks	16	0.51	0.275	0.216	D
remember	remember	member	6	0.061	0.171	0.768	DS
resembles his father	favors	looks like	81	0.395	0.459	0.146	FDS
rheumatism	arthritis	rheumatism	29	0.007	0.051	0.942	DS
rhododendron	rhododendron	azalea	13	0.013	0.229	0.757	DS
rim	rim	tire	19	0.061	0.13	0.808	DS
rind	skin	rind	17	0.026	0.065	0.909	DS
rising	picking up	rising	57	0.436	0.399	0.165	D
roasting ears	roasting ears	corn	32	0.045	0.121	0.834	DS
rock	rock	stone	4	0.033	0.141	0.826	DS
rock fence	rock fence	rock wall	12	0.378	0.367	0.255	D
sample	sample	scrap	20	0.006	0.391	0.603	DS
sauce	sauce	syrup	16	0.033	0.187	0.78	DS
sawhorse	sawhorse	horse	96	0.005	0.215	0.78	DS
scrapple	scrapple	hash	9	0.009	0.386	0.605	DS
screech owl	screech owl	owl	50	0.004	0.08	0.916	DS
second cutting	volunteer	second cutting	69	0.095	0.324	0.58	DS
seesaw	seesaw	teeter-totter	17	0	0.526	0.474	FDS
setting hen	setting hen	setting	36	0.034	0.428	0.538	DS
shades	shades	window shades	16	0.049	0.182	0.77	DS

shafts	shafts	trace	14	0	0.967	0.033	FDS
sheaf	bundle	sheaf	21	0.007	0.248	0.745	DS
shell	shell	hull	6	0.011	0.192	0.798	DS
shell verb	shell	hull	19	0	0.418	0.582	DS
shin	shin(s)	shinbone	15	0.019	0.13	0.851	DS
shopping	shopping	trading	8	0.011	0.164	0.825	DS
shuck	shuck (n.)	husk (n.)	19	0.01	0.251	0.739	DS
shut the door	shut	close	16	0.021	0.047	0.932	DS
sic him	sic (him/it/them)	get (him/it/them)	29	0.011	0.145	0.844	DS
sidewalk	sidewalk	banquette	14	0.001	0.431	0.568	DS
single tree	singletree	swingletree	16	0.001	0.488	0.511	DS
sit down	sit down	have a seat	21	0.111	0.226	0.663	DS
skillet	skillet	frying pan	35	0.038	0.147	0.815	DS
skunk	skunk	polecat	29	0.004	0.05	0.946	DS
sleepy	sleepy	drowsy	23	0.001	0.156	0.842	DS
slop bucket	slop bucket	bucket	47	0.001	0.432	0.566	DS
small predator	varmint	rodent	13	0.003	0.49	0.507	DS
snack	snack	lunch	37	0	0.697	0.303	FDS
snits	dried apples	dried fruit	27	0.001	0.122	0.876	DS
sofa	sofa	couch	23	0.276	0.28	0.443	DS
somersault	somersault	somerset	25	0.073	0.164	0.763	DS
souse	souse	hoghead cheese	27	0.303	0.243	0.455	DS
spanking	whipping (n)	spanking (n)	54	0.003	0.088	0.91	DS
spider web	spider web	cobweb	10	0.252	0.177	0.571	DS
spoiled meat	spoilt	rancid	33	0.068	0.412	0.52	DS
stairs	steps	stairs	11	0.406	0.186	0.408	DS
stallion	stallion	stud	27	0.05	0.119	0.831	DS
start when does school	start	begin	12	0.165	0.191	0.644	DS

steady drizzle	drizzle (n.)	sprinkle (n.)	62	0.125	0.313	0.562	DS
stole	stole	took	34	0.175	0.334	0.492	DS
stovepipe or flue	stovepipe	flue	20	0.097	0.137	0.767	DS
stranger	stranger	foreigner	29	0.004	0.175	0.821	DS
streetcar	streetcar	trolley	13	0.103	0.2	0.698	DS
string of beads	necklace	beads	20	0.238	0.211	0.551	DS
strong	strong	stout	55	0.002	0.109	0.888	DS
stubborn	stubborn	hardheaded	78	0.027	0.411	0.562	DS
student	student	pupil	11	0.138	0.139	0.723	DS
sumac	sumac	poison sumac	24	0	0.63	0.37	FDS
sunday before last	Sunday before last	Sunday a week ago	36	0.102	0.344	0.554	DS
sunday next week	Sunday week	Sunday after next	41	0.008	0.262	0.73	DS
sundown	sundown	sunset	10	0.172	0.259	0.57	DS
sunup	sunup	sunrise	14	0.348	0.291	0.361	DS
swamp	swamp	swampland	72	0	0.708	0.292	FDS
sweat preterite	sweated	perspired	4	0.112	0.096	0.791	DS
sweet potatoes	sweet potatoes	yams	48	0	0.036	0.964	DS
sycamore	sycamore	buttonwood	9	0	0.986	0.014	FDS
syrup	syrup	long sweetening	6	0	0.915	0.085	FDS
tattletale	tattletale	gossip	58	0	0.388	0.612	DS
team of mules	team (of mules)	pair (of mules)	29	0.005	0.065	0.929	DS
terrapin	terrapin	gopher turtle	41	0.034	0.288	0.678	DS
theater	theater	picture show	22	0.04	0.27	0.69	DS
threw	threw	chunked	17	0.014	0.159	0.827	DS
thunderstorm	thunderstorm	electrical storm	28	0.113	0.149	0.738	DS
tightwad	tightwad	miser	53	0.006	0.123	0.871	DS
tired	tired	exhausted	60	0	0.692	0.308	FDS
toadstool	toadstool	frogstool	22	0.002	0.293	0.705	DS

tongue	tongue	tree	13	0.016	0.198	0.786	DS
tote	tote	carry	9	0.348	0.289	0.364	DS
touchy	touchy	high-tempered	89	0.024	0.577	0.399	FDS
towel	towel	drying rag	5	0	0.939	0.061	FDS
trousers	pants	overalls	35	0.459	0.331	0.21	D
turned him down	turned him down	refused him	134	0.02	0.239	0.741	DS
valley	valley	gutter [sic]	20	0	0.733	0.267	FDS
vase	vase	flowerpot	37	0.002	0.105	0.893	DS
vest	vest	jacket	6	0.013	0.239	0.748	DS
vomit crude	puke	throw up	61	0.005	0.4	0.595	DS
vomit neutral	vomit	throw up	44	0.003	0.1	0.897	DS
wardrobe	wardrobe	chifforobe	33	0.012	0.093	0.895	DS
warmed up	leftovers	warmed-overs	16	0	0.572	0.428	FDS
washcloth	washrag	washcloth	16	0.174	0.235	0.591	DS
washpot	washpot	kettle	30	0.239	0.195	0.566	DS
water in blister	water	fluid	13	0.016	0.305	0.679	DS
waterfall	waterfall	fall	14	0.007	0.096	0.897	DS
weather boarding	weatherboarding	siding(s)	30	0.041	0.098	0.861	DS
what time is it	what time is it?	what time do you have?	60	0	0.948	0.052	FDS
wheat bread	light bread	bread	35	0.372	0.33	0.298	D
wheelbarrow	wheelbarrow	wheelbar	17	0.012	0.146	0.842	DS
whetrock	whetrock	whetstone	42	0.018	0.152	0.83	DS
while	while	time	8	0.08	0.184	0.736	DS
whinny	nicker	bray	38	0.008	0.472	0.52	DS
whip	whip	switch	15	0	0.594	0.406	FDS
wild hog	wild hog	wild boar	49	0.024	0.29	0.686	DS
wire fence	barbwire fence	wire fence	34	0.002	0.046	0.951	DS
woman teacher	teacher	schoolteacher	29	0.054	0.126	0.82	DS

yellow jaundice	yellow jaundice	jaundice	11	0.129	0.142	0.729	DS
yolk	yolk	yellow (=yolk)	10	0.056	0.122	0.822	DS

245

Table S3. Top ten words by posterior probability for each model

Drift	Frequency-dependent selection	Directional selection
0237_LAGS_relatives (noun)	0305_LAGS_syrup (noun)	0223_LAGS_pulley_bone (noun)
0146_LAGS_hay_shed (noun)	0083_LAGS_died_crude (verb)	0240_LAGS_rheumatism (noun)
0199_LAGS_nice_day (expression)	0214_LAGS_poached_egg (noun)	0100_LAGS_field_patch (noun)
0174_LAGS_liver_pudding (noun)	0318_LAGS_towel (noun)	0272_LAGS_skunk (noun)
1062_LAMSAS_parlor (noun)	0133_LAGS_grits (noun)	1001_LAMSAS_a_little_ways (exp)
0145_LAGS_have_a_calf (verb)	0334_LAGS_what_time_is_it (exp)	0045_LAGS_cemetery (noun)
0226_LAGS_push_the_baby (verb)	0258_LAGS_shafts (noun)	0342_LAGS_wire_fence (noun)
0173_LAGS_like_as_not (expression)	1008_LAMSAS_biscuits (noun)	1033_LAMSAS_froze_over (verb)
0234_LAGS_rather (expression)	0304_LAGS_sycamore (noun)	0186_LAGS_merry_go_round (noun)
1023_LAMSAS_cobbler (noun)	0013_LAGS_axle (noun)	0303_LAGS_sweet_potatoes (noun)

Note: across all meanings, parts of speech do not differ significantly among the model 'winners' (the model with the highest posterior probability of a given meaning): χ^2=7.8, *df*=8, *p*>0.45

Table S4 Delete and merge criteria

Reason for Filtering	Type	Example	no. of occurrences	no. of meanings affected
merging plurals	MERGE	waterfalls -> waterfall (0332_LAGS_waterfall)	392	76
merging different verb tenses	MERGE	shelling, shelled -> shell (0261_LAGS_shell_verb)	274	78
merging different pronouns	MERGE	sure of yourself -> sure of himself (0294_LAGS_stubborn)	16	4
merging typos, incorrect or different spelling to correct form	MERGE	night miler -> night miller (0037_LAGS_candle_fly)	191	71
removing irrelevant context around concept	MERGE	Bib apron, red apron -> apron (0011_LAGS_apron)	2286	363
removing articles (a, the)	MERGE	the daddy -> daddy (0032_LAGS_bull)	92	25
remove inappropriate word classes	DELETE	greedy [adj.] (0312_LAGS_tightwad)	193	19
removing inappropriate or incorrect response that doesn't fit concept	DELETE	delinquents (0048_LAGS_children)	757	249
merging subtype or subcategory of main concept	DELETE	heater flue -> flue (0289_LAGS_stovepipe_or_flue)	24	6
removing inappropriate non-English responses	DELETE	meulon (F) (0146_LAGS_hay_shed)	8	8
merging uneccessary brackets or implied context	MERGE	rain shower (light) -> light rain shower (0287_LAGS_steady_drizzle)	143	51

Note: The total number of filtering operations was 4376 or roughly 1% of the approximately 417,000 responses given for the 418 meanings.

References

1. Kretzschmar WA (1993 (2007)) *Handbook of the linguistic atlas of the Middle and South Atlantic States* (University of Chicago Press).
2. Davies M (2009) The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics* 14(2):159-190.
3. Brysbaert M, Warriner AB, & Kuperman V (2014) Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods* 46(3):904-911.
4. Coltheart M (1981) The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology* 33(4):497-505.
5. Kuperman V, Stadthagen-Gonzalez H, & Brysbaert M (2012) Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods* 44(4):978-990.
6. Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution and Systematics* 41(379-406):1.
7. Beaumont MA & Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics* 5(4):251-261.
8. Ohta T (1977) Extension to the neutral mutation random drift hypothesis. *Molecular evolution and polymorphism. National Institute of Genetics, Mishima, Japan*:148-167.
9. Beaumont MA, Cornuet J-M, Marin J-M, & Robert CP (2009) Adaptive approximate Bayesian computation. *Biometrika*:asp052.