Université de Montréal

Prediction of Alzheimer's disease and semantic dementia from scene description: toward better language and topic generalization

par

Ilya Ivensky

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Discipline

September 12, 2020

Université de Montréal

Faculté des études supérieures et postdoctorales

Ce mémoire intitulé

Prediction of Alzheimer's disease and semantic dementia from scene description: toward better language and topic generalization

présenté par

Ilya Ivensky

a été évalué par un jury composé des personnes suivantes :

Mitliagkas Ioannis
(président-rapporteur)

Philippe Langlais
(directeur de recherche)

Pierre Bellec
(membre du jury)

Résumé

La segmentation des données par la langue et le thème des tests psycholinguistiques devient de plus en plus un obstacle important à la généralisation des modèles de prédiction. Cela limite notre capacité à comprendre le cœur du dysfonctionnement linguistique et cognitif, car les modèles sont surajustés pour les détails d'une langue ou d'un sujet particulier.

Dans ce travail, nous étudions les approches potentielles pour surmonter ces limitations. Nous discutons des propriétés de divers modèles de plonjement de mots FastText pour l'anglais et le français et proposons un ensemble des caractéristiques, dérivées de ces propriétés. Nous montrons que malgré les différences dans les langues et les algorithmes de plonjement, un ensemble universel de caractéristiques de vecteurs de mots indépendantes de la langage est capable de capturer le dysfonctionnement cognitif. Nous soutenons que dans le contexte de données rares, les caractéristiques de vecteur de mots fabriquées à la main sont une alternative raisonnable pour l'apprentissage des caractéristiques, ce qui nous permet de généraliser sur les limites de la langue et du sujet.

Mots clés Déficience cognitive, Caractéristiques multilingues, Plongement de mots, Linguistique informatique, Traitement automatique des langues

Abstract

Data segmentation by the language and the topic of psycholinguistic tests increasingly becomes a significant obstacle for generalization of predicting models. It limits our ability to understand the core of linguistic and cognitive dysfunction because the models overfit the details of a particular language or topic.

In this work, we study potential approaches to overcome such limitations. We discuss the properties of various FastText word embedding models for English and French and propose a set of features derived from these properties. We show that despite the differences in the languages and the embedding algorithms, a universal language-agnostic set of word-vector features can capture cognitive dysfunction. We argue that in the context of scarce data, the hand-crafted word-vector features is a reasonable alternative for feature learning, which allows us to generalize over the language and topic boundaries.

Keywords Cognitive impairment, Multilingual features, Word embedding, Computational linguistics, Natural language processing

Contents

List of tables

List of figures

Acronyms

AD: Alzheimer's Disease. 21–24, 26–29, 31, 33, 36, 37, 39, 45, 49, 55–57
ANN: Artificial Neural Network. 23, 48, 53
AUC-ROC: Area Under The Curve Receiver Operating Characteristics. 49, 50

CBOW: Continuous Bag of Words. 11, 23, 31, 37–40, 42–44, 54–59
CFG: Context-Free Grammar. 54, 61
CNN: Convolutional Neural Network. 11, 23, 58, 59
CS: Connected Speech. 21, 23, 26
CSU: Clause-like Semantic Unit rate. 23

DEPID: Dependency-based Propositional Idea Density. 34
DL: Deep Learning. 21, 22

FN: false negative. 50
FNN: Feed-Forward Neural Network. 48, 53, 54
FP: false positive. 50, 51

HC: Healthy Control group. 22, 24, 26–28, 36, 41, 44, 49, 55–57

ICU: Information Content Unit. 22, 24, 28, 34
ID: Idea Density. 13, 24, 34, 36–38, 55, 56

KDE: Kernel Density Estimation. 13, 36, 38, 41–43

MATTR: Moving-Average Type-Token Ratio. 31, 32, 47
MCI: Mild Cognitive Impairment. 24
ML: Machine Learning. 21, 22
MLP: Multilayer Perceptron. 48
MMSE: Mini-Mental State Examination. 27, 49

NB: Naive Bayes. 48

NLP: Natural Language Processing. 21, 23, 29, 30, 33, 42, 57

OOV: out-of-vocabulary. 30

PID: Propositional Idea Density. 24, 34
POS: part-of-speech. 23, 29–32
PTB: Penn Treebank. 32

ROC: Receiver Operating Characteristics. 49, 50

SCNN: Self Normalizing Convolutional Neural Network. 57, 58
SD: Semantic Dementia. 36, 49, 55–57
SG: skip-gram. 11, 13, 23, 30, 31, 36–39, 41–43, 55, 56, 58, 59
SID: Semantic Idea Density. 24, 34
SNN: Self-Normalizing Neural Network. 54–57
SVM: Support-Vector Machine. 23, 48
svPPA: semantic variant of Primary Progressive Aphasia. 26, 27, 36, 41, 44

TN: true negative. 50
TP: true positive. 50, 51
TTR: Type-Token Ratio. 23, 31, 32, 34, 47

UD: Universal Dependencies. 32

Glossary

Brunét's Index: a measure of the lexical richness. It is calculated as $N^{V^{-0.165}}$, where N is the length of the text (after removing punctuation), V is the number of unique lemmas. Lower values correspond to richer texts.. 23, 31, 47

Clause-like Semantic Unit rate: quantifies the subject's ability to form phrases. A good low score implies that the subject can form longer phrases. 15

Honoré's statistic: a measure of the lexical richness. It is calculated as $100 * \log_2 N/(1 - V_1/V)$, where N is the length of the text (after removing punctuation), V is the number of unique lemmas and V_1 is the number of lemmas used only once in the text. Note that this statistic is not defined when all lemmas are unique (i.e. when $V_1 = V$) due to division by zero. That happens for the short texts like *Je ne sais pas. Non.* 23, 31, 47

TTR: Type-Token Ratio (TTR) is vocabulary size divided by the length of the text. 16

Dedication

A thesis to the memory of my late father, Prof. Gregory Ivensky, who inspired me for the academic research, and to my family, for their unconditional support.

Chapter 1

Introduction

Alzheimer's Disease (AD) is an irreversible and progressive brain disease that destroys brain cells and results in memory loss and cognitive impairment. Mostly, it affects older adults in their 60-70s, but in rare cases, people get affected in their 30-40s. According to the Alzheimer Society of Canada, over half a million Canadians were living with AD in 2019. In the next 15 years, this number is expected to be close to one million[**1**]. The annual cost of care in 2019 exceeded 10 billion dollars.

Despite the technological progress in medical imaging and biochemistry, which allowed AD diagnostics in the very early stages, neuropsychological examinations remain the primary tool to evaluate the disease because it is the only tool that can assess mental decline.

Linguistic analysis of Connected Speech (CS) is an integral part of these examinations for the last several decades. It was noticed that subtle changes in the language could occur before clinical symptoms become apparent. Modern Natural Language Processing (NLP) helps to detect dementia earlier than any other type of neuropsychological evaluation. It may differentiate types of dementia and measure the progress of the disease.

Recent resurgence of Deep Learning (DL) sparked revolution in NLP. Deep algorithms outperform classical ones by a significant margin in any domain of application. There are successful applications of DL in psycholinguistic as well, including the prediction of AD. However, most of those applications are concentrated on a single dataset Cookie Theft task[**2**] from DementiaBank corpus in English[**3**], and their objective is binary discrimination. They remain to be a proof of the concept rather than a multilingual diagnostic tool, which is able to differentiate a range of conditions.

In general, the training of deep algorithms requires much data. However, medical data, by its nature, is sparse. In psycholinguistic, contrarily to other medical specialties, the data is fragmented by the language, which further complicates the application of DL methods.[1] Still, classical Machine Learning (ML) algorithms running on the hand-crafted features extracted from the text are essential for psycholinguistics. They better generalize on sparse data

[1]Language agnostic of deep algorithms applies to their code, but not to the learned model.

and are more straightforward to tune and train. They are better positioned for finding diagnostically useful features than DL algorithms because their features are engineered by humans and, therefore, more interpretable for humans.

A major limitation of classical ML algorithms is the dependency of their features on the underlying language. Since deep algorithms learn features from a raw input or its embedded representation, they automatically adapt to the underlying language during the training. Contrarily to that, hand-crafted features of classical ML algorithms should be manually adjusted for new languages. In many cases, that requires research, which is impossible without sufficient medical data in the targeted language and local economic and human resources.

Topic-constrained tasks like picture description highlight another limitation of classical ML algorithms - dependency of their features on a specific task. Participants of the test are expected to mention a list of topics, a.k.a. Information Content Units (ICUs). The algorithm should be aware of those ICUs, regardless of how they are retrieved from the picture - manually or learned by some external algorithm. Such a limitation - a predefined set of static features - leads to additional segmentation of available data. Topic-constrained tasks become non-compatible with each other if the classifier cannot have the same features across tasks.

We argue that both problems caused by static features could be mitigated using properties of the word-vector space and careful selection of lexical and syntactical features.

Some lexical and syntactic features, especially production rules, poorly translate to other languages. For instance, one production rule can have an excellent prediction power in one language but be not applicable in the other. Over-relying on such features is a sort of overfitting of the algorithm for one language. We believe that the core language dysfunction does not depend on the language. That is an additional motivation for reducing language dependency on the features.

An analysis of word-vector space properties allows us to unlock some insights into the data, which otherwise are accessible only for DL techniques. We try to use those insights when the size of the data has no capacity for DL, for instance in Picnic dataset (see 1.2.2).

We suggest that if word-vector space were pre-trained on a sufficiently diverse content (e.g., Wikipedia or Common Crawl), then mapping text to that space would reveal distinctive patterns for Healthy Control group (HC) and AD participants. While features extracted from these mappings depend on the word-vector model, they are language-agnostic. Word-vector models represent the contextual meanings of words. A sequence of words in the text corresponds to a path in the word-vector space. Properties of that path could hint on the richness of the text content (though these properties cannot indicate whether the content of the text makes any sense) and provide a workaround for a fixed list of ICUs.

In the current work, we study properties of word-vector space of Continuous Bag of Words (CBOW) and Skip-gram (SG) models of two languages - English and French - and propose new features based on those properties for predicting AD. This work is a proof of the concept rather than a complete practical solution. More research on a more substantial amount of data and other language families is needed.

The central part of this work - chapters 2 and 3 - describe details of the suggested features and classification methods. In chapter 4, we describe our experiments and compare their results with some deep models (various implementations of Convolutional Neural Networks (CNNs)) and state-of-the-art achieved by other feature-based models on the same data.

1.1. Related work

Prediction of dementia from the CS is essentially a text classification problem - one of the most indispensable problems in NLP. Each year brings a continuously increasing number of new papers, which causes rapid progress in all applications of text classification - from Information Retrieval to Sentiment Analysis to Knowledge Management and Recommender Systems. Prediction of dementia is not an exception in that process.

Roots of performing text analysis for psychological evaluation go back to the very beginning of the previous century. In the 1901 Freud wrote in his book "Psychopathology of Everyday Life"[**4**] that tongue slips, also known as parapraxis, can reveal unconscious thoughts.

The first computerized text analysis in psychology was described by Stone et al. in 1966[**5**], and in 1997 S. Singh published a paper describing the quantitative classification of conversational analysis with Artificial Neural Network (ANN)[**6**] - probably, one of the first attempts of quantification of the degree of linguistic deficit based on the lexical features extracted from the CS transcripts. In that work he used 8 lexical features such Part-of-speech (POS) tag frequencies for nouns, pronouns, adjectives and verbs, as well Type-Token Ratio (TTR), Clause-like Semantic Unit (CSU) rate, Brunét's Index and Honoré's statistic for classifying dysplasia patients.

Bucks et al.[**7**] applied the same lexical features over the first 1000 words to conduct a linear discriminant analysis of spontaneous speech for AD patients.

The next era of text analysis started with the appearance of standardized NLP tools. They allowed fast and reliable feature extraction, and greatly facilitated research in applied NLP.

Orimaye et al.[**8**] used Stanford Parser[**9**] to extract syntactic features. Their Support-Vector Machine (SVM) model with a carefully selected and very limited set of lexical and syntactic features achieved an $F1$ score of 0.74 in predicting AD and related Dementias group on DementiaBank corpus (Cookie Theft task[**2**]). However, that study did not focus specifically on AD. Its dementias group included 189 probable AD, 8 possible AD, 37

Mild Cognitive Impairment (MCI), 3 memory problems, 4 Vascular dementia, and 1 other participant with an unidentified form of dementia.

Fraser et al.[**10**] enhanced the model with acoustic features and also added binary features of information content. They used a manual set of expected ICUs enhanced with synonyms to match for relevant lexical items. Their model reached an $F1$ score of 0.82 on the same dataset. Their AD group included only “possible AD” or “probable AD” participants.

Yancheva et al.[**11**] used pre-trained word-vector space models to cluster nouns and verbs of HC group and learn ICUs rather than using a human-supplied list. Representation of words and ICUs in continuous space allowed fuzzy matching of words to topics. They calculated Idea Density (ID) as a number of expected topics, which were mentioned in the transcript, divided by the total number of words. Similarly, they defined Idea Efficiency as the number of the mentioned expected topics divided by the total duration of the recording (in seconds). Using these features together with the cluster matching distance features and set of linguistic, syntactic and acoustic features from Fraser et al.[**10**], their model achieved an $F1$ score of 0.80 on DementiaBank corpus (Cookie Theft task[**2**]). In section 2.4 we will analyze details of their variant of ID and we will compare it with other interpretations.

Sirts et al.[**12**] coined names Propositional Idea Density (PID) and Semantic Idea Density (SID), though the concepts behind those names existed earlier. They experimented with both versions of ID on the topic-constrained corpus (DementiaBank) and topic-free autobiographical memory interviews (AMI). They showed that the PID in combination with SID and cluster distance features (two latter were proposed by Yancheva et al.[**11**]) yields better results on the topic-constrained corpus, and SID combined with cluster distance features without PID yields better results on the topic-free corpus.

Rentoumi et al.[**13**] suggested that some of the language features could be shared between languages on samples from Cookie Theft task[**14**] in English and Greek. Their preliminary findings were promising, however they did not share classification results.

Fraser et al.[**15**] generalized the work of Yancheva et al.[**11**] for multilingual topic modeling. Using pre-trained FastText word embeddings[**16**] of English and Swedish and applying transformation matrices to align two different word spaces ([**17**]), they showed that adding data from the different language is more effective for training topic clustering model than adding data from the same language. They argued that when the picture is relatively simple, an increase of sampling in a single language does not provide new information but concentrates topic clusters around highly frequent words and reduces the fuzzy matching of topics to keyword lookup. That finding supports our claim that classifiers trained on a single language are prone to overfitting for that language.

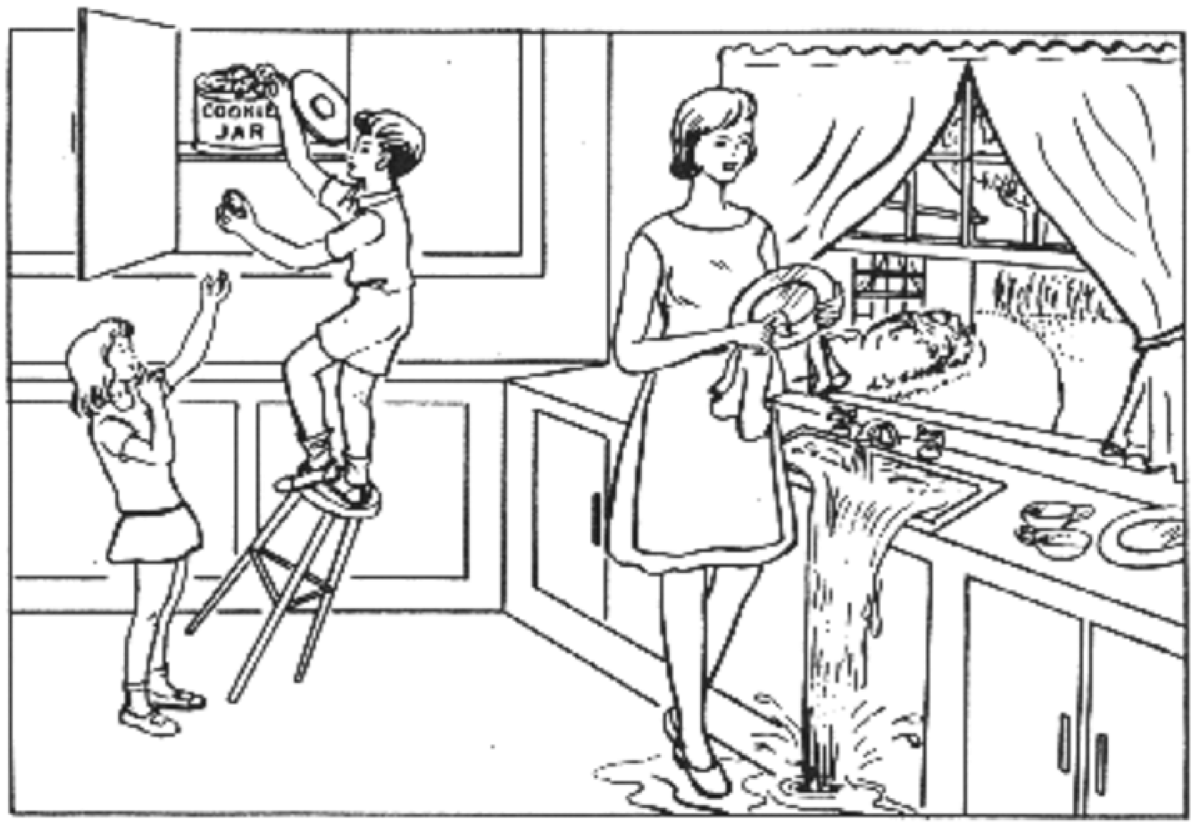


Fig. 1.1. Cookie Theft scene

1.2. Data

This research focuses on three datasets - two in French and one in English. Both French datasets were collected by the Center de Recherche of Institut Universitaire de Gériatrie de Montréal (CRIUGM) with the approval of the Ethics-Neuroimaging Research Ethics Committee of the CIUSSS of Center-Sud-de-l'ile-de-Montréal, approval number $CERVN17-18-12$. The third dataset is the Pitt (a.k.a. "Cookie Theft") corpus of DementiaBank [**3**] (in English). All datasets are manual transcripts of a picture description task to elicit speech. Each dataset differs from the others in terms of the number of participants, age and education balance between dementia and control groups, and the severity of the mental impairment in dementia groups. Thus, each dataset presents its unique challenges.

1.2.1. DementiaBank corpus ("Cookie Theft")

This corpus is a part of DementiaBank. It was gathered as a part of a protocol administrated by the Alzheimer and Related Dementias Study at the University of Pittsburgh

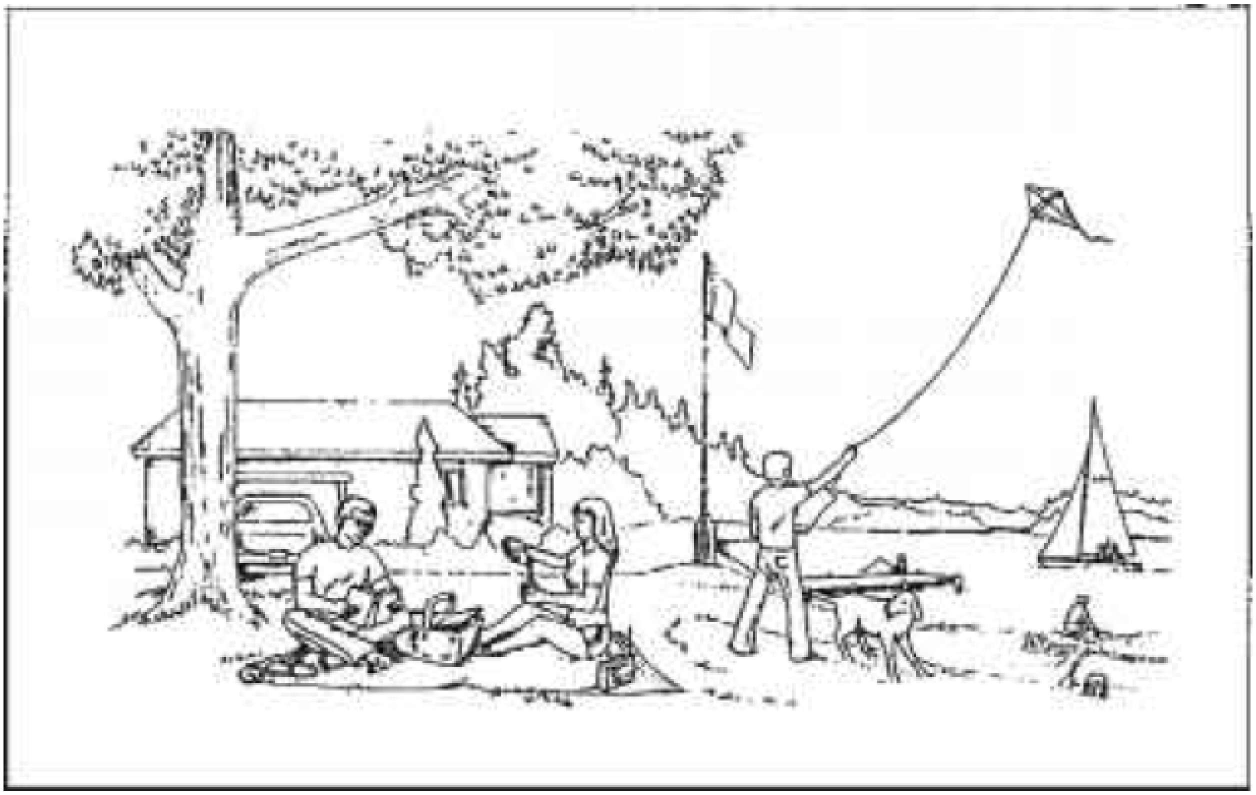


Fig. 1.2. Picnic scene

School of Medicine and described by Becker et al.[**3**]. It is probably the most popular and well-studied corpus in that domain.

The corpus comprises 536 transcripts of 104 controls, 208 dementia (different types, including AD) and 85 with an unknown diagnosis. All transcripts are descriptions of the Cookie Theft scene (Fig 1.1) from the Boston Diagnostic Aphasia Examination[**2**], which are manually transcribed according to the CHAT protocol[**18**].

1.2.2. Picnic corpus

This small corpus in French comprises 33 short, manually created transcripts of CS elicited through a description of a Picnic scene (Fig 1.2) from Western Aphasia Battery[**19**]. All participants have been matched by age and education. The group includes 12 patients with AD, 9 patients with a semantic variant of Primary Progressive Aphasia (svPPA), and the remaining 12 are cognitively intact older adults - a HC group. The protocol of transcripts is described in detail by Montembeaul et al.[**20**].

Contrarily to other corpora, which consider only two classes, this one distinguishes participants between three categories: one control group and two impaired. The challenge here is to separate the AD group correctly, not only from controls but also from svPPA. Both impaired groups equally present deficits in naming unique entities, but AD participants have significantly better performance in calling non-unique ones. For example:

svPPA: *Il y a un chien il y a trois personnes dont une petite fille il y a un autre gars ici il y a une auto ici ça a l'air de une maison ça Lui est assis elle est assise aussi là elle vidait quelque chose de la eau ou quelque chose à boire C'est pour mettre ses pieds dessus Lui là il a une affaire pour lire il n'y a rien à lire là-dessus Il ouvre ça il regarde il lit. Il y a quelque chose dans la eau.*

AD: *une maison de la campagne avec un arbre un jeune homme qui court avec un cerf-volant un bateau sur l' eau avec deux personnes à l' intérieur un couple faire un pique-nique avec une bonne bouteille de vin un chien une petite mademoiselle qui fait un château avec une pelle une petite chaudière un pêcheur sur un quai je vois l' auto en avant de la maison le drapeau à côté de la maison le poisson qui est attrapé par le pêcheur. la madame se sert un bon verre de vin la madame écoute la musique. les sandales du monsieur les oiseaux dans le nuage*

HC: *il y a un arbre devant la maison une voiture près du garage un drapeau à droite de la maison un enfant joue avec le cerf-volant il court près de la plage le chien suit l' enfant un homme est à la pêche sur le quai il vient d' attraper un poisson la fillette fait des châteaux de sable avec une pelle et une chaudière on voit un voilier à l' horizon il porte le numéro 470 un couple fait un pique-nique en écoutant de la musique l' homme fait de la lecture la femme prend un breuvage.*

That suggests that to separate AD and svPPA groups, the classifier has to distinguish between unique and non-unique entities.

1.2.3. MACS corpus

This corpus comprises the Montreal Assessment of Connected Speech (MACS)[**21**]. It is the most significant and most diverse corpus. It contains 1659 manual transcripts of spontaneous oral descriptions of pictures representing 15 situations from everyday life: in the bank, countryside, room, cinema, beach, grocery, kitchen, living room, party, fishing, mountains, park, restaurant, zoo and street. There are two groups of participants - HC and AD.

The HC group includes 101 young and older francophone adults. It comprises 3 subgroups of approximately equal size, divided by the participant's age: 18 - 30 years old, 31 to 50 years old, and the last subgroup with the ages between 51 and 70. Educational levels of the majority of healthy controls vary between 10 and 25 years, with one extreme case of 3 years. Mini-Mental State Examination (MMSE)[**22**] scores are between 25 and 30[2].

Contrarily to the HC group, the AD group is small, with only 10 participants. There is no information available about their age and MMSE scores.

[2]Scores 24 and higher are considered normal; 30 is the maximal possible score in that test.

All healthy controls and all AD participants except one completed descriptions of all scenes.

The length of transcripts varies from 14 up to 1190 tokens, while the vast majority falls in the range of 100-400. Transcripts of the AD group are very distinct from the controls. They contain much fewer details, more repeated words and self-corrections:

HC: *il semble y avoir une dame qui est en train de de recevoir du change, sur un comptoir à un comptoir de change. C'est une banque. Il y a une autre personne qui est au guichet, qui se prépare à mettre ou retirer sa carte, pour faire transaction. Il y a quelqu'un qui a l'air d'être en train de checker l'heure, et qui tient portefeuille, j'imagine, et qui vient d'aller à l'épicerie, semble-t-il. Ça c'est ce qui se passe. Bien, ça dépend. La question c'était... Bien, la consigne était de de dire ce qui se passait sur l'image. Donc, je mentionnerais pas la présence d'une plante.*

AD: *la première image, c'est des gens qui sont en en en activité de de de de de de de de de des de de de de de demander de l'argent à la machine, ok. voilà. euh il y a aussi une une une une dame qui attend. euh euh elle a elle a la la la dame, ici, regarde son son l'heure. euh et une autre dame est à à au au au au comptoir. de de de de au du du du d'une caissière, finalement. puis euh puis euh je peux tu aller aller en arrière, aussi? je veux dire que c'est comme euh... non.*

As we mentioned earlier, the MACS corpus is not limited to the description of a single scene. That provides a unique opportunity to learn to recognize AD from general texts when a predefined set of the main ICUs is impossible to create. Since the complexity of the topic varies between different scenes, it is impossible to rely on the absolute values of many traditional metrics. We had to normalize them and use dependencies between metrics to predict AD from the previously unseen situations. For instance, the expected number of unique words dramatically depends on the length of the text[3], which in turn depends on the complexity of the topic.

Such generic tools which are subject-independent are crucial, for example, for mental health screening on social network platforms and public health analysis.

[3]According to Heap's law, $V = KN^{\beta}$, where V is vocabulary size, N is the length of the text, K is the constant and β is some value between 0 and 1,

Chapter 2

Features

In the age of deep learning algorithms, the old-fashioned machine learning methods may look outdated. We can argue that it is too early to dismiss them. A feature-based approach is still relevant when the data is too small for deep learning. Feature-based methods also allow more control over the features. Since deep models learn high-level representations rather than have them specified by the feature space, these models are challenging to interpret. We need to know which features are good predictors of the diagnosis not only to make more accurate predictions but also to understand better the impairment.

We do not draw a hard boundary between feature-based and deep learning methods. Instead, we will show that neural networks with a reasonable number of hidden layers could be optimal solutions for the problems described by hand-crafted features. We will also show that feature-based methods can use features manually derived from word-vectors.

2.1. Data preparation

The goal of data preparation is to remove factors that can complicate or corrode feature extraction by existing tools. It is an essential step in any machine learning algorithm and any domain, and NLP is not an exception. Indeed, natural languages usually have very complex and diverse grammar rules and a high level of ambiguity. Any anomalies in the sentence structure can be very confusing for sentence annotators like part-of-speech (POS) taggers and parsers.

K. Fraser et al.[**10**] reported a higher error ratio of Stanford CoreNLP POS tagger on AD data compared with control data in DementiaBank corpus. One can get the impression that it is possible to reduce the error rate by cleaning and "fixing" the text since some errors are apparent, such as repetitions of words. We can say that it is not a good idea.

First of all, we have to distinguish between the flaws in the speech and legitimate usage of the language. Natural languages are incredibly complex and diverse, and such distinction is not always trivial. Any rule-based algorithm has high false-positive and false-negative rates, and it is language-dependent. Machine learning algorithms have a better performance than hard-coded sets of rules, but they have to be trained, and they have to parse the sentence in order to distinguish errors.

The second and the most crucial point is that by changing the text, we corrupt it and introduce the bias of the judgment. That can be our own bias, like in the case of hand-crafted rules or the bias of the machine learning algorithm.

Instead of looking for the best possible cleaning algorithm, we suggest to accept the errors in sentence annotation and keep the original text, fixing only the spelling. It is a benign modification since the texts are the transcripts, not the original documents. The errors in parsing could be the features, and we are going to explain how.

Classification of the transcripts in order to detect signs of dementia is not a regular classification task. The objective is not to find a correct parsing or retrieve an information which the author intended to communicate. Here the objective is to use the text to identify the patterns which could suggest some mental disorders of the author. In such a setting, the content of the scene becomes less critical. Much more important is to preserve and identify the patterns, often subtle, which will allow discriminating transcripts of healthy controls from the dementia group. In such settings, we can tolerate some parser errors. Usually, especially in French texts, agrammatical or corrupted sentences result in abnormally flat parsing trees. We can learn to associate such features with the condition of the participant. Similarly, a high ratio of out-of-vocabulary (OOV) words in this research can suggest some mental conditions.

Given these points, we decided not to clean the texts except for the spelling. The results confirmed that it did not compromise the accuracy of predictions, and in some cases, improved it.

2.2. NLP Tools

The scope of this research did not allow to build and train customized NLP tools. We had to resort to pre-trained models for sentence annotation and for word embeddings.

We have found only two packages that provide a full NLP pipeline for sentence annotation with multilingual support. One of them is CoreNLP suite[**23**] with StanfordNLP Python package, which implements an interface to the CoreNLP Java server as well as a fully neural pipeline from the CoNLL 2018 Shared Task competition[**24**]. The second one is Spacy[**25**] - a Python package for sentence annotation, which supports over 50 languages. Both packages implement tokenization, lemmatization, POS tagging and dependency parsing. However, CoreNLP also offers constituency parsing, which is missing in Spacy. Besides, CoreNLP is slightly more accurate at the expense of speed. So we selected CoreNLP.

In our experiments, we have used four English and two French word-space models. All of them obtained from FastText. Two of these models (one English and one French) were trained on Wikipedia with the SG algorithm using subword information. All other models (three English and one French) were trained with the CBOW algorithm[**26**] on a Common Crawl and Wikipedia, one model with the conventional CBOW and the others one with subword information[**27**] - see table 2.1 for a short summary. We make a detailed analysis of those models in section 2.4.2.

Besides sentence annotation and word-vectors, we utilized word frequencies, which we obtained from the Python Wordfreq package[**28**].

	SG-subwrd	CBOW	CBOW-subwrd
EN	*Wiki*	*CC*	*Wiki*,*CC*
FR	*Wiki*	-	*Wiki*

Table 2.1. Summary of word-vector models.

Wiki stands for Wikipedia, *CC* stands for Common Crawl

2.3. Lexical and syntactic features

Lexical and syntactic features are the oldest and the most studied group of linguistic features for predicting AD. Long before the beginning of the computer era, in the 1970s, Schwartz et al.[**29**] described the effects of the changes in the language function of the brain on the lexical forms of language structure. Since that time, this topic remains an active area of research.

2.3.1. Lexical features

This group is the primary group of linguistic features used to detect and discriminate different cognitive impairments. Multiple studies reported that patients with AD have the ratio of pronouns to full noun phrases higher than the control group [**30, 10, 31**]. Other features like TTR, Moving-Average Type-Token Ratio (MATTR)[**32**], open:closed word ratio, verb rates, vocabulary size, average word length and syntactic complexity were also have been studied and found to be predictors of cognitive impairments, but not all of them were markers of AD. For example, a reduced vocabulary size can suggest a form of Primary Progressive Aphasia, and we have seen it in the Picnic corpus. Non-words and out-of-vocabulary words are indicators of disfluency. However, they can be a symptom of other types of dementia as well.

In our experiments, we have implemented two groups of lexical features. The first group is independent of the POS tagger, and it includes features which describe vocabulary:

- **vocabulary size** - we have collected two variants of the statistics - original, non-normalized tokens as they appeared in the text with inflections, and their lemmas.
- **text length** - the total of all tokens after removing punctuation
- **Honoré's statistic**
- **Brunét's Index**

- **TTR** - calculated separately for original tokens and their lemmas
- **MATTR10** - average TTR in a moving window of size 10, after filtering punctuation (calculated separately for original tokens and their lemmas)
- **MATTR3** - average TTR in a moving window of size 3, after filtering punctuation (separate statistics for original tokens and their lemmas)
- **BiMATTR50** - average number of unique bi-grams (two consecutive tokens) in a moving window of size 50, after filtering punctuation (separate statistics for original tokens and their lemmas)
- **average word length** - total number of characters in all tokens divided by the text length
- **average word Zipf frequency** - Zipf frequency is base-10 logarithm of the number of times a word appears per billion words

The second group, which is tagger-dependent (see section 2.2), describes POS statistics:

- **pronoun proportion** - a ratio of pronouns to the sum of nouns and pronouns according to Penn Treebank (PTB) or Universal Dependencies (UD) tag sets. Nouns group includes common and proper nouns (tags starting with *NN* if using PTB tag set or *NOUN* and *PROPN* if using UD tag set , pronouns include personal and possessive pronouns (tags starting with *PRP* if using PTB tag set or *PRON* if using UD tag set)
- **verb proportion** - the ratio of verbs to the sum of nouns, pronouns and verbs, according to PTB or UD tag sets. Note that in UD tagset auxiliary verbs (*AUX*) is a separate group of verbs, which is not included in the group of verbs. The same is true about the modal verbs (*MD*) in PTB tag set
- **interjection proportion** - the ratio of interjections (tags *UH* and *INTJ*) to all other parts of speech, excluding punctuation, symbols and spaces
- **modal proportion** - the ratio of modal and auxiliary verbs (tags *MD* and *AUX*) to all verbs, including modal and auxiliary
- **open:closed ratio** - a ratio of open class tokens to closed class tokens, i.e. ratio of nouns, verbs, adverbs, and adjectives to the number of pronouns, prepositions, conjunctions and determiners

2.3.2. Sentence structure

A strong syntactic structure is the main factor, which differentiates a bag of words from the language. Syntax organizes words into hierarchical structures of phrases, greatly expanding the expressiveness of otherwise would be isolated words. It is intuitive to suggest that the sentence structure would be sensitive to various forms of aphasia. Indeed, an inability to produce syntactically well-formed sentences, reduced sentence length or lack of embedded clauses are markers of Broca's agrammatic aphasia [**33, 34**]. K. Fraser et al.[**10**] noted that the tendency of AD patients to replace nouns with pronouns and the frequent usage adverbs *here* and *there* in place of nouns change the distribution of syntactical patterns.

The role of the parsers is to recover sentence structures and describe them through parse trees. Two different formalisms exist in the organization of trees: constituency and dependency. The former split the sentence into phrases, which are the nodes on the tree, while the leaves - terminals - are literals, i.e. words. Production rules describe the hierarchy of phrases. Contrarily to the constituency trees, dependency trees do not split the sentence into phrases but build the hierarchy directly from the words.

In general, the constituency trees, due to their phrase-oriented structure, are more suitable for the languages with the fixed order of words like English. Since dependency trees do not have a concept of phrase, they allow more freedom in word linearisation. That makes dependency trees the natural choice for the parsing languages where the order of words is defined pragmatically, like Slavonic languages. However, that does not mean that dependency trees do not suit for the languages with the predefined order of words. In the same way, the languages having a more flexible order of words can be parsed with constituency parsers. Though, the flexibility and more diverse patterns of the sentence make it more challenging to derive the features from production rules.

Basing on the analysis of production rules in English texts, K. Fraser et al.[**10**] pointed on an elevated frequency of $NP \rightarrow PRP$ phrases in the AD group. Such a feature becomes less prominent in French because French has more possibilities to create a noun phrase with a pronoun as its head.

Since we aim to find a universal set of features, which is not language-dependent, we decided not to consider features derived from production rules[1]. We replaced such features with the features that describe the content of the dependency subtree, e.g., the size of subtrees having a noun or pronoun as their root.

The full list of dependency tree features includes:

- **average max depth of the tree** - the longest path from the root
- **average max width of the tree** - the maximal number of children in the node
- **average number of trees with nsubj** - the average number of the sentences that have a subject
- **proportion of noun chunks** - the ratio of the size of all subtrees having the noun or pronoun as their root (the first noun or pronoun node in the path from the tree root) to the total size of all trees. This feature describes the richness of the noun clause. For example, the ratio of noun chunks in the sentence *The window was open* is 2/4, but in the sentence *The big window to a cosy garden was open* such ratio would be 7/9, and in the sentence *It was open* such ratio would be only 1/3.

[1]It is remarkable that multi-lingual NLP packages like Spacy and StanfordNLP neural pipeline do not include constituency parsers. CoreNLP provides constituency parser, but it is not a part of its multi-lingual pipeline

- **average dependency distance** - the average distance between the governor and dependent.
- **average dependency diversity** - the number of unique dependencies divided by the number of all dependencies (similar to TTR, but it describes dependencies)

In addition, we collected statistics for each dependency type: the count of one dependency type divided by the number of all dependencies.

2.4. Word-vector space features

2.4.1. Using word-vector space for measuring Idea Density

ID measures the rate of ideas expressed per word. Two alternative definitions of ID exist in neurological studies. PID[**35**] counts *any* expressed ideas (a.k.a. propositions) with no restrictions to the topic. Each proposition corresponds to a boolean question, which is possible to ask based on the text. In contrast to this, SID [**36**] is a recall metric of predefined ICUs such as an object or action. SID is naturally more applicable to closed-topic domains like the description of the picture. Both variants of idea density are complementary to each other, and each of them requires semantic comprehension of the text.

In recent years, much of research has been conducted to automatize the extraction of both forms of ID and potentially redefine its previously existing definitions. K. Sirts et al.[**12**] proposed Dependency-based Propositional Idea Density (DEPID), a method for computing PID from semantic dependencies. They noted that it is possible to make a reliable approximation of the PID score by counting the number of dependencies.

Yancheva and Rudzicz[**11**] have proposed a method of calculating SID, which brings SID closer to PID regarding restrictions of topics. Their method does not require a human-supplied set of ICUs. Thus, it is applicable for a free-topics dataset, similar to PID. They cluster vector representations of nouns and verbs found in the transcript into a predefined number of clusters and interpret centroids of those clusters as automatic ICUs. Since this method does not require a human-supplied set of ICUs, it is also applicable for free-topics datasets, similarly to PID.

While these new methodologies eliminate a significant part of manual processing and simplify ID scoring, none is completely free of human supervision. In most cases, they are also language-specific. The relation of propositions to semantic dependencies is not symmetrical, and not each dependency corresponds to a valid proposition. Thus, the set of valid dependencies has to be hand-crafted or learned by some other algorithm. Automatic extraction of ICUs still requires a careful selection of the number of clusters and the alignment of transcripts with these clusters.

We propose a new version of ID scoring, which is more computational friendly and language-independent. It does not require any structured analysis, and it is based solely on word embeddings. We assume that words corresponding to similar concepts will have higher cosine similarities of their embeddings. Then, measuring the average cosine (dis-)similarities in a moving window, we can assess the density of ideas. Such a score of density will not represent the pragmatism of ideas expressed in the text because it will treat all words within the window as a bag of words. Both, it will not measure the significance of ideas because cosine similarities consider only directions of word vectors and ignore their lengths. The main advantage of this method is its extreme simplicity.

Our interpretation of idea density is similar to the Repetitiveness feature proposed by K.Fraser et al.[**10**]. They also used the cosine relation of word embeddings to score average distance between utterances after removing stopwords. In contrast, we do not group words into utterances and do not look for similarities (or dissimilarities). Instead, we treat each word separately and measure diversity within the sliding window of fixed size. Since we do not compare the signatures of the sentences, there is no need to remove stopwords. In our interpretation of the metric, repeated words do not harm the accuracy of results either. In opposite, that will make results even more expressive because people with dementia tend using more words with broad senses.

We calculate the score only locally within a moving window of size 10 (this number is arbitrary). That reduces computational complexity from $O(N^2)$ to $O(NW)$, where N is the number of words, and W is the the window size. Indeed, the number of word pairs in the window is $\frac{W(W-1)}{2}$. The window can be applied to $N-W+1$ positions in the text. When moving the window, there is no need to recompute all pairs again. On each step, there are $W-1$ new pairs added, and the same number of pairs is removed from the window. There are $N-W$ positions in the text where we can get the advantage of that and only one where we have to compute all possible pairs from scratch. Therefore the total number of pairs in the text is

$$\frac{W(W-1)}{2} + (N-W)(W-1) \tag{2.4.1}$$

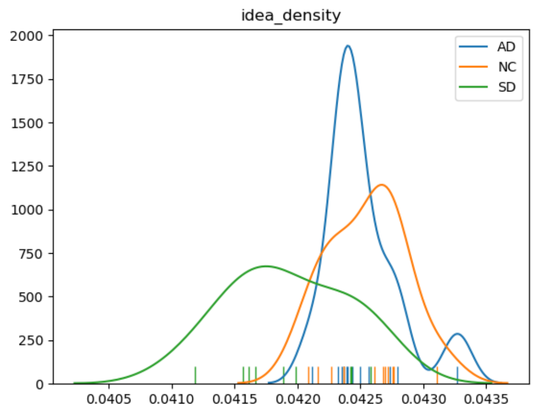
When $N >> W$, then the complexity of this expression is linear.

In our experiments, we have used the following formula for calculating the score of the window:

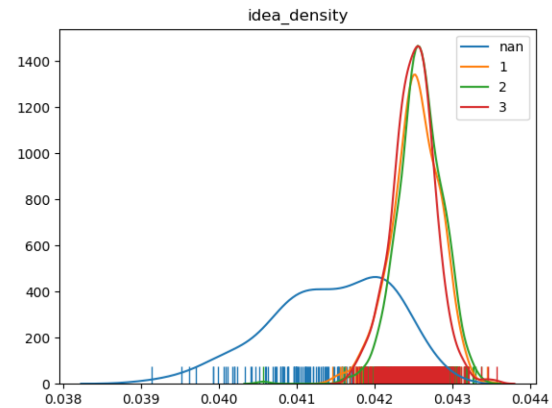
$$idea\_density(window) = \frac{\Sigma_{i=1}^{N-1}\Sigma_{j=i+1}^{N}(\sigma(k \times sim(i,j))}{\frac{N(N-1)}{2}} \tag{2.4.2}$$

where N is the number of tokens in the window, $k < 0$ is a negative constant, $sim(i,j)$ is cosine similarity of i-th and j-th tokens and σ is the sigmoid function.

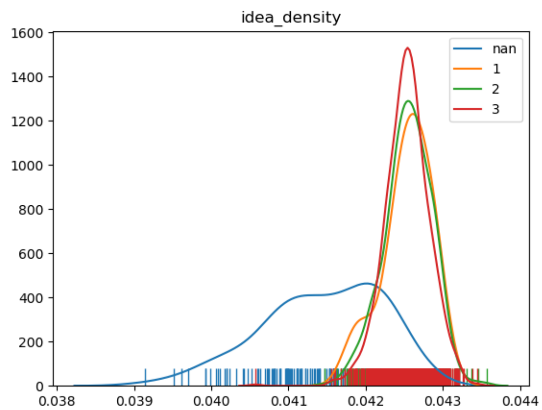
Requirement $k < 0$ flips the sign of cosine similarity such that similar words would contribute to the idea density less than dissimilar. Sigmoid function converts (dis-)similarities to probabilities, such that all results would be in the range (0,1); the magnitude of k controls skewness of sigmoid.



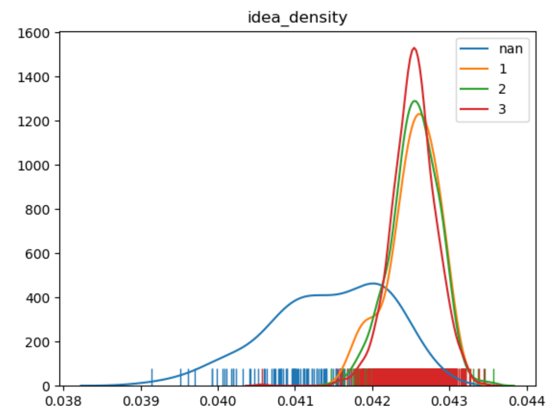
(a) Picnic dataset - AD vs svPPA vs HC



(b) MACS dataset - Dementia ("nan") vs HC by age group (1, 2, 3)



(c) MACS dataset - Dementia ("nan") vs HC by educational groups (1, 2, 3)



(d) MACS dataset - Dementia ("nan") vs HC by sex (0, 1)

Fig. 2.1. Gaussian Kernel Density Estimation (KDE) of ID scores SG word-vector model, Picnic and MACS corpora

Graphs describe distributions of ID scores for each of the classes according to the SG word vector models. Horizontal axis is shows ID scores, vertical axis shows the density estimation. The data points are the rug plots on horizontal axis.

Fig. 2.1 and fig. 2.2 show a clear distinction of ID score distributions between dementia and control groups in all corpora. We can see particularly low scores in the svPPA group (class Semantic Dementia (SD) in fig. 2.1a) of the Picnic dataset. We cannot say that these two groups have significantly different scores due to the small size of this dataset (only 33 transcripts divided between 3 classes) and similarity between KDEs of AD and HC groups. However, significantly bigger MACS dataset with 1659 transcripts (fig.2.1b, 2.1c and 2.1d) and DementiaBank dataset with 536 transcripts (fig.2.2) confirm lower ID scores for dementia group (De) comparing to controls (Co). MACS dataset does not distinguish between different types of dementia. However, its control group is divided into subgroups by age, level of education and Sex. All participants of the dementia group of that corpus have a severe form of dementia. One can observe there that the distribution of ID scores has no significant variation between control groups. In all cases, the factor which most affects ID score is dementia.

Dementia group (De) of DementiaBank corpus includes patients with different severity of the impairment, and the differences in distributions of ID scores are less pronounced there than in other corpora. Nevertheless, we can see that the distinction between the two groups does not depend on the selected word-vector model. We obtained similar results using SG and several CBOW models trained on different sources. In all cases, the ID scores in the De group were meaningfully lower than in the controls. We have observed the same phenomena in both French datasets when replaced SG model with the CBOW.

In section 4.1, we will return to the analysis of the proposed interpretation of ID. We will examine there how this feature interacts with other features and its impact on a prediction power of classifier.

2.4.2. Using word-vector space for the analysis of proposition specificness

2.4.2.1. *Motivation*

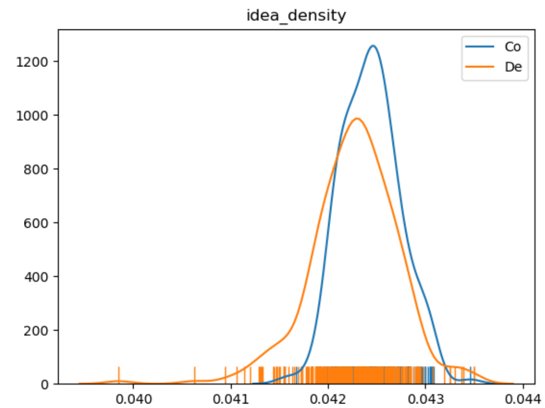
One of the early signs of dementia or AD is a vague speech. It is expressed not only in the augmented ratio of pronouns to nouns but also by the reduced explicitness of used words. Aronoff et al.[**37**] observed that the impairment of the relational knowledge often accompanies impaired naming abilities of AD patients. Yi et al.[**38**] described difficulties of AD patients to select the verb that was best illustrated by the description. Kim et al.[**39**] showed that verb deficits in the AD are based on semantic complexity. Our study shows that such deficiencies of AD patients can be traced through the anomalies in the profiles of word-vectors, associated with the patient's vocabulary. Namely, the average of word-vector magnitudes can be significantly lower or higher in the AD group than in controls, depending on the specific word-vector model used for the evaluation of both groups.

We find vector norms (a.k.a. magnitudes) to be a convenient tool for measuring word significance because they are rotationally invariant.[2]

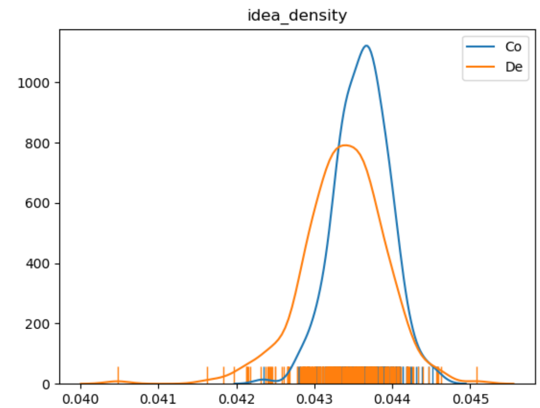
2.4.2.2. *Background*

Word-vectors are trained on word co-occurrences, which ensures that similar words (or words frequently used together) would have similar representations[**26**]. As a consequence of such training, words that appear in diverse contexts are represented by a weighted average of various meanings. In turn, the averaging over vectors pointing in different directions results

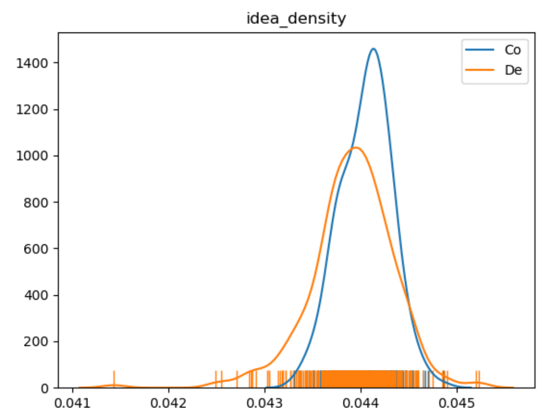
[2]The angle of the rotation of word-vector space around its origin is purely random and meaningless.



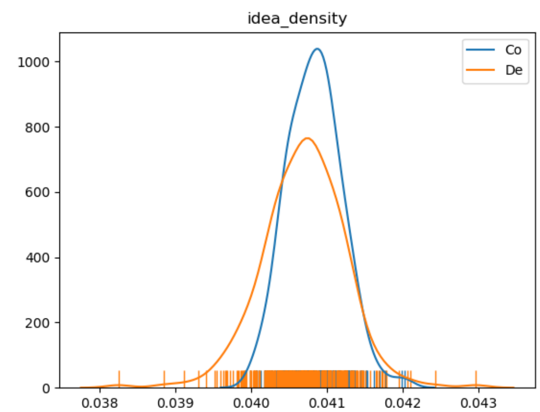
(a) SG Wikipedia model with subword info



(b) CBOW Wikipedia model with subword info



(c) CBOW model Common Crawl without subword info



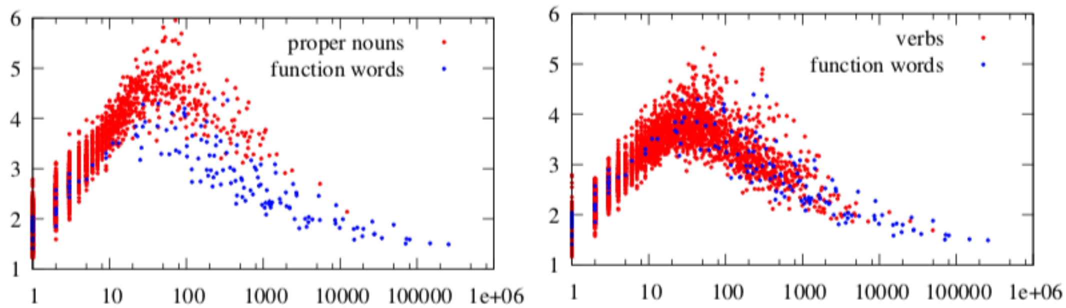
(d) CBOW model Common Crawl with subword info

Fig. 2.2. Gaussian KDE of ID scores - DementiaBank corpus

Four different word-vector models trained with SG or CBOW algorithms. The changes in the model do not have a significant impact on the differences of KDE between two groups

in a shorter vector[**40**]. Frequent words with fixed contexts tend to be represented by longer vectors with higher norm values. However, when the word becomes more prevalent, it starts to appear in more diverse settings, and the effect of the context diversity overweights the impact of frequency. The length of the vector begins gradually decreasing. We can observe these opposite trends in fig 2.3, which visualizes distributions of vector lengths by term frequency for nouns and verbs.

This behavior is consistent between CBOW and SG models, even though models have different optimization objectives. CBOW optimizes the prediction of the word from the summary of surrounding words (e.g., a weighted average of surrounding words). In contrast, SG optimizes the prediction of the surrounding words given the word (each surrounding word



(a) Nouns **(b)** Verbs

Fig. 2.3. Word vector norm v versus term frequency *tf*

Note axes of *tf* are in logarithmical scale. *Images from Adriaan M. J. Schakel, Benjamin J. Wilson "Measuring Word Significance using Distributed Representations of Words"*[**40**]

is predicted individually). Both models generate word-vector space as a by-product of their optimization. Due to the different optimization objectives, the resulted word-vector spaces have different properties. For example, SG model, in contrast to the CBOW model, can capture the different semantics of a single word and make distinct vector representations for each of them[**41**]. However, if the word has a broad general usage (e.g. the verb *get*) rather than multiple distinctive narrow usages (e.g. *apple (the fruit)* vs. *Apple*), then it will have a single vector representation. Its norm will be decreased in both models. Such properties of the word-vector space allow us to conclude that the length of the vector is related to its explicitness. That suggests that we can use the average of vector norms as a predictor of the AD.

2.4.2.3. *Subword information*

Enrichment of word vectors with the subword information has complicated interpretation of the word-vector space. The resulted space largely depends on how this information has been integrated within the model. We have studied two different methods of integration of subword information

In one method, which is described in [**16**], the word was represented by the sum of the vectors associated with its character n-grams. For example, the character 3-grams (i.e. $n = 3$) representation of the word *where* will be

<wh, whe, her, ere, re>

If we denote word's n-gram vector as z_g where $g \in G_w$, a set of all word's n-grams and the context vector as v_c, then we can write the scoring function of SG model, which matches word w and a context word c as

$$s(w,c) = \Sigma_{g \in G_w} z_g^T v_c \qquad (2.4.3)$$

Optimization of such function (i.e., maximization of its value for the matching context and minimization for the negative samples) has a similar effect on the distribution of n-gram vectors. That was consistent with the function which operated on word-vectors. More diverse matched contexts lead to lower vector norms. Such an effect is propagated to the resulted word-vectors.

The situation changes dramatically when the subword vectors are pre-trained, like in the recent CBOW model[**27**] of FastText. That model obtains vector representations of n-grams from an external algorithm, and enriches the word-vector v_w representing the word w with the average of its n-gram vectors:

$$v_w + \frac{1}{|G|}\Sigma_{g \in G_w} z_g \tag{2.4.4}$$

This expression substitutes v_w in the original scoring function of CBOW model, which scores word w and the context C:

$$s(w, C) = \frac{1}{|C|}\Sigma_{c \in C} v_c^T v_w \tag{2.4.5}$$

Since long words comprise a higher number of n-grams than short ones, the average of their n-gram vectors is more likely to include vectors pointing in different directions. As a result, those vectors cancel each other. Longer words tend to have a smaller weight of their subword component in 2.4.4 than short ones. Accordingly, vectors associated with longer words became more dependent on their context. That partially overrides the effect of word specificity in the CBOW model when enriched with subword information.

Appendix A.0.1 illustrates the impact of subword information on CBOW models. Two word-vector spaces were trained on the same data (Common Crawl), both with the CBOW algorithm. One was enriched with subword vectors, while the other was trained with the conventional CBOW. We compare these models using the vocabulary of the DementiaBank corpus. Noticeably, the model with subword information has a significantly lower range of vector norms than the model without subword information. Morphologically complex words in the model with subword info are represented with shorter vectors than simple words.

We can observe the same phenomena when we compare models with imported n-gram vectors (the scoring function as in eq. 2.4.5 with substituted 2.4.4) versus models with embedded n-gram vectors (scoring function as in eq. 2.4.3). Appendices A.0.2 (the former scoring) and A.0.3 (the latter) illustrate the differences in the distribution of vector norms between these two approaches on examples of English and French models. In both languages, the models with the imported subword information have the lower vector norms associated with more complex words. In a contrast, the models with embedded subwords have distribution consistent with the models without subword information.

2.4.2.4. *Observations*

We have studied whether any meaningful difference in the distributions of vector norms exists between participant groups and whether these differences are consistent between various word-vector models.

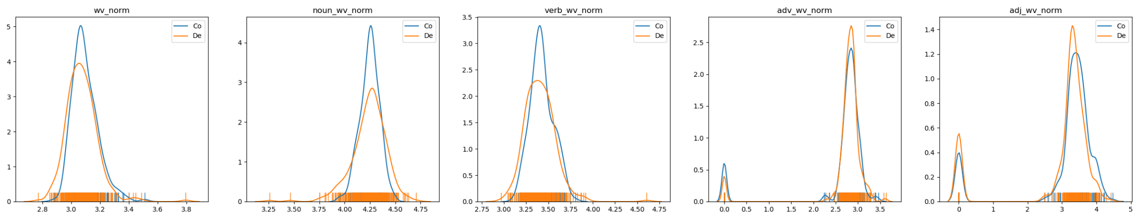
For each transcript, we have calculated the average of vector norms of four classes of open-class words, separately for each class - nouns, verbs, adjectives and adverbs. The fifth group of words included all alphabetical words together. We have measured the statistical significance of the differences between participant groups with the Kolmogorov-Smirnov and Anderson-Darling tests for k-samples. Figures 2.4 and 2.5 visualize those distributions, and tables 2.2 and 2.3 provide measurements of statistical significance of the observed difference.

		AD-Statistic	AD-p-value	KS-alpha	KS-p value
Nouns	PicnicSD-NC	-0.242	0.25	0.333	0.547
	DementiaBank	**6.595**	**0.001**	**0.132**	**0.018**
	MACS3	**35.099**	**0.001**	**0.290**	**0.001**
Verbs	PicnicSD-NC	-0.688	0.25	0.278	0.73
	DementiaBank	**3.992**	**0.008**	**0.143**	**0.008**
	MACS3	**7.912**	**0.001**	**0.128**	**0.046**
Adj	PicnicSD-NC	-1.038	0.25	0.222	0.932
	DementiaBank	**8.375**	**0.001**	**0.174**	**0.001**
	MACS3	**40.904**	**0.001**	**0.299**	**0.001**
Adv	PicnicSD-NC	-0.534	0.25	0.25	0.851
	DementiaBank	-0.339	0.25	0.051	0.866
	MACS3	**29.617**	**0.001**	**0.281**	**0.001**
All	PicnicSD-NC	-0.112	0.25	0.333	0.547
	DementiaBank	**6.911**	**0.001**	**0.157**	**0.003**
	MACS3	**20.454**	**0.001**	**0.241**	**0.001**

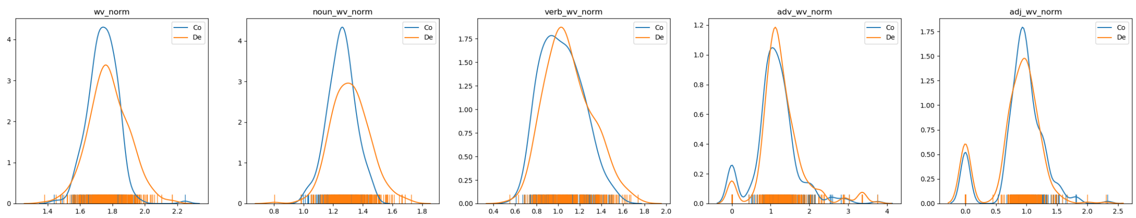
Table 2.2. AD- and KS-tests for k-samples, SG models with subword info

p-value columns show the probability of obtaining the observed result due to a sampling error if the *null hypothesis* is true (i.e. if results of Dementia and HC originate from the same distribution); **Bold** font indicates results with $p \leq 0.05$. p-values of AD-test are floored / capped at 0.001 / 0.25. p-values of KS-test are floored at 0.001, but not capped. *PicnicSD-NC* is Picnic dataset, svPPA (a.k.a. SD) group vs. HC (a.k.a. NC group), *MACS3* is MACS dataset, dementia group vs. age group 3

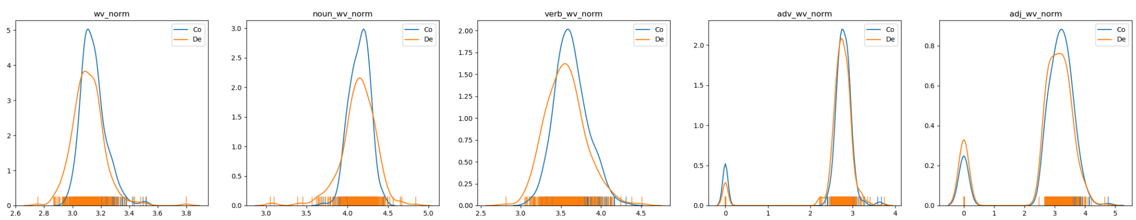
We can see that for DementiaBank and MACS corpora, all groups except adverbs in DementiaBank corpus have p-values lower than the commonly accepted threshold 0.05 for statistical significance. Those numbers are consistent with KDEs plotted in Figures 2.4 and 2.5. While Picnic dataset statistics are at least high as the cupped 0.25 significance level, the plots of their KDEs still demonstrate differences of distributions. We assume that high p values for that dataset were due to the tiny size of that corpus, which has only 33 data points distributed between three classes.



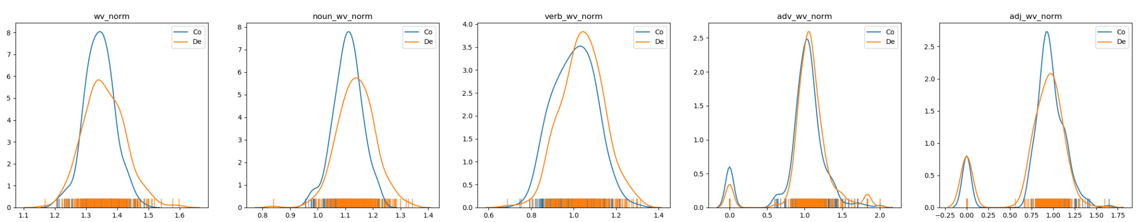
(a) DementiaBank dataset - SG model with subword info, Wikipedia



(b) DementiaBank dataset - CBOW model with subword info, Wikipedia



(c) DementiaBank dataset - CBOW model without subword info, Common Crawl

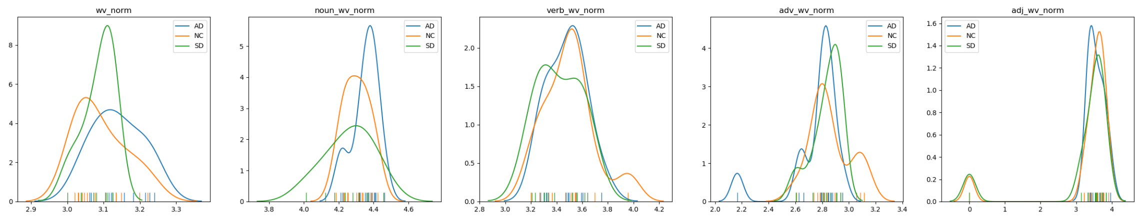


(d) DementiaBank dataset - CBOW model with subword info, Common Crawl

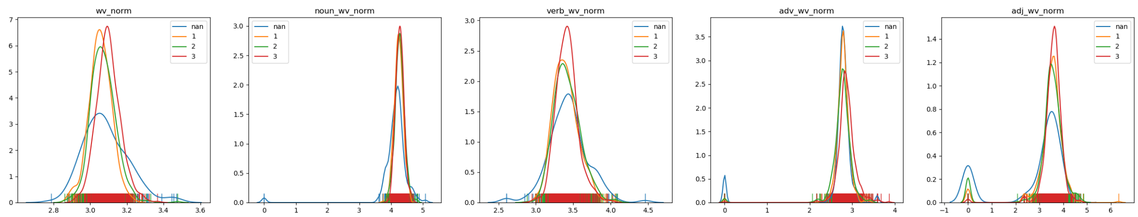
Fig. 2.4. Gaussian KDE of English word-vectors norms

The data points are the rug plots on horizontal axis. Columns from left to right: nouns, verbs, adverbs and adjectives together; only nouns; only verbs; only adverbs; only adjectives. Both CBOW models with subword info have higher vector norms for dementia groups than for controls, while SG model as well as CBOW model without subword info have the lower vector norms for dementia group than for controls.

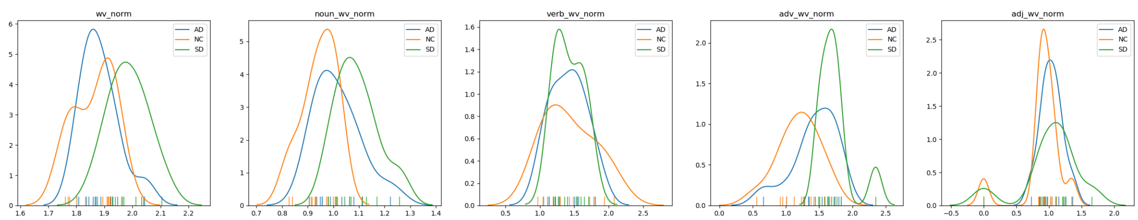
Remarkably, there was a significant difference in the group where we did not regard word's part-of-speech. One can explain it by a high ratio of pronouns and non-words, as well by the reduced sharpness of the speech of people with dementia in general. This metrics is essential because it provides a numerical estimation of such change and does not depend on any NLP tool except word embeddings.



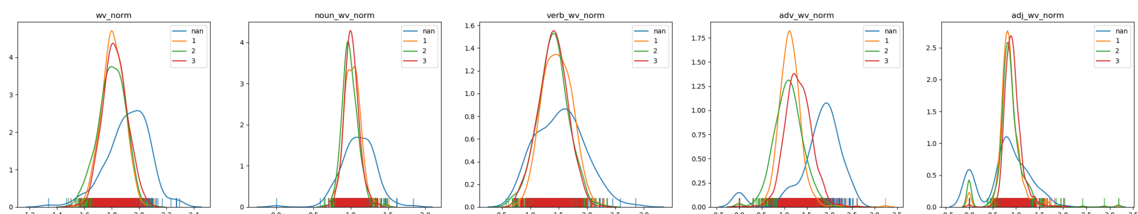
(a) Picnic dataset - SG model with subword info



(b) MACS dataset - SG model with subword info



(c) Picnic dataset - CBOW model with subword info



(d) MACS dataset - CBOW model with subword info

Fig. 2.5. Gaussian KDE of French word-vectors norms

The data points are the rug plots on horizontal axis.Columns from left to right: nouns, verbs, adverbs and adjectives together; only nouns; only verbs; only adverbs; only adjectives. Both CBOW models with subword info have higher vector norms for dementia groups than for controls, while SG model as well as CBOW model without subword info have the lower vector norms for dementia group than for controls.

Another noticeable fact is that these results have been confirmed with various word-vector models in two languages. There is a statistically significant difference between distributions of word-vector norms between healthy controls and dementia groups. It does not depend on whether the model included subword info or not, and regardless of the type of the model (SG vs. CBOW).

		AD-Statistic	AD-p-value	KS-alpha	KS-p value
Nouns	PicnicSD-NC	1.625	0.069	0.444	0.206
	DementiaBank	**21.651**	**0.001**	**0.256**	**0.001**
	MACS3	**43.824**	**0.001**	**0.365**	**0.001**
Verbs	PicnicSD-NC	-0.712	0.25	0.25	0.851
	DementiaBank	**6.604**	**0.001**	**0.136**	**0.014**
	MACS3	**20.451**	**0.001**	**0.255**	**0.001**
Adj	PicnicSD-NC	-0.542	0.25	0.278	0.729
	DementiaBank	**2.479**	**0.031**	0.108	0.083
	MACS3	**38.892**	**0.001**	**0.304**	**0.001**
Adv	PicnicSD-NC	1.543	0.075	0.5	0.114
	DementiaBank	**5.531**	**0.002**	**0.152**	**0.004**
	MACS3	**125.595**	**0.001**	**0.622**	**0.001**
All	**PicnicSD-NC**	**4.641**	**0.005**	**0.639**	**0.018**
	DementiaBank	**12.401**	**0.001**	**0.209**	**0.001**
	MACS3	**60.569**	**0.001**	**0.416**	**0.001**

Table 2.3. AD- and KS-tests for k-samples, CBOW models with subword info

p-value columns show the probability of obtaining the observed result due to a sampling error if the *null hypothesis* is true (i.e. if results of Dementia and HC originate from the same distribution); **Bold** font indicates results with $p \leq 0.05$. p-values of AD-test are floored / capped at 0.001 / 0.25. p-values of KS-test are floored at 0.001, but not capped. *PicnicSD-NC* is Picnic dataset, svPPA (a.k.a. SD) group vs. HC (a.k.a. NC group), *MACS3* is MACS dataset, dementia group vs. age group 3

In section 4.1, we will turn to the analysis of the proposed word-vector norm features in the context of other features and see their performance in the classifier.

Chapter 3

Classification

3.0.1. Feature selection

3.0.1.1. *Background*

Feature selection process is an integral part of the model, regardless of whether it is embedded into the classifier (e.g., regularization) or performed as a separate step. Since selected features significantly impact the model's performance, a feature selection algorithm should not obtain any information from the data, which we use to measure the performance of the model. If we use cross-validation, then feature selection has to be done separately within each fold; otherwise, we contaminate test data and bias the model.

There are different criteria for feature selection. The most popular one - at least in the literature describing linguistic features for identifying AD - is statistical filtering. The authors apply Student's T-test to measure the statistical significance of the differences in the value of the feature between different groups. We have to note that Student's T-test measures the significance of the mean's differences, and it is valid only for normal distributions. If two distributions have similar means but radically different variances, Student's T-test will not find these distributions significantly different. The second point is the type of distribution, which has to be Normal, i.e., has to pass a goodness-to-fit test for normality. In our experiments, only a few features have passed such a test.

Statistical significance tests are not limited to Student's T-test, and they are not limited only to univariate analysis. However, regardless of the method of analysis, all filtering methods are disconnected from the learning algorithms. In result, selected features might be not optimal for the target learning tasks.[**42**].

One of the alternatives for the feature filtering, which addresses the connection between feature selection and the learning algorithm, would be wrapper methods. They can be seen as an interactive system alternating between feature selection and its evaluation by the learning algorithm. Unfortunately, such methods are not practical due to the exponential complexity of an exhaustive search, which is prone to overfitting.

Another option would be embedded methods, i.e., regularization and decision trees. Such methods are integral parts of the learning algorithm. They have low complexity ($\log n$) and help avoid overfitting. Their drawback is dependence on the classifier.

3.0.1.2. *Proposed method*

We selected features using a blend of all the above methods. As an initial step, we ranked all features with the Robust Feature Selection (RFC) algorithm[**43**]. That algorithm jointly minimizes $l_{2,1}$ norm of the loss function and regularization:

$$\min_{W} J(W) = ||X^T W - Y||_{2,1} + \gamma ||W||_{2,1} \tag{3.0.1}$$

where γ is a regularization parameter, W is feature weights matrix and $||W||_{2,1}$ is the $l_{2,1}$ norm of matrix W defined as

$$||W||_{2,1} = \Sigma_{i=1}^{n} ||w^i||_2 \tag{3.0.2}$$

Since that algorithm has a dual objective - loss function and weights, we consider this method of feature selection as an embedded method, even though we deploy it as a filtering. In order to remove noisy features and obtain meaningful feature rankings, we utilized very aggressive values of γ. Depending on the dataset, its values ranged from 32 to 1024.

Then we tested different thresholds for the number of selected features using the wrapper method, with the top N features according to their ranking. On that step, we did only a mild pruning and eliminated at most $10 - 15\%$ of the total 80+ features.

The last step was the embedding feature filtering method of the main classifier. One can ask why do we need two embedded feature selection methods applied sequentially, especially if we prune only a few features after the ranking. The main reason is that we could filter the noisiest features before they reached the main classifier. Empirically, some pre-filtering was beneficial for the accuracy of predictions. That observation was consistent for all datasets. Two-step classification allowed us to treat their dual optimizations objectives differently. In the first step (ranking), we prioritized minimizing weights on account of the loss. In the second step, the main objective was to minimize the loss. It is important to note that none of these steps completely ignored the secondary optimization objective. That was the main advantage of our paradigm comparing with the classical feature filtering. Ranking of features also converted the exponential complexity of the wrapper method to the linear.

3.0.1.3. *Selected features*

As we have mentioned earlier, cross-validation requires an independent feature selection process for each fold of the data. Exact feature ranking in cross-validation is impossible, but simple methods like statistics of selection or average ranking can provide a coarse approximation of feature importance.

Top features that have been selected in all folds of cross-validation (ordered by feature ID, not by importance):

- vocabulary size
- lemmatized vocabulary size
- text length
- number of sentences
- TTR
- lemmatized TTR
- MATTR with window size 10 tokens (MATTR10)
- lemmatized MATTR with window size 10 tokens
- bigram MATTR with window size 50 tokens
- lemmatized bigram MATTR with window size 50 tokens
- MATTR with window size 3 tokens (MATTR3)
- lemmatized MATTR with window size 3 tokens
- avgerage zipf frequency
- Honoré's statistic
- Brunét's Index
- pronoun proportion
- verb proportion
- modal proportion
- interjection proportion
- open:closed class ratio
- idea density
- average word length
- average sentence length
- average dependence tree height
- average dependence tree width
- average number of dependence trees having a subject
- proportion of noun chunks in dependency tree
- average word-vector norm of all alphabetical words
- average word-vector norm of all nouns
- average word-vector norm of all verbs
- average word-vector norm of all adverbs
- average word-vector norm of all adjectives
- average dependence distance in the sentence
- dependency diversity

3.0.2. Classification methods

The main objective is to identify dementia from the text. It is not a traditional classification problem, where we need to distinguish between discrete categories. As we discussed in chapter 1, dementia starts gradually, and there is no distinct boundary exist between normal and abnormal. However, the labels in all corpora indicate only the class. They do not provide information about how strongly a transcript belongs to its class[1]. That creates a situation where a continuous score of dementia has to be mapped into discrete classes. In effect, some border cases could belong to different classes, despite having very similar characteristics.

There are multiple algorithms exist which can predict the class - Naive Bayes (NB), decision trees (including ExtraTree and ensembles of trees), SVM, logistic regressions. Many of these algorithms (SVM, logistic regression) maximize the margin between predicted classes, which might be suboptimal for the problems where border cases are frequent. Other algorithms, like decision trees, only predict the class without maximizing a margin between classes. The third kind of algorithms only ranks samples and leaves the interpretation of the ranking to the user.

Decision trees could be the best match for this specific problem because they separate data into classes without amplifying the margin between them. Unfortunately, Decision trees are unstable - a small change in the data can lead to a completely different structure of the tree. Their decision boundaries are not smooth - each split cuts the space by a straight line, and the intersection of such lines create angles. Many other algorithms, despite being theoretically suboptimal for the described problem, often provide a significantly better classification than trees.

We have selected for our experiments the ANNs framework, mainly because of its flexibility. With this framework, we have implemented a softmax regression - a generalization of logistic regression to multi-class problems. In the context of binary classification, softmax regression is equivalent to logistic regression. However, it allowed us to handle also a 3-class problem in the Picnic dataset. The types of the features - summaries of the text, which have no spatial information - suggested a Feed-Forward Neural Network (FNN) architecture known as Multilayer Perceptron (MLP).

3.0.2.1. *MACS classification*

Another reason for selecting ANN as an implementation of the classifier is their extreme flexibility in general. Besides the ability to tune the number of layers, neural networks

[1] Since the DementiaBank dataset was created during a longitudinal study, where participants have been interviewed multiple times during a prolonged period, we can see the progression of the symptoms when we compare transcripts of the same participant. That suggests only a relative order by severity between the transcripts of the same participant. It does not provide an order between different participants.

support various loss functions and are the natural choice for multi-target learning. For instance, the network can be trained to predict not only the probability of AD but at the same time, also MMSE score, educational level, age and sex of the participant. It is very intuitive to suggest that additional information about participant can benefit the quality of predictions. Despite the fact that such information is provided as targets rather than input features, it helps classifier to separate healthy aged participants from the potentially younger participants but having dementia. Additional outputs - predicted age and educational level - can explain some features of the text, which otherwise could be explained as dementia symptoms.

3.0.2.2. *Evaluation*

A choice of evaluation criteria depends on the objective of classification and distribution of classes. Since there is no distinct boundary exist between the classes, the goal of classification is to rank samples rather than pool them to the opposite poles. That means we have to maximize the probability that in any randomly chosen pair of positive and negative samples, a positive sample would be ranked higher than a negative one:

$$A = \frac{\Sigma_{i=1}^{m} \Sigma_{j=1}^{n} \mathbb{1}_{x_i > y_j}}{mn} \tag{3.0.3}$$

In the formula above, m is the number of positive samples, n is the number of negative samples, mn is the number of all possible pairs, and $\mathbb{1}_{x_i > y_j}$ is a boolean identity function, which yields 1 if and only if a score of a positive sample x_i is greater than a score of a negative sample y_j. It is a probabilistic interpretation of the Area Under The Curve Receiver Operating Characteristics (AUC-ROC) metric. Such an interpretation suggests using AUC-ROC metric for classifier evaluation. It looks attractive because it does not require a threshold, which defines a mapping from the predicted probability of the class to a discrete class label.

However, there are two known limitations for AUC-ROC:

- AUC-ROC metric exists only for binary problems. Indeed, there are three classes in the Picnic dataset. However, that problem can be mitigated when there is an ordered relationship between the classes. In case of Picnic dataset that condition holds: HC > AD > SD. We can decompose it into two binary problems by evaluating the ordering of combined AD > SD versus HC group and then evaluating combined HC > AD groups versus AD group.
- AUC-ROC is not descriptive in cases where classes are imbalanced. That point is not obvious from (3.0.3). In reality, the AUC-ROC metric is an area under the Receiver Operating Characteristics (ROC) curve. Each point on that curve is defined by

$$\frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{3.0.4}$$

where True positive (TP) is the number of correctly predicted positive samples, False negative (FN) is the number of positive samples which are predicted as negative, True negative (TN) - a number of correctly predicted negative samples and False positive (FP) - a number of negative samples predicted as positive. That formula is not sensitive to the ratio of positive and negative samples since their counts are in separate terms of eq.3.0.4 and proportional increase of TN and FP counts cancel itself. Fig.3.1 illustrates that effect: ROC remained the same when the number of negative samples increased 10 times - from 160 to 16000, while the precision of prediction dropped from 0.76 to 0.24. Such a property of AUC-ROC makes it not indicative when applied for the heavily imbalanced MACS dataset.

An alternative for AUC-ROC is F score. Contrarily to AUC-ROC, it requires a pre-defined threshold if classifier predicts probability of the class rather than class itself. However, the main advantage of F score its independence for from the balance of the data. Its value remains invariant when the number of TN samples changes if all other confusion matrix entries remain the same[**45**]. Indeed, F score is defined as:

$$F_\beta = (1+\beta)\frac{precision * recall}{precision + recall} \tag{3.0.5}$$

where

$$precision = \frac{TP}{TP+FP} \tag{3.0.6}$$

$$recall = \frac{TP}{TP+FN} \tag{3.0.7}$$

and β defines a desired balance between precision and recall. $F1$ metric is a special case of F_β metric where $\beta = 1$.

None of its components includes TN values.

Mainly due to reliability of F score in the context of imbalanced data, we have adopted that measure for evaluation of our experiments.

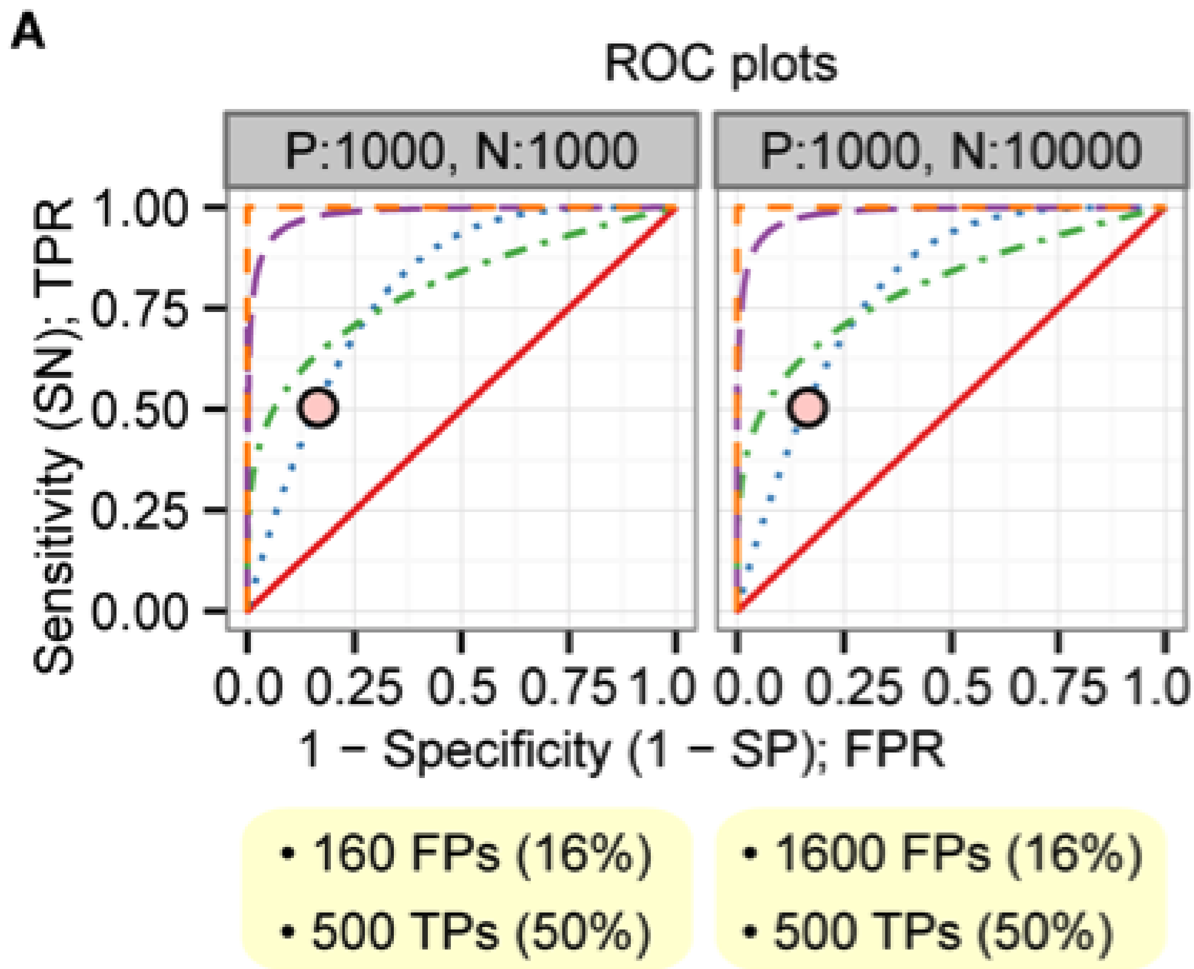


Fig. 3.1. ROC curves for balanced and imbalanced data

Five curves represent five different performance levels: Random (Rand; red), Poor early retrieval (ER-; blue), Good early retrieval (ER+; green), Excellent (Excel; purple), and Perfect (Perf; orange). A point on the left plot corresponds to 500 TP predictions (i.e., 50% of all positive samples) and 160 FP predictions (i.e. 16% of all negative samples), which makes precision score 0.76. The same point on the right side corresponds to 1600 FP makes precision score 0.24. *Images from Saito T., Rehmsmeier M. "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets."*[**44**]

Chapter 4

Experiments and analysis

We conducted our experiments with three different objectives:

- validate the performance of the engineered word-vector features
- compare different classification algorithms with engineered features (fixed feature set)
- experiment with the models, which learn features from the word-vector space

4.1. Engineered word-vector features

We have described the engineered word-vector features earlier in section 2.4, where we analyzed their separation power basing on a purely statistical approach. The feature selection algorithm repeatedly selected these features 3.0.1.3 in all datasets and all cross-validation splits of data. However, the most important examination - their performance in combination with all available features - was missing. In this chapter, we closely examine these features from a perspective of the classifier. For each combination of the dataset and a relevant word-vector model (see section 2.2), we train a FNN with various policies of feature selection:

- with any features from the full list of engineered features, specifying only the number of features that the feature selection algorithm has to select
- the same as above, but masking word-vector features from the classifier after feature selection (that prevents feature selection algorithm from selecting other features in place of masked) and retraining the classifier.
- train the classifier exclusively with word-vector features.

Since the training algorithm for ANN is not deterministic, we repeat each experiment 10 times and average metrics from all runs. We report a macro-$F1$ score and its standard deviation.

4.1.1. Implementation of FNN

For our experiments with engineered features, we implemented a Self-Normalizing Neural Network (SNN)[**46**], a variant of FNN with self-normalizing activations. We found empirically that for all datasets the best results are obtained when the network has only one hidden layer, which has a size of about one half of the dimensionality of the input.

For us, an essential factor in the configuration of the classifier was to avoid a linear classifier even if such a classifier potentially has a comparable performance with a more complex model. Since we focus on the analysis of particular features, our priority was to avoid the XOR situation - a mutual exclusive combination of two features, which is not linearly separable. That problem can potentially degrade performance of a linear classifier. In such a situation, removing one of the features could cause an improvement, which can be falsely attributed to a mal-behavior of the removed feature.

On the other side, deeper networks tend to have a higher variance, especially when trained on small datasets. In turn, the variance of results makes it more challenging to assess the impact of individual features. For that reason, we find that a single hidden layer is the right balance.

4.1.2. Training and validation

We have tuned hyperparameters of the model with HpBandSter, a Python package that implements Hyperband[**47**] - a bandit-based approach to hyperparameters optimization.

We have splat Picnic and DementiaBank datasets in folds of equal size in such a way that each fold would have the same proportion of the classes. In the case of the MACS dataset, we used its intrinsic partition into scenes - each scene became a fold.
We had held out one fold for the evaluation (test set), one fold for the validation during the training, and the rest folds we used for the training. We reported the average result obtained on all test sets.

4.1.3. Ovservations

Macro-$F1$ metrics of our model with 50 features on the DementiaBank dataset are similar to the metrics (up to 74%) achieved by Orimaye et al.[**8**] on the same data with only lexicosyntactic features. However, our model did not use any syntactic features derived from Context-Free Grammar (CFG).

We have observed an improvement about $0.5 - 1\%$ in $F1$ score in DementiaBank dataset (table 4.1) and over 1% in MACS dataset (table 4.2) with all word-vector models except an English CBOW model without sub-word info trained on Common Crawl. We cannot definitely state that there is an improvement in Picnic dataset due to a high variance of results, where standard deviation is about 1.5% and often significantly exceeds the difference between scores (table 4.3).

		DB-CBOW	DB-CBOW-CC	DB-CBOW-CCs	DB-SG
70 feat	no mask	0.733 ± 0.005	0.717 ± 0.008	0.732 ± 0.006	0.731 ± 0.009
	masked ID	0.734 ± 0.006	0.723 ± 0.005	0.728 ± 0.001	0.731 ± 0.004
	masked WVN	0.725 ± 0.006	0.725 ± 0.005	0.724 ± 0.008	0.723 ± 0.005
	masked WVN, ID	0.729 ± 0.004	0.724 ± 0.005	0.723 ± 0.006	0.728 ± 0.009
50 feat	no mask	0.737 ± 0.003	0.722 ± 0.007	0.727 ± 0.003	0.722 ± 0.005
	masked ID	0.734 ± 0.006	0.719 ± 0.005	0.728 ± 0.005	0.723 ± 0.006
	masked WVN	0.728 ± 0.006	0.727 ± 0.004	0.725 ± 0.002	0.723 ± 0.004
	masked WVN, ID	0.728 ± 0.003	0.727 ± 0.005	0.721 ± 0.003	0.723 ± 0.003
WVN, ID only		0.640 ± 0.007	0.551 ± 0.008	0.624 ± 0.007	0.564 ± 0.008

Table 4.1. DementiaBank dataset macro-$F1$ metrics

SNN with one hidden layer was trained with CBOW and SG word-vector models and various policies of feature selection. *ID* denotes ID feature, *WVN* denotes word-vector norm features for nouns, verbs, adverbs, adjectives and all alphabetical words (see section 2.4.2).

		MACS-CBOW	MACS-SG
top 60	no mask	0.935 ± 0.001	0.922 ± 0.002
	masked ID	0.929 ± 0.002	0.917 ± 0.001
	masked WVN	0.922 ± 0.002	0.921 ± 0.001
	masked WVN, ID	0.921 ± 0.001	0.916 ± 0.001
top 50	no mask	0.932 ± 0.001	0.930 ± 0.001
	masked ID	0.932 ± 0.001	0.926 ± 0.001
	masked WVN	0.920 ± 0.001	0.927 ± 0.002
	masked WVN, ID	0.925 ± 0.001	0.926 ± 0.002
top 40	no mask	0.924 ± 0.001	0.916 ± 0.002
	masked ID	0.925 ± 0.002	0.918 ± 0.001
	masked WVN	0.919 ± 0.001	0.917 ± 0.003
	masked WVN, ID	0.920 ± 0.002	0.917 ± 0.001
WVN, ID only		0.731 ± 0.003	0.718 ± 0.003

Table 4.2. MACS dataset macro-$F1$ metrics

SNN was trained with CBOW and SG word-vector models and various policies of feature selection. *ID* denotes ID feature, *WVN* denotes word-vector norm features for nouns, verbs, adverbs, adjectives and all alphabetical words (see section 2.4.2).

However, we have observed that word-vector features alone can discriminate between the SD group and the (AD and HC) groups. A model with only 6 features achieved a score of 0.553 in a 3-class problem. That score is very similar to the score achieved by the model with 70 features.

		Picnic-CBOW	Picnic-SG
top 70	no mask	0.551 ± 0.015	0.539 ± 0.018
	masked ID	0.563 ± 0.015	0.536 ± 0.015
	masked WVN	0.535 ± 0.013	0.535 ± 0.014
	masked WVN, ID	0.549 ± 0.018	0.548 ± 0.016
top 50	no mask	0.562 ± 0.023	0.522 ± 0.015
	masked ID	0.546 ± 0.018	0.508 ± 0.020
	masked WVN	0.519 ± 0.015	0.529 ± 0.010
	masked WVN, ID	0.513 ± 0.011	0.525 ± 0.012
WVN, ID only		0.553 ± 0.023	0.418 ± 0.019

Table 4.3. Picnic dataset macro-$F1$ metrics (cross-validation)

SNN with 1 hidden layer was trained with CBOW and SG word-vector models and various policies of feature selection. *ID* denotes ID feature, *WVN* denotes word-vector norm features for nouns, verbs, adverbs, adjectives and all alphabetical words (see section 2.4.2).

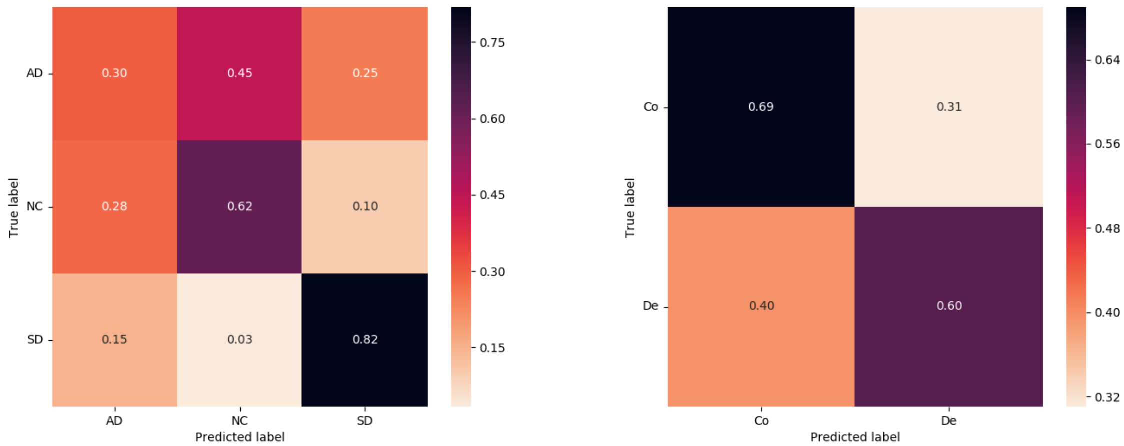
Confusion matrix resulted in that experiment (fig. 4.1a) showed that word-vector features predicted SD correctly in 82% of the cases (averaging between all folds in cross-validation). The most significant confusion was in the AD group, which semantically lies between HC and SD groups. We can see that the model could predict that group correctly only in 30% of cases. 45% of AD participants were classified as HC and the other 25% as SD. The HC group was predicted correctly in 62% of the cases, and its most significant confusion was with the AD group - 28%. Only 10% of that group were predicted as SD group. Such an observation of the performance is consistent with the statistical analysis of ID value distribution. Fig. 2.1a shows that distributions of ID scores for HC and AD groups are much similar to each other than to SD group.

Statistical analysis of feature distribution also explains why the CBOW variant of the model trained with only word-vector features performed significantly better than its SG counterpart. Fig. 2.5a, which corresponds to SG model, shows substantially less pronounced differences between the groups than fig. 2.5c, which corresponds to CBOW model.

The same experiment with the DementiaBank dataset (fig. 4.1b) had less noticeable results. Only one CBOW model achieved results that are statistically better than random. Word-vector features were able to classify correctly 60% of the Dementia group and 69% of the HC. Again, as in Picnic case, slightly better performance of CBOW model comparing to other word-vector models can be attributed to more pronounced differences between the groups (comparing fig. 2.1b vs other plots in 2.2, and fig. 2.4b vs other plots in 2.4)

4.2. Deep Learning methods

These methods provide an alternative to feature engineering - they learn features rather than engineer them. The main focus has shifted to the design and training models. In



(a) Picnic **(b)** DementiaBank

Fig. 4.1. Confusion matrix of Picnic and DementiaBank datasets

The main diagonal (top left to bottom right) indicates correct predictions, the rest of the cells are the errors. Saturation of the color indicates the number of samples in each group - a higher saturation corresponds to a higher population. Numbers in the cells show distribution of predicted classes for each of the ground truth classes. *NC* denotes HC group. SNN with 1 hidden layer trained exclusively on word-vector features obtained from CBOW model. Results show that the classifier has a remarkable separation between SD group and the rest, however it struggled to predict correctly AD group, which is between HC ans AD

this section, we will not discuss features anymore but focus on the analysis of different architectures and compare their performances. At the same time, that will provide a reference point for evaluating the results achieved by feature-engineering methods.

4.2.1. Self-normalizing convolutional network

We have implemented a Self Normalizing Convolutional Neural Network (SCNN) model as it was described by [**48**]. It is a small shallow network with 3 parallel convolutional branches, which are responsible for the reading of 3-, 4- and 5-grams. The concatenation layer collects their outputs, and a fully connected dense layer predicts the class (softmax regression). Figure 4.2 visualizes this network.

4.2.2. CNN-Incept

An Inception module (fig. 4.3) was originally proposed in GoogLeNet[**49**] for the ImageNet competition in 2014.

We have adopted this design to NLP by replacing two-dimensional convolutional filters with one-dimensional ones and increased the depth of the input from 3 channels (RGB) to 300 - the size of word-vector. With this modified module, we created a network of the shape

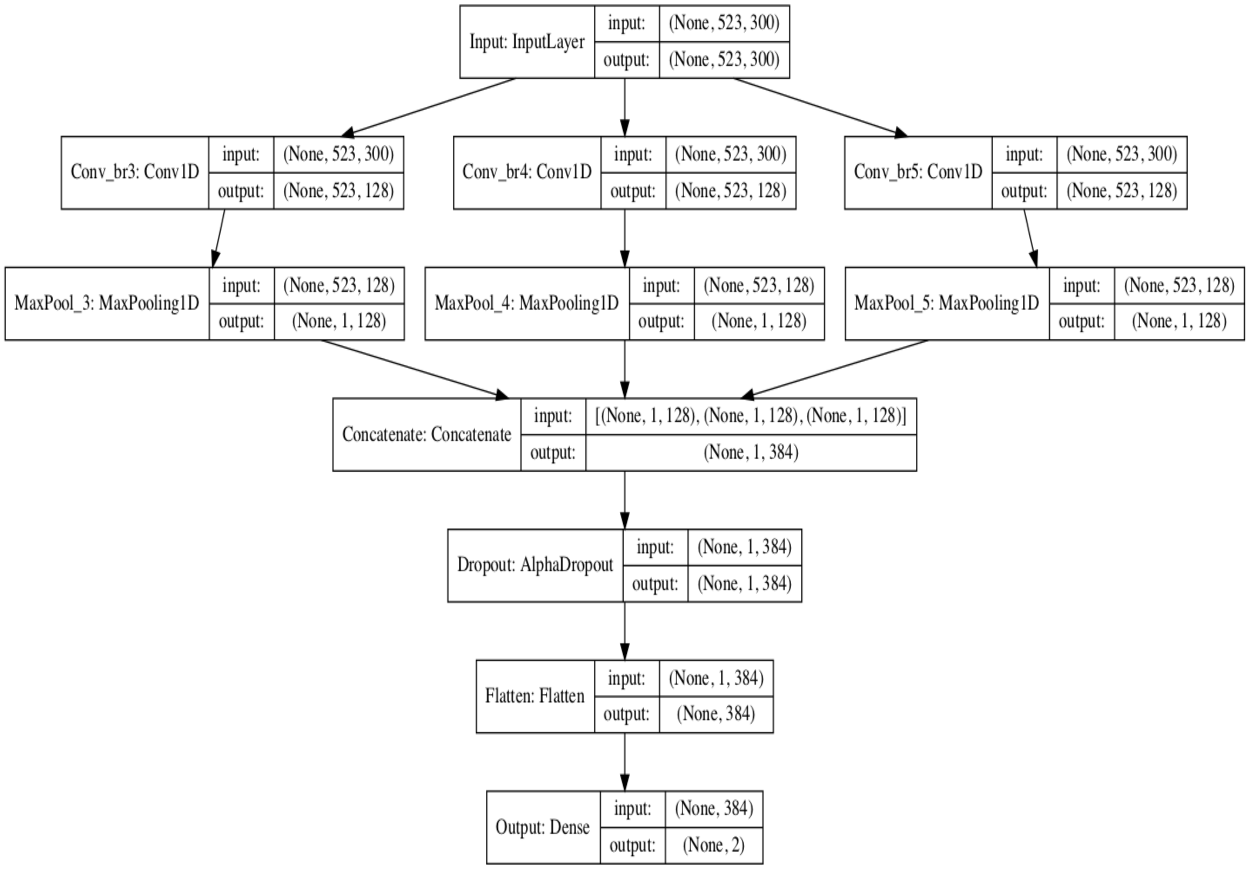


Fig. 4.2. Self-normalizing Convolutional Neural Network (SCNN)

very similar to SCNN (fig. 4.4). Both networks have a module of parallel branches with convolutional layers and a dense feed-forward layer on top. However, the critical difference between these two networks is an additional convolutional layer with the convolution of size 1. That layer precedes the main convolutional layer in each of the branches and causes a significant compression of the input depth. Indeed, a single convolution step with the filer of size 1 outputs a scalar for each position in the text, effectively transforming a 300-dimensional input (the size of word-vector) to a 1-dimensional output. In our implementation, there are 128 parallel filters. Together, they output a 128-dimensional vector, i.e., compressing input's depth more than twice. Beyond the benefit of compressing, that layer, due to its non-linear activations, increases solution space of the network.

Tables 4.4 and 4.5 show that both classifiers have a very similar performance. However, SCNN network has 461,954 parameters, and CNN-Incept has only 145,082.

	DB-CBOW	DB-SG
SCNN	0.742 ± 0.003	0.779 ± 0.005
CNN-Incept	0.742 ± 0.007	0.781 ± 0.005

Table 4.4. DementiaBank macro-$F1$ metrics - CNN models

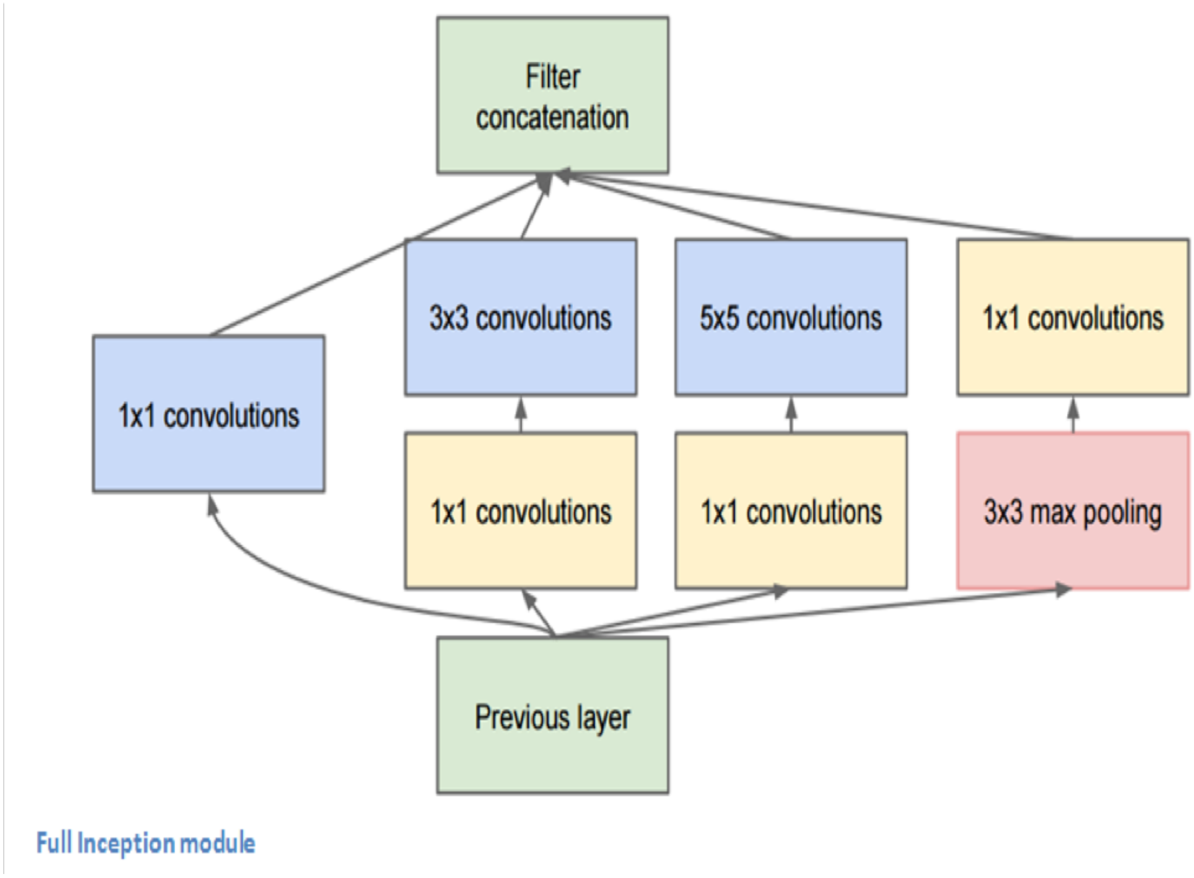


Fig. 4.3. Inception module with dimension reduction

Image from "Going Deeper with Convolutions"[**49**]

	DB-CBOW	DB-SG
SCNN	0.979 ± 0.002	0.970 ± 0.003
CNN-Incept	0.977 ± 0.001	0.971 ± 0.005

Table 4.5. MACS macro-$F1$ metrics - CNN models

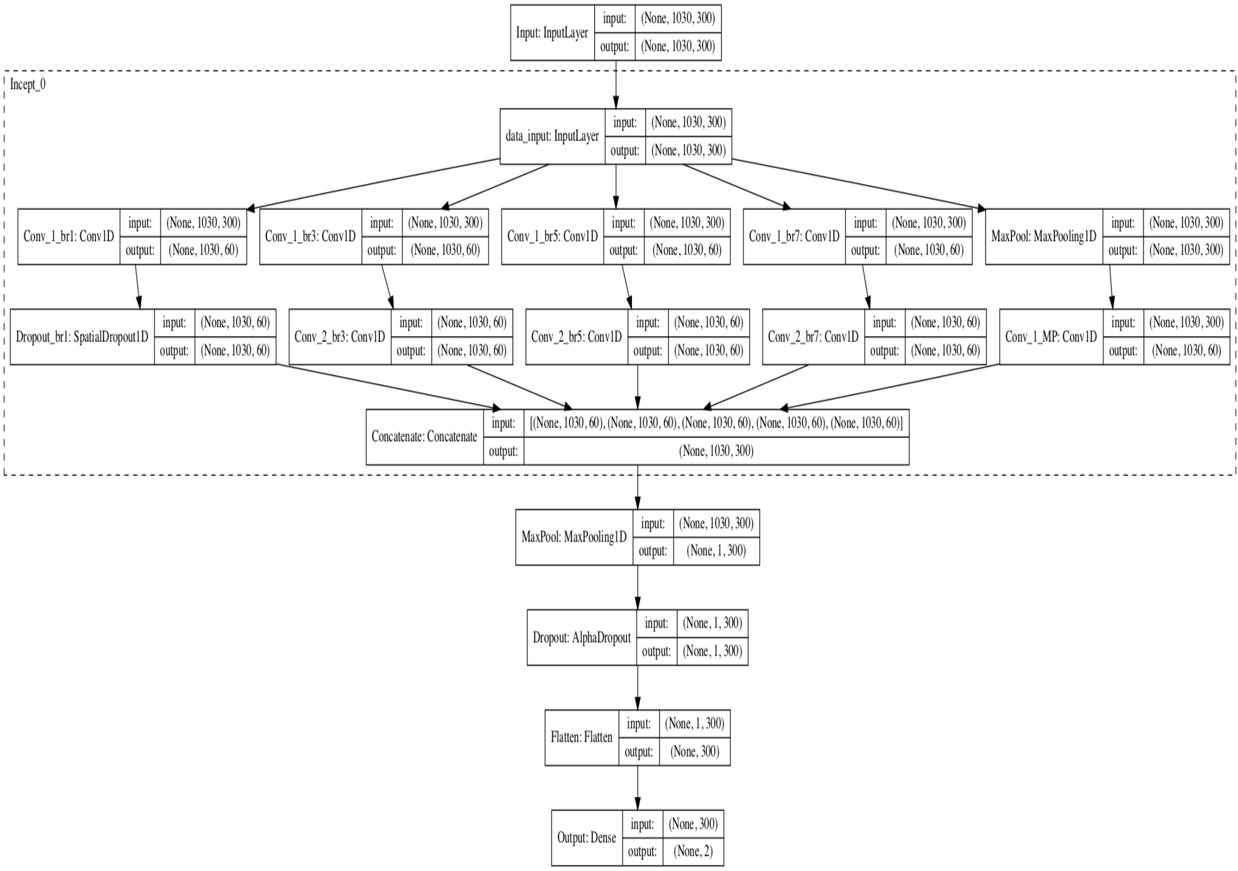


Fig. 4.4. An actual configuration of Self-normalizing Convolutional Neural Network with Inception module (SCNN-Incept)

Chapter 5

Contributions and future work

We have proposed new language- and topic-independent methods of measuring the number of expressed ideas and their specificness. Our experiments demonstrated that the scores derived from the word-vector model are valuable features that improve classification results in all datasets.

We have successfully applied those methods in binary and trinary classification scenarios, in single-topic datasets (DementiaBank and Picnic), and in the multi-topic MACS corpus.

In our experiments with DementiaBank, we have reproduced results reported by Orimaye et al.[**8**] - a state-of-the-art model based solely on lexicosyntactic features - though our model did not use features derived from CFG in order to eliminate feature dependency on English grammar and be able to deploy the same set of features for French. We have compensated occurred loss of information with the language-agnostic semantic features derived from the word embeddings.

Though our semantic features are weaker than the features proposed by Yancheva et al. [**11**], they allow semantic evaluation in topic-free datasets and in multi-topic datasets like MACS. We believe that diversification of topics in a topic-constrained dataset is a more efficient way of increasing data size than an increase of data on the same topic (a similar point was made by Fraser et al.[**15**] about multilingual data). Future work would be to show that.

We also should further develop an idea of reducing the language dependency of features. We selected our lexicosyntactic features with English and French in mind. Rentoumi et al.[**13**] used a similar approach for English and Greek. Is it possible to extend to other languages? How far? Can we generalize to different language families?

References

[1] Dementia numbers in Canada.

[2] Kaplan E. Barresi B.. Goodglass, H. *Boston Diagnostic Aphasia Examination-Third Edition (BDAE-3)*. 2000.

[3] J. T. Becker, F. Boller, O. L. Lopez, J. Saxton, and K. L. McGonigle. The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology, 51(6), 585-594*, 1994.

[4] S. Freud. *Zur Psychopathologie des Alltagslebens: (über Vergessen, Versprechen, Vergreifebn. Aberglaube und Irrtum)*. Internationale Psychoanalytische Bibliothek. Internationaler Psychoanalytischer Verlag, 1920.

[5] Dunphy D. C. Smith M. S. Stone, P. J. *The general inquirer: A computer approach to content analysis.* M.I.T. Press., 1966.

[6] Sameer Singh. Quantitative classification of conversational language using artificial neural networks. *Aphasiology*, 11(9):829–844, 1997.

[7] R. S. Bucks, S. Singh, J. M. Cuerden, and G. K. Wilcock. Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance. *Aphasiology*, 14(1):71–91, 2000.

[8] Sylvester Olubolu Orimaye, Jojo Sze-Meng Wong, and Karen Jennifer Golden. Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 78–87, Baltimore, Maryland, USA, June 2014. Association for Computational Linguistics.

[9] Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, page 423–430, USA, 2003. Association for Computational Linguistics.

[10] Kathleen C Fraser, Jed A Meltzer, and Frank Rudzicz. Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2):407–422, 2016.

[11] Maria Yancheva and Frank Rudzicz. Vector-space topic models for detecting Alzheimer's disease. In *ACL*, 2016.

[12] Kairit Sirts, Olivier Piguet, and Mark Johnson. Idea density for predicting Alzheimer's disease from transcribed speech. *CoRR*, abs/1706.04473, 2017.

[13] Vassiliki Rentoumi, George Paliouras, Dimitra Arfani, Katerina Fragkopoulou, Spyridoula Varlokosta, and Peter Garrard. Automatic detection of linguistic indicators as a means of early prediction of alzheimer's and of related dementias: A cross-linguistics analysis. *Alzheimer's & Dementia*, 14:P1302, 07 2018.

[14] Edith Kaplan Goodglass, Harold and Barbara Barresi. *Boston Diagnostic Aphasia Examination*. Lippincott Williams & Wilkins, Baltimore, MD, 3 edition, 2001.

[15] Kathleen Fraser, Kristina Lundholm Fors, and Dimitrios Kokkinakis. Multilingual word embeddings for the assessment of narrative speech in mild cognitive impairment. *Computer Speech & Language*, 53, 08 2018.

[16] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.

[17] Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2979–2984, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[18] Diane E. Beals. The childes project: Tools for analyzing talk: Vol. 1. transcription format and programs; vol. 2. the database (3rd ed.). b. macwhinney. mahwah, nj: Erlbaum, 2000. pp. 366 (vol. 1); pp. 418 (vol. 2). *Applied Psycholinguistics*, 23(2):304–306, 2002.

[19] Kertesz A. Shewan, C. M. Reliability and validity characteristics of the Western Aphasia battery (wab). *Journal of Speech & Hearing Disorders*, 45(3):308–324, 1980.

[20] M. Montembeault, S.M. Brambati, S. Joubert, M. Boukadi, M. Chapleau, R.Jr. Laforce, M.A. Wilson, J. Macoir, and I. Rouleau. Naming unique entities in the semantic variant of primary progressive aphasia and alzheimer's disease: Towards a better understanding of the semantic impairment. *Neuropsychologia*, 95:11 – 20, 2017.

[21] Anna Zumbansen, Joséphine Frachon, Dorothée Quiquempois, Sylvie Hébert, and Alexander Thiel. Abstract tp148: The Montreal Assessment of Connected Speech offers good psychometric properties to monitor ecological language recovery in post-stroke aphasia. *Stroke*, 47, 2016.

[22] M F Folstein, S E Folstein, and P R McHugh. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*, 12(3):189–198, Nov 1975.

[23] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014.

[24] Peng Qi, Timothy Dozat, Yuhao Zhang, and Christopher D. Manning. Universal dependency parsing from scratch. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 160–170, Brussels, Belgium, October 2018. Association for Computational Linguistics.

[25] Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.

[26] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013.

[27] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. Advances in pre-training distributed word representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[28] Robyn Speer, Joshua Chin, Andrew Lin, Sara Jewett, and Lance Nathan. Luminosoinsight/wordfreq: v2.2, October 2018.

[29] Myrna F. Schwartz, Oscar S.m. Marin, and Eleanor M. Saffran. Dissociations of language function in dementia: A case study. *Brain and Language*, 7(3):277–306, 1979.

[30] Amit Almor, Daniel Kempler, Maryellen C. MacDonald, Elaine S. Andersen, and Lorraine K. Tyler. Why do alzheimer patients have difficulty with pronouns? working memory, semantics, and reference in comprehension and production in alzheimer's disease. *Brain and Language*, 1999.

[31] Gitit Kavé and Mira Goral. Word retrieval in picture descriptions produced by individuals with alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 38(9):958–966, 2016. PMID: 27171756.

[32] Michael A. Covington and Joe D. McFall. Cutting the gordian knot: The moving-average type–token ratio (mattr). *Journal of Quantitative Linguistics*, 17(2):94–100, 2010.

[33] Naama Friedmann. Speech production in broca ' s agrammatic aphasia : Syntactic tree pruning. 2004.

[34] Frank Burchert, Nadine Meißner, and Ria De Bleser. Production of non-canonical sentences in agrammatic aphasia: Limits in representation or rule application? *Brain and Language*, 104:170–179, 2008.

[35] A. Turner, E. Greene, Boulder. Institute for the Study of Intellectual Behavior University of Colorado, and American Psychological Association. *The Construction and Use of a Propositional Text Base*. Documents in psychology. American Psychological Association, 1977.

[36] S Ahmed, CA de Jager, A-M Haigh, and P Garrard. Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed alzheimer's disease. *Neuropsychology 27(1), 79-85*, 2013.

[37] Justin M. Aronoff, Laura M. Gonnerman, Amit Almor, Sudha Arunachalam, Daniel Kempler, and Elaine S. Andersen. Information content versus relational knowledge: semantic deficits in patients with alzheimer's disease. *Neuropsychologia*, 44 1:21–35, 2006.

[38] Hyon-Ah Yi, Peachie Moore, and Murray Grossman. Reversal of the concreteness effect for verbs in patients with semantic dementia. *Neuropsychology*, 21(1):9, 2007.

[39] Mikyong Kim and Cynthia K Thompson. Verb deficits in alzheimer's disease and agrammatism: Implications for lexical organization. *Brain and language*, 88(1):1–20, 2004.

[40] Adriaan M. J. Schakel and Benjamin J. Wilson. Measuring word significance using distributed representations of words. *CoRR*, abs/1508.02297, 2015.

[41] Sergey Bartunov, Dmitry Kondrashkin, Anton Osokin, and Dmitry P. Vetrov. Breaking sticks and ambiguities with adaptive skip-gram. In Arthur Gretton and Christian C. Robert, editors, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016, Cadiz, Spain, May 9-11, 2016*, volume 51 of *JMLR Workshop and Conference Proceedings*, pages 130–138. JMLR.org, 2016.

[42] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P Trevino, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. *arXiv preprint arXiv:1601.07996*, 2016.

[43] Feiping Nie, Heng Huang, Xiao Cai, and Chris Ding. Efficient and robust feature selection via joint L2,1-norms minimization. In *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 2*, NIPS'10, pages 1813–1821, USA, 2010. Curran Associates Inc.

[44] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *Plos One*, 10(3), Apr 2015.

[45] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, July 2009.

[46] Guenter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In *NIPS*, 2017.

[47] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(185):1–52, 2018.

[48] Avinash Madasu and Vijjini Anvesh Rao. Effectiveness of self normalizing neural networks for text classification. *ArXiv*, abs/1905.01338, 2019.

[49] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2014.

[50] Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

Appendix A

Word-vector norms

Lists of word-vector norms from various models, sorted by vector norms in ascending order

A.0.1. English CBOW models Common Crawl

See T. Mikolov et al[**27**] for the implementation details of these models

A.0.1.1. *English CBOW model Common Crawl with subword information*

('unfortunately', 0.37924102), ('apparently', 0.44693032), ('essentially', 0.4646008), ('fortunately', 0.46674728), ('particular', 0.46695006), ('neighborhood', 0.4717642), ('presumably', 0.48134735), ('obviously', 0.48179045), ('practically', 0.48207688), ('interesting', 0.48654154), ('perspective', 0.49185047), ('definitely', 0.49400023), ('supposedly', 0.5114017), ('everything', 0.5115667), ('otherwise', 0.5120028), ('description', 0.51303476), ('significance', 0.5133884), ('understand', 0.51683587), ('sometimes', 0.52297884), ('something', 0.52436996), ('certainly', 0.52447015), ('eventually', 0.527465), ('basically', 0.52813554), ('absentmindedly', 0.5286176), ('including', 0.5302935), ('instructions', 0.5304592), ('presuming', 0.53335965), ('completely', 0.5335946), ('considered', 0.5379829), ('background', 0.5381742), ('naturally', 0.53893995), ('underneath', 0.5392244), ('beginning', 0.54092), ('catastrophe', 0.544896), ('reference', 0.54498637), ('responsible', 0.54580754), ('encouraging', 0.5482303), ('grandmother', 0.5483914), ('indication', 0.55237395), ('forgetting', 0.5527692), ('connection', 0.55466837), ('obliterated', 0.5590712), ('evidently', 0.5606899), ('actually', 0.5608256), ('descriptive', 0.5629673), ('although', 0.5632526), ('different', 0.5637488), ('discovered', 0.5646451), ('difference', 0.5658845), ('housekeeping', 0.57078207), ('proceeding', 0.5711169), ('possible', 0.5713138), ('indicating', 0.57230866), ('possibly', 0.57293284), ('elsewhere', 0.5740826), ('beautiful', 0.5741009), ('activities', 0.57636213), ('distinguish', 0.57724154), ('permitting', 0.577277), ('addition', 0.5791804), ('instructing', 0.5793305), ('preoccupied', 0.58002263), ('interested', 0.581062), ('important', 0.5816941), ('somewhere', 0.5849502), ('distressing', 0.5901473), ('attempting', 0.59047747), ('meanwhile', 0.5924103), ('directions', 0.5947615), ('unconcerned', 0.59698796), ('expecting', 0.5978635), ('identifying', 0.5987296), ('supposed', 0.59892356), ('mentioned', 0.59894377), ('summertime', 0.5998814), ('criticizing', 0.60016114), ('admonishing', 0.6021986), ('probably', 0.60357094), ('incomplete', 0.60397094), ('consider', 0.6064052), ('starting', 0.6065456), ('present', 0.606647), ('distracted', 0.60688835), ('expression', 0.60722536), ('direction', 0.607507), ('nondescript', 0.607992), ('housekeeper', 0.60889316), ('because', 0.6090477), ('condition', 0.60914195), ('situation', 0.61004627), ('youngsters', 0.610336), ('conscious', 0.6107299), ('extension', 0.61281914), ('correctly', 0.61289084), ('overflowing', 0.61325), ('overlooking', 0.6138964), ('thought', 0.61398846), ('unconscious', 0.61433643), ('thinking', 0.61581945), ('assuming', 0.6166173), ('landscaping', 0.61696005), ('daydreaming', 0.6189315), ('another', 0.6190533), ('remember', 0.62064), ('selection', 0.6210094), ('certain', 0.62206095), ('though', 0.62221384), ('either', 0.62411445), ('youngster', 0.62458295), ('stretching', 0.6257667), ('straight', 0.62598145), ('finished', 0.6312772), ('neglecting', 0.63150495), ('happening', 0.63226265), ('outstretched', 0.63444483), ('instead', 0.6345894), ('indicates', 0.63503104), ('position', 0.63632387), ('together', 0.6366157), ('perhaps', 0.6388349), ('someplace', 0.6421725), ('personal', 0.6421821), ('snickering', 0.643605), ('anything', 0.6439479), ('backwards', 0.64402384), ('partially', 0.6444785), ('overturning', 0.6454221), ('building', 0.6475595), ('overrunning', 0.6484362), ('attentively', 0.65005004), ('whatever', 0.65069795), ('wondering', 0.6524826), ('disturbed', 0.6543045), ('terrible', 0.65542704), ('whispering', 0.6557252), ('precarious', 0.65693), ('standing', 0.6578416), ('difficult', 0.65798813), ('observed', 0.6601474), ('already', 0.6602966), ('stretches', 0.66033584), ('neglected', 0.6610783), ('abstractly', 0.66128045), ('cautioning', 0.66237885), ('decorating', 0.663359), ('dangerous', 0.66388977), ('signaling', 0.6639524), ('suppose', 0.66427004), ('presume', 0.6645394), ('happenings', 0.66506344), ('extended', 0.6655517), ('listening', 0.6666338), ('designate', 0.6669461), ('believe', 0.66741717), ('sprinkling', 0.66764086), ('oblivious', 0.6712931), ('happens', 0.67143047), ('upsetting', 0.6730031), ('according', 0.67399156), ('discharged', 0.67414945), ('anyway', 0.67420954), ('distinct', 0.67488194), ('activity', 0.6762544), ('counting', 0.67804575), ('opposite', 0.67846036), ('preparing', 0.68115276), ('appears', 0.68177235), ('sleeveless', 0.68533486), ('children', 0.68580216), ('working', 0.6861506), ('country', 0.686495), ('christmas', 0.68653333), ('correct', 0.6872481), ('interest', 0.6876274), ('slightly', 0.68766224), ('somewhat', 0.68835783), ('assisting', 0.68874323), ('ordinary', 0.6887591), ('mention', 0.6890089), ('whereas', 0.6892932), ('pleasant', 0.6893852), ('daughter', 0.68958014), ('receiving', 0.6896879), ('somebody', 0.6907642), ('finally', 0.69295824), ('struggle', 0.6937497), ('threesome', 0.69423103), ('unbalances', 0.6943708), ('course', 0.69597113), ('strictly', 0.6962344), ('anymore', 0.6962485), ('somehow', 0.6966525), ('nothing', 0.6973274), ('meantime', 0.6997621), ('impervious', 0.6997693), ('pictured', 0.70010257), ('insisted', 0.700438), ('dishwasher', 0.70112014), ('perplexed', 0.7014099), ('continue', 0.70141697), ('looking', 0.7018768), ('inattentive', 0.7020282), ('suspect', 0.70313656), ('fluttering', 0.7038705), ('splattered', 0.7040921), ('through', 0.7043873), ('evergreens', 0.70459414), ('breakfast', 0.7069496), ('sometime', 0.70739377), ('goodness', 0.707787), ('special', 0.707999), ('similar', 0.7081308), ('accident', 0.7083932), ('countertop', 0.7086381), ('realize', 0.7087108), ('process', 0.70887035), ('watching', 0.711038), ('backyard', 0.7113124), ('anyhow', 0.7113171), ('movement', 0.7117306), ('stumbling', 0.71175396), ('leading', 0.71270764), ('allowing', 0.713392), ('talking', 0.7135379), ('started', 0.7137297), ('further', 0.7141971), ('furiously', 0.71492493), ('project', 0.7156546), ('running', 0.71627986), ('indicate', 0.71716267), ('noticing', 0.71741605), ('picture', 0.7185647), ('attention', 0.7187314), ('setting', 0.7198235), ('deciding', 0.7211768), ('written', 0.7228928), ('alright', 0.7233997), ('implies', 0.7237328), ('worrying', 0.7242107), ('halfway', 0.7245038), ('overfilled', 0.72464013), ('pineapple', 0.72520113), ('others', 0.72529393), ('neighbor', 0.7258615), ('carrying', 0.7265936), ('sideways', 0.7278926), ('changing', 0.7279078), ('imagine', 0.7286394), ('touching', 0.72975326), ('stepping', 0.7303701), ('attentive', 0.73102266), ('describe', 0.7313685), ('accidents', 0.73159134), ('terribly', 0.73180574), ('pointing', 0.7318338), ('anywhere', 0.7318732), ('disaster', 0.73206264), ('darling', 0.7342243), ('windowsills', 0.73431796), ('typical', 0.7346404), ('average', 0.7354631), ('dreaming', 0.7357395), ('strange', 0.7358148), ('section', 0.7360111), ('earlier', 0.73606396), ('someone', 0.7365559), ('really', 0.7372851), ('housewife', 0.7374888), ('missing', 0.73804295), ('except', 0.73809934), ('second', 0.73819745), ('enjoying', 0.7388873), ('engrossed', 0.7389055), ('cascading', 0.7389415), ('ignoring', 0.74030477), ('trouble', 0.74058443), ('grandson', 0.74122095), ('passing', 0.7422716), ('reason', 0.7424996), ('breaking', 0.7425563), ('pleasures', 0.7428624), ('writing', 0.74322265), ('conflict', 0.7432845), ('throwing', 0.74368334), ('gesturing', 0.74409294), ('outside', 0.74410224), ('outdoors', 0.7459401), ('affected', 0.74675405), ('machine', 0.7476699), ('dropping', 0.74769866), ('overflowed', 0.7479014), ('airplane', 0.74845415), ('adjacent', 0.74884737), ('quickly', 0.7490714), ('telling', 0.7493286), ('putting', 0.7499041), ('sneaking', 0.7499767), ('problem', 0.7505295), ('nineteen', 0.7507148), ('things', 0.751297), ('partial', 0.75166297), ('daylight', 0.7518294), ('little', 0.75212765), ('details', 0.75262344), ('problems', 0.7530127), ('painting', 0.7532718), ('morning', 0.75361013), ('person', 0.7536677), ('turning', 0.75414133), ('shoulder', 0.7552401), ('labelled', 0.75529575), ('leaving', 0.75561076), ('variety', 0.75602776), ('reading', 0.75603616), ('showing', 0.7563084), ('hazardous', 0.7564243), ('artistic', 0.75673836), ('lopsided', 0.7574939), ('opposed', 0.7589586), ('silencing', 0.75934416), ('identify', 0.7598118), ('slipping', 0.76014954), ('copyright', 0.76030886), ('getting', 0.7604505), ('heading', 0.7610695), ('realizes', 0.7611373), ('calamity', 0.7614957), ('splashing', 0.7621962), ('catching', 0.7625221), ('saying', 0.7627116), ('removing', 0.76492447), ('shrubbery', 0.76494765), ('sentence', 0.7654494), ('observant', 0.7655663), ('blooming', 0.7655732), ('counter', 0.76626676), ('kitchen', 0.7672736), ('himself', 0.7673845), ('almost', 0.76783067), ('venture', 0.7680946), ('husband', 0.7682419), ('cupboards', 0.76851994), ('family', 0.7688596), ('sidewalk', 0.76926625), ('visible', 0.7704162), ('noticed', 0.7707799), ('spreading', 0.7707968), ('rather', 0.7712191), ('playing', 0.77219176), ('stopped', 0.772686), ('inside', 0.77332765), ('brother', 0.7735467), ('overboard', 0.7737295), ('unusual', 0.77386254), ('reaching', 0.7744445), ('collapse', 0.7747354), ('flexible', 0.77506393), ('portion', 0.77560914), ('handful', 0.77638304), ('careless', 0.7772592), ('shortly', 0.7776569), ('pointed', 0.7797032), ('petroleum', 0.7809854), ('cleaning', 0.7811384), ('laughing', 0.78117996), ('always', 0.7812257), ('younger', 0.78141975), ('matter', 0.7815268), ('carpentry', 0.78230923), ('climbing', 0.78309304), ('giggling', 0.78374404), ('resting', 0.7839505), ('theirs', 0.78395396), ('letting', 0.78455913), ('overflows', 0.78494966), ('whoever', 0.7861033), ('removed', 0.7861436), ('immersed', 0.7866583), ('actions', 0.7882047), ('moment', 0.7883403), ('mostly', 0.7885141), ('survived', 0.78915215), ('grabbing', 0.7896265), ('outfitted', 0.79047453), ('allowed', 0.79183614), ('quietly', 0.7918452), ('holding', 0.7920984), ('again', 0.79216796), ('pattern', 0.7924425), ('wanting', 0.7934567), ('reacting', 0.79371154), ('peaceful', 0.79442406), ('persons', 0.7945582), ('draining', 0.79456407), ('stockings', 0.79486203), ('means', 0.79492307), ('tumbling', 0.7949567), ('place', 0.79500765), ('rolling', 0.7951991), ('coming', 0.7952896), ('rocking', 0.7954602), ('moments', 0.7954652), ('frowning', 0.79591745), ('around', 0.7959427), ('receive', 0.79653394), ('herself', 0.79680383), ('sharing', 0.79778516), ('better', 0.79823107), ('driveway', 0.79873085), ('snitching', 0.79899037), ('fricking', 0.79943746), ('number', 0.80071807), ('staring', 0.8013463), ('crawling', 0.80139565), ('single', 0.8017624), ('squirting', 0.8018149), ('stepped', 0.80199975), ('enough', 0.8021036), ('crashing', 0.802527), ('unaware', 0.80291975), ('growing', 0.80300444), ('recessed', 0.8030056), ('motioning', 0.80439395), ('point', 0.8046729), ('teetering', 0.8047384), ('myself', 0.80495113), ('mistake', 0.80620915), ('should', 0.80724597), ('overflown', 0.8074293), ('overflow', 0.80873895), ('stealing', 0.8092476), ('people', 0.80948603), ('sneakers', 0.80950433), ('billowing', 0.8099476), ('cupboard', 0.80999535), ('happen', 0.81048816), ('called', 0.81064326), ('against', 0.81156677), ('mistakes', 0.8117004), ('having', 0.8118889), ('flooding', 0.8120606), ('dripping', 0.81213415), ('figure', 0.8132489), ('ground', 0.8134994), ('healthy', 0.8139854), ('there', 0.8160459), ('distract', 0.8160482), ('managed', 0.816489), ('careful', 0.816544), ('walking', 0.8166341), ('cabinets', 0.8171261), ('whether', 0.817206), ('school', 0.8173237), ('lovely', 0.81745726), ('finish', 0.81891006), ('waiting', 0.81946206), ('extent', 0.82081664), ('wobbling', 0.82109135), ('whisper', 0.8216696), ('dropped', 0.8222257), ('normal', 0.82241875), ('beside', 0.8225988), ('trousers', 0.82316387), ('strapped', 0.82322013), ('counts', 0.823512), ('action', 0.8235971), ('which', 0.82548803), ('handles', 0.8264248), ('slender', 0.8286433), ('minutes', 0.82970077), ('checked', 0.83001816), ('miniskirt', 0.8301259), ('hanging', 0.83060116), ('sitting', 0.8307406), ('right', 0.8311832), ('before', 0.8312473), ('curtains', 0.831277), ('tripodal', 0.8315751), ('notice', 0.8316453), ('turned', 0.83215594), ('diamonds', 0.8322723), ('anybody', 0.8323691), ('stopper', 0.83272356), ('dishwater', 0.83349407), ('fountain', 0.8336447), ('animals', 0.8337015), ('wonder', 0.83377236), ('balance', 0.8338045), ('handing', 0.83430284), ('active', 0.8345029), ('expect', 0.83553976), ('charged', 0.8357346), ('culprit', 0.8369345), ('double', 0.8372341), ('slopping', 0.83870035), ('splashed', 0.83911085), ('falling', 0.8392017), ('forget', 0.8397881), ('amounts', 0.8402332), ('wearing', 0.84049773), ('while', 0.84155697), ('helping', 0.84218514), ('assume', 0.84228396), ('first', 0.8423735), ('excited', 0.8424732), ('pathway', 0.84258366), ('windows', 0.84260815), ('making', 0.84270906), ('smiling', 0.84344536), ('cutting', 0.8434799), ('picking', 0.8439051), ('mother', 0.8442042), ('fiddling', 0.84447336), ('strong', 0.8450146), ('father', 0.84546626), ('hitting', 0.8466477), ('appear', 0.847635), ('begging', 0.84790075), ('singing', 0.8479773), ('letters', 0.84812534), ('couple', 0.8492232), ('salient', 0.84978795), ('clothes', 0.85163707), ('between', 0.85196424), ('center', 0.85329825), ('depends', 0.8535286), ('longer', 0.8537313), ('friend', 0.8537339), ('wanted', 0.85404867), ('slanting', 0.8543159), ('silence', 0.8546967), ('think', 0.8550298), ('starts', 0.85511625), ('square', 0.8561739), ('sweater', 0.8568104), ('still', 0.8570909), ('pompadour', 0.85740185), ('bordered', 0.8579058), ('outdoor', 0.858059), ('whole', 0.8584284), ('dumping', 0.85846764), ('leaning', 0.85873467), ('shocked', 0.8588832), ('weather', 0.8595819), ('neither', 0.85958266), ('donating', 0.8599591), ('yelling', 0.860248), ('stacked', 0.8602554), ('shining', 0.8604699), ('causing', 0.8605241), ('thing', 0.861333), ('seeing', 0.86149514), ('please', 0.8621255), ('other', 0.8622579), ('unless', 0.86235553), ('sister', 0.86255413), ('bottom', 0.8632316), ('solidly', 0.8633557), ('moving', 0.8637664), ('filling', 0.8643577), ('house', 0.86450475), ('slipped', 0.8648296), ('flowing', 0.8652536), ('asking', 0.8659207), ('missed', 0.86614835), ('corner', 0.86628973), ('thicker', 0.8663202), ('sudden', 0.8665419), ('slippers', 0.8666179), ('moisture', 0.8682308), ('behind', 0.86831135), ('cooking', 0.86836845), ('overrun', 0.8691638), ('hahaha', 0.86928177), ('towards', 0.86973584), ('hundred', 0.87081885), ('summer', 0.87120205), ('memorize', 0.8715809), ('stuff', 0.8717215), ('start', 0.87220633), ('least', 0.8722781), ('minute', 0.87248194), ('scenery', 0.87280047), ('giving', 0.8729341), ('flowers', 0.8733345), ('draperies', 0.8733526), ('affair', 0.8742917), ('knocked', 0.8753597), ('found', 0.8769682), ('world', 0.87716526), ('signing', 0.8771773), ('bumping', 0.8772824), ('since', 0.87738824), ('percent', 0.87767535), ('reached', 0.8785287), ('pretty', 0.8785883), ('cabinet', 0.8801636), ('crossed', 0.88040435), ('taking', 0.88088477), ('jumping', 0.88097095), ('slanted', 0.8812841), ('signal', 0.8816844), ('forgot', 0.88179934), ('spilling', 0.8824104), ('great', 0.88262004), ('bother', 0.8829291), ('silent', 0.88331115), ('spring', 0.8836108), ('foliage', 0.88370055), ('hardly', 0.88370246), ('leaves', 0.8839638), ('messed', 0.88455313), ('overturns', 0.8850087), ('glimpse', 0.88533086), ('handle', 0.88676447), ('gadgets', 0.8873029), ('secret', 0.8873801), ('awfully', 0.88755095), ('picked', 0.887607), ('fingers', 0.8876757), ('sounds', 0.88793343), ('beyond', 0.88847435), ('platter', 0.88858706), ('pouring', 0.88901573), ('keeping', 0.8901261), ('middle', 0.8902316), ('living', 0.89033896), ('crooked', 0.89067715), ('looked', 0.8912211), ('walkway', 0.8917426), ('maybe', 0.8918831), ('close', 0.8927765), ('lengths', 0.8940728), ('window', 0.89565486), ('handed', 0.8960105), ('dinner', 0.8962617), ('damaged', 0.89632595), ('times', 0.8966847), ('trying', 0.89868027), ('smashed', 0.8998579), ('planted', 0.9001321), ('garden', 0.90132433), ('reaches', 0.90160567), ('dressed', 0.9016737), ('blowing', 0.9017059), ('broken', 0.90184397), ('state', 0.9019006), ('shaking', 0.90209013), ('coaxing', 0.90251476), ('toppling', 0.9029119), ('torrent', 0.9032936), ('raining', 0.9035775), ('winter', 0.90482074), ('going', 0.904829), ('closed', 0.90553695), ('fellow', 0.9055969), ('upside', 0.9057069), ('guess', 0.9057215), ('protrude', 0.9065765), ('might', 0.90692955), ('hidden', 0.9079343), ('slight', 0.90836895), ('gravity', 0.9088404), ('gentle', 0.90911037), ('meant', 0.90956825), ('often', 0.90960836), ('after', 0.90999895), ('drooping', 0.91060144), ('remain', 0.9106795), ('washing', 0.9112295), ('raising', 0.9114217), ('junior', 0.9117226), ('define', 0.9120723), ('leaking', 0.91239184), ('front', 0.9126677), ('stand', 0.91321814), ('hoping', 0.9133958), ('seemed', 0.91388357), ('rolled', 0.9144463), ('pieces', 0.9144959), ('clear', 0.9149009), ('seventy', 0.9152437), ('then', 0.9153143), ('jogging', 0.91585237), ('excuse', 0.9162439), ('soaking', 0.9170361), ('sticks', 0.91739213), ('round', 0.9176374), ('small', 0.91764843), ('fairly', 0.9179715), ('nuthouse', 0.9179886), ('short', 0.9180852), ('comes', 0.91844124), ('tipping', 0.9188088), ('event', 0.91882753), ('shoulda', 0.9189912), ('glasses', 0.9194119), ('bending', 0.9202636), ('detail', 0.9208051), ('just', 0.92319804), ('mirror', 0.92332387), ('pushed', 0.92478716), ('shows', 0.9254398), ('tumbles', 0.9264121), ('that', 0.92651874), ('sandals', 0.92686313), ('force', 0.92745996), ('crocker', 0.9278228), ('weight', 0.92917997), ('nicely', 0.92939), ('grading', 0.9297501), ('usual', 0.9300671), ('paying', 0.9300683), ('amount', 0.9302801), ('homely', 0.9307292), ('female', 0.9307879), ('bemused', 0.930823), ('deeply', 0.93120354), ('above', 0.93191206), ('tilting', 0.9329413), ('even', 0.9330752), ('medium', 0.93356603), ('only', 0.9341601), ('looks', 0.9344382), ('pissed', 0.9355861), ('about', 0.93687737), ('along', 0.9373741), ('large', 0.9377472), ('grasses', 0.93814236), ('wrong', 0.93835276), ('pulled', 0.9389903), ('flying', 0.93920445), ('killed', 0.9394628), ('gushing', 0.939479), ('crawled', 0.93988496), ('third', 0.9401271), ('length', 0.9402089), ('light', 0.9408756), ('opened', 0.94098514), ('count', 0.9412144), ('cleanup', 0.9412185), ('where', 0.9417837), ('shushing', 0.9420213), ('style', 0.94305885), ('quite', 0.9436462), ('robbing', 0.9437633), ('doing', 0.9438854), ('both', 0.944725), ('danger', 0.9449132), ('climbed', 0.94491404), ('total', 0.9454569), ('glance', 0.94560283), ('like', 0.9462767), ('colors', 0.94738084), ('autumn', 0.94753194), ('hurry', 0.9476221), ('cause', 0.9481112), ('gotten', 0.9491131), ('silly', 0.94926214), ('gather', 0.9492883), ('curved', 0.9495549), ('caught', 0.9495951), ('garage', 0.9501102), ('being', 0.9502108), ('motion', 0.9506182), ('haywire', 0.9506667), ('every', 0.9508417), ('plus', 0.9513655), ('spells', 0.95221895), ('drawers', 0.9522524), ('bandaids', 0.9528834), ('titled', 0.95386237), ('caused', 0.954417), ('parted', 0.9551879), ('seems', 0.95539474), ('crying', 0.95550394), ('would', 0.9556407), ('sleeves', 0.95590556), ('hammer', 0.95708483), ('shades', 0.95725316), ('pencils', 0.9580964), ('sight', 0.9582711), ('fallen', 0.9587819), ('ruining', 0.95917964), ('and', 0.9594953), ('turns', 0.9606776), ('flower', 0.9611754), ('overran', 0.9612305), ('ruffling', 0.9615978), ('commode', 0.9622965), ('string', 0.9636273), ('filled', 0.9647268), ('those', 0.96493864), ('sound', 0.9653766), ('pudding', 0.96615905), ('fact', 0.9667485), ('eating', 0.9669601), ('slower', 0.9672011), ('losing', 0.9676195), ('nephew', 0.96779984), ('spiral', 0.9678388), ('write', 0.9683873), ('wrote', 0.9692844), ('hands', 0.96951586), ('sloppy', 0.97009605), ('splash', 0.9701156), ('speak', 0.9706128), ('sense', 0.97085845), ('plain', 0.9710406), ('cockeyed', 0.97121656), ('three', 0.9712262), ('hiding', 0.97127473), ('cover', 0.9717841), ('board', 0.97212595), ('plants', 0.9724991), ('here', 0.9734464), ('steady', 0.97367626), ('rubber', 0.9744973), ('curtain', 0.97498536), ('time', 0.9751825), ('touch', 0.9751846), ('stick', 0.97579956), ('these', 0.9758142), ('neatness', 0.97619826), ('johnny', 0.97675204), ('laterals', 0.9782258), ('space', 0.9783224), ('words', 0.9792256), ('crazy', 0.9792589), ('nobody', 0.9812003), ('plates', 0.9812199), ('across', 0.98256606), ('this', 0.98302084), ('finger', 0.9839251), ('cookies', 0.98434013), ('height', 0.9846957), ('laying', 0.98479134), ('moves', 0.98487496), ('image', 0.9848952), ('hungry', 0.98491204), ('tipped', 0.9853406), ('fixed', 0.9856183), ('below', 0.98580176), ('dishes', 0.9859463), ('aware', 0.9861843), ('styles', 0.98623466), ('happy', 0.98643243), ('swiping', 0.9866799), ('vaseline', 0.9868932), ('spread', 0.9875842), ('clogged', 0.98762035), ('lucky', 0.98766375), ('bunch', 0.9877894), ('button', 0.9884114), ('young', 0.9891667), ('mixed', 0.9899394), ('breeze', 0.9907215), ('could', 0.9912266), ('floors', 0.99202096), ('table', 0.9926637), ('awful', 0.9928052), ('enjoy', 0.9931078), ('groomed', 0.9934997), ('mean', 0.99365556), ('toppled', 0.9936724), ('well', 0.9941577), ('ladder', 0.994344), ('vased', 0.99463147), ('moved', 0.9946632), ('calmly', 0.99468243), ('part', 0.99483943), ('depart', 0.99729186), ('work', 0.99797434), ('share', 0.99803054), ('raiding', 0.99814683), ('fancy', 0.9985899), ('break', 0.99883467), ('woman', 0.99981314), ('alike', 1.0003489), ('makes', 1.0004572), ('upward', 1.0006286), ('water', 1.0007659), ('wetting', 1.000856), ('child', 1.0009179), ('galore', 1.0039914), ('coffee', 1.0051734), ('seven', 1.0072091), ('doubt', 1.0073406), ('spilled', 1.008483), ('closet', 1.008597), ('kicked', 1.0086781), ('oldest', 1.0098497), ('watch', 1.0114564), ('backs', 1.0117217), ('facing', 1.0117719), ('tilted', 1.0118335), ('cracks', 1.0119462), ('ahead', 1.0128298), ('waisted', 1.0133804), ('aside', 1.0134362), ('kiddos', 1.0134878), ('ruffled', 1.0135585), ('paper', 1.01364), ('until', 1.0144409), ('kind', 1.0148616), ('bring', 1.0150646), ('scene', 1.0169222), ('good', 1.017302), ('casement', 1.0174179), ('clause', 1.0175483), ('throat', 1.0178002), ('sleeve', 1.0179912), ('drive', 1.0182651), ('shorts', 1.0183616), ('raised', 1.0186352), ('green', 1.0186864), ('home', 1.019406), ('eight', 1.0201244), ('color', 1.0202613), ('never', 1.0208601), ('dream', 1.0221213), ('scallop', 1.0221658), ('sort', 1.0222024), ('heard', 1.0224252), ('climbs', 1.0224493), ('know', 1.0249481), ('badly', 1.028732), ('back', 1.0289537), ('bushes', 1.0293151), ('build', 1.0293742), ('straps', 1.0305486), ('toward', 1.0306848), ('bitchy', 1.0307722), ('left', 1.0311024), ('under', 1.0313413), ('takes', 1.0313727), ('older', 1.0315683), ('snowing', 1.0328596), ('also', 1.0329975), ('quick', 1.0337238), ('tried', 1.0340089), ('with', 1.0343186), ('lousy', 1.0344746), ('drawer', 1.0345155), ('shape', 1.0370396), ('upraised', 1.0373832), ('real', 1.0374353), ('ready', 1.037489), ('harmed', 1.0378224), ('line', 1.0379072), ('needs', 1.0380136), ('valance', 1.0380342), ('sure', 1.040629), ('years', 1.0410844), ('wants', 1.043808), ('soaked', 1.0443023), ('them', 1.0445123), ('lower', 1.044635), ('fresh', 1.0449891), ('blouse', 1.0451005), ('faucets', 1.0452586), ('hand', 1.0456475), ('fine', 1.0456996), ('form', 1.0457505), ('angle', 1.0463811), ('blind', 1.0465909), ('apart', 1.0473642), ('tumble', 1.0477186), ('piece', 1.0479989), ('crack', 1.0486774), ('sorry', 1.0487802), ('girls', 1.0495174), ('mantle', 1.049543), ('baking', 1.0495595), ('huhuh', 1.0508449), ('drying', 1.0512568), ('catch', 1.0513576), ('pantry', 1.0519637), ('upset', 1.0531094), ('knows', 1.0534556), ('tennis', 1.0547259), ('legged', 1.0549523), ('fully', 1.0551221), ('turkey', 1.0552452), ('look', 1.055878), ('laugh', 1.0566279), ('cloudy', 1.0567222), ('over', 1.0569116), ('floor', 1.0569849), ('throw', 1.0573279), ('ones', 1.0581231), ('sleeved', 1.0584918), ('said', 1.0585128), ('cross', 1.0595818), ('falled', 1.0598601), ('show', 1.0598766), ('taken', 1.0602919), ('washed', 1.0616262), ('dress', 1.0619886), ('dishpan', 1.0622537), ('empty', 1.0628889), ('lunch', 1.0637933), ('clean', 1.0645841), ('come', 1.0646718), ('knock', 1.0646808), ('tasty', 1.0649643), ('reach', 1.0659305), ('topple', 1.0660203), ('same', 1.0667112), ('broke', 1.0674344), ('gloves', 1.0689398), ('drops', 1.06944), ('spigots', 1.0696234), ('mouths', 1.0697879), ('washes', 1.0702105), ('apples', 1.0706816), ('jersey', 1.0718418), ('quiet', 1.0720227), ('anklets', 1.0722235), ('call', 1.0728996), ('hairdo', 1.073237), ('damn', 1.0739226), ('they', 1.0746726), ('tired', 1.0756708), ('some', 1.0761596), ('ample', 1.07678), ('their', 1.0769415), ('drawn', 1.0775757), ('dance', 1.077634), ('full', 1.077764), ('done', 1.0800169), ('cookie', 1.0800782), ('pompoms', 1.0805386), ('dirty', 1.0806704), ('adult', 1.0808561), ('shelf', 1.0814654), ('heck', 1.0822301), ('area', 1.0828935), ('apron.', 1.0840763), ('rest', 1.0841151), ('plant', 1.0842987), ('rooms', 1.0846531), ('says', 1.0846852), ('wreck', 1.084989), ('turn', 1.0851203), ('drapes', 1.0855806), ('falls', 1.0858095), ('life', 1.086271), ('plate', 1.0872082), ('momma', 1.0885589), ('pardon', 1.0885932), ('allow', 1.0897518), ('forth', 1.0900344), ('past', 1.0906503), ('does', 1.0906606), ('washer', 1.0908153), ('long', 1.0915345), ('stop', 1.0928432), ('darn', 1.0936673), ('angry', 1.0958935), ('now', 1.0959494), ('mommy', 1.0967091), ('type', 1.0968283), ('grassy', 1.0984237), ('came', 1.0989401), ('saved', 1.1001238), ('steps', 1.1009102), ('gotta', 1.1009868), ('twins', 1.1013072), ('talk', 1.1039081), ('waving', 1.1054258), ('very', 1.1057844), ('find', 1.106778), ('went', 1.1077718), ('tries', 1.108134), ('okay', 1.108764), ('need', 1.1091574), ('towels', 1.109254), ('shrubs', 1.1098969), ('index', 1.1112156), ('but', 1.1120784), ('worst', 1.1126226), ('more', 1.112717), ('will', 1.1138145), ('untied', 1.1144487), ('toying', 1.114472), ('shiny', 1.1150689), ('yeah', 1.1152545), ('from', 1.1154596), ('game', 1.1158334), ('women', 1.1163032), ('puddle', 1.116343), ('fouled', 1.1163563), ('read', 1.1167995), ('minds', 1.1169972), ('outer', 1.1171191), ('mercy', 1.1182947), ('upper', 1.1196858), ('faced', 1.1204507), ('view', 1.1206756), ('room', 1.1208456), ('shall', 1.1210494), ('wiping', 1.1215571), ('billy', 1.1216764), ('goes', 1.1221805), ('book', 1.1222068), ('play', 1.1222134), ('pass', 1.1222918), ('cloth', 1.1225371), ('nice', 1.1229538), ('away', 1.1230738), ('want', 1.123196), ('gives', 1.1233509), ('devil', 1.1233903), ('bloom', 1.1247337), ('smile', 1.124851), ('mouth', 1.1263894), ('bench', 1.1281555), ('side', 1.1282747), ('when', 1.1294427), ('strap', 1.1302767), ('hell', 1.130872), ('many', 1.1314539), ('flood', 1.1319773), ('page', 1.1329879), ('spigot', 1.1331712), ('mind', 1.1336834), ('aprons', 1.1337446), ('horse', 1.133869), ('each', 1.1348543), ('trees', 1.1352775), ('next', 1.1354933), ('karen', 1.1364056), ('glass', 1.1374521), ('spills', 1.1387995), ('combed', 1.1393074), ('clock', 1.1397371), ('pants', 1.1399292), ('haste', 1.1400157), ('hard', 1.14042), ('chair', 1.141342), ('faucet', 1.1414394), ('gone', 1.1418995), ('elbow', 1.1421196), ('made', 1.1426562), ('open', 1.14456), ('word', 1.1454713), ('must', 1.1460946), ('shirt', 1.146276), ('boots', 1.1463616), ('base', 1.1464037), ('skirt', 1.1467761), ('hedges', 1.1470234), ('near', 1.147087), ('shoes', 1.147581), ('honey', 1.1477998), ('head', 1.148062), ('main', 1.150194), ('tripod', 1.1507753), ('girl', 1.1512889), ('have', 1.1525956), ('sissy', 1.1528176), ('shush', 1.1531485), ('dozen', 1.1539018), ('crash', 1.1547948), ('grass', 1.1556107), ('doors', 1.1558926), ('seen', 1.157926), ('make', 1.1593764), ('basin', 1.1602603), ('half', 1.1605276), ('steal', 1.1606122), ('blown', 1.1645162), ('sledge', 1.1647722), ('mess', 1.1653624), ('down', 1.1654), ('soon', 1.1656657), ('road', 1.1664286), ('pick', 1.1678593), ('face', 1.1680242), ('hoyle', 1.1682917), ('trunk', 1.1686071), ('hickey', 1.1692924), ('drugs', 1.1698124), ('inner', 1.1713463), ('tell', 1.1713871), ('all', 1.1714565), ('drain', 1.17209), ('sinks', 1.1727148), ('chaos', 1.1730452), ('much', 1.1735847), ('cream', 1.1737231), ('year', 1.1741233), ('thank', 1.1756953), ('one', 1.1765473), ('four', 1.1771078), ('ends', 1.1785779), ('mouse', 1.1796592), ('hold', 1.1796845), ('stools', 1.1797649), ('flat', 1.1804521), ('care', 1.1804705), ('nails', 1.1805568), ('lady', 1.1811293), ('shove', 1.18172), ('wreath', 1.1823727), ('climb', 1.1832656), ('dried', 1.183338), ('noise', 1.1837062), ('land', 1.1837869), ('spill', 1.1838086), ('payed', 1.1839496), ('jelly', 1.1860915), ('sunny', 1.1871629), ('gosh', 1.1875927), ('fall', 1.1883353), ('take', 1.1884282), ('socks', 1.1888095), ('fast', 1.1890727), ('ever', 1.1891694), ('runs', 1.18961), ('hope', 1.1896921), ('eaten', 1.1923082), ('miss', 1.1929555), ('else', 1.1937039), ('help', 1.1944249), ('puffed', 1.1954205), ('saucers', 1.196283), ('high', 1.196685), ('hears', 1.1971306), ('told', 1.1984695), ('knees', 1.1987506), ('sign', 1.1988833), ('sill', 1.1997471), ('ankle', 1.2008775), ('towel', 1.2028006), ('five', 1.2034627), ('knew', 1.2041223), ('sets', 1.2055238), ('days', 1.2061027), ('nine', 1.2061626), ('quit', 1.2063276), ('kids', 1.2066808), ('ours', 1.20685), ('cease', 1.2092435), ('uhoh', 1.2097107), ('wahoo', 1.2100259), ('mini', 1.2101073), ('want.', 1.2124289), ('birds', 1.2127173), ('wrist', 1.2159581), ('took', 1.2167367), ('step', 1.2169455), ('dumb', 1.2180243), ('trip', 1.2201048), ('lost', 1.2201535), ('yes', 1.2202809), ('sees', 1.2206818), ('wowie', 1.2214061), ('too', 1.2214142), ("boy's", 1.2226644), ('boys', 1.2226644), ('till', 1.2234482), ('baby', 1.2238654), ('swung', 1.2243401), ('saucer', 1.2254983), ('way', 1.2268589), ('into', 1.2282257), ('tops', 1.2298937), ('weeds', 1.2303883), ('the', 1.230557), ('give', 1.2305626), ('beat', 1.2306197), ('apron', 1.2313571), ('brisk', 1.2314935), ('what', 1.2316092), ('stay', 1.2317547), ('keeps', 1.2321999), ('drop', 1.232484), ('sugar', 1.2330256), ('wall', 1.2340901), ('puffy', 1.2381833), ('pair', 1.2385648), ('too.', 1.2397493), ('spout', 1.2400308), ('your', 1.2410147), ('sizes', 1.2452043), ('puts', 1.2453169), ('foot', 1.2454197), ('jerk', 1.2462668), ('than', 1.2468488), ('mamas', 1.247017), ('walk', 1.2480398), ('yards', 1.2485148), ('whom', 1.2506374), ('snowy', 1.251066), ('hear', 1.2513602), ('dear', 1.2530162), ('kill', 1.2536532), ('bowls', 1.2542688), ('fell', 1.256283), ('goody', 1.2578719), ('stool', 1.2581117), ('busy', 1.2585725), ('slip', 1.2586776), ('wife', 1.2588258), ('wipes', 1.2588563), ('pull', 1.2592257), ('rear', 1.2603906), ('lose', 1.2606928), ('food', 1.2607746), ('dump', 1.2631966), ('lots', 1.264659), ('spank', 1.2650241), ('hedge', 1.2662516), ('hits', 1.270385), ('were', 1.2708571), ('race', 1.2714516), ('hurt', 1.2744952), ('seem', 1.274802), ('cocoa', 1.275008), ('door', 1.2756296), ('geese', 1.2761785), ('kick', 1.2762755), ('yard', 1.2770821), ('wiped', 1.2771057), ('send', 1.2791084), ('dries', 1.2795304), ('gave', 1.2800729), ('you', 1.2801901), ('say', 1.2808808), ('man', 1.2824067), ('bang', 1.2833142), ('mama', 1.2834789), ('path', 1.285175), ('daft', 1.2881134), ('edge', 1.2897214), ('wing', 1.2921067), ('warm', 1.2924509), ('panes', 1.292642), ('dryer', 1.2926893), ('wait', 1.29296), ('for', 1.2978898), ('draw', 1.2982454), ('use', 1.2995971), ('ball', 1.2999672), ('male', 1.3003346), ('whew', 1.3027304), ('shrub', 1.3030668), ('duck', 1.3044127), ('junk', 1.3046943), ('yet', 1.3048207), ('did', 1.3056672), ('uhuh', 1.3057353), ('dish', 1.3070569), ('copy', 1.3070838), ('end', 1.308235), ('pool', 1.3090571), ('wind', 1.3091376), ('sole', 1.3113471), ('gets', 1.3122107), ('tilts', 1.3125856), ('ties', 1.3130225), ('bent', 1.3153384), ('hers', 1.3175772), ('skin', 1.3188738), ('see', 1.3191342), ('plug', 1.3192052), ('tree', 1.3198215), ('okey', 1.3203424), ('out', 1.3213763), ('sink', 1.3214929), ('wash', 1.3223765), ('day', 1.3237994), ('upend', 1.3245116), ('plum', 1.3251568), ('new', 1.3253512), ('weak', 1.3276707), ('size', 1.3281653), ('knobs', 1.3285915), ('grab', 1.3291416), ('poor', 1.3299392), ('set', 1.3310578), ('feet', 1.3322077), ('yep', 1.333427), ('legs', 1.3347299), ('laid', 1.3363482), ('leaf', 1.3376695), ('meal', 1.3380159), ('keep', 1.3386363), ('own', 1.3390464), ('drew', 1.3400286), ('trap', 1.3414611), ('roof', 1.3441595), ('wood', 1.3469175), ('cook', 1.3469325), ('put', 1.3515332), ('daze', 1.3518715), ('shut', 1.3521553), ('kept', 1.3541173), ('tied', 1.3566577), ('nose', 1.3584498), ('act', 1.3631625), ('pipe', 1.3695898), ('tippy', 1.3702191), ('guy', 1.3719331), ('big', 1.3723544), ('pour', 1.3727893), ('onto', 1.3746554), ('run', 1.3753797), ('prong', 1.3769118), ('lawn', 1.3773203), ('seat', 1.379585), ('been', 1.3807276), ('knots', 1.3823454), ('was', 1.3834878), ('tilt', 1.3841432), ('bush', 1.3845423), ('lips', 1.3854825), ('hey', 1.3871447), ('lane', 1.3876965), ('slid', 1.3897282), ('loud', 1.3904005), ('hair', 1.3916408), ('top', 1.3937479), ('scan', 1.3939617), ('lot', 1.3940411), ('neck', 1.3958008), ('knee', 1.3963838), ('dokey', 1.4015663), ('saw', 1.4019163), ('off', 1.4047867), ('tear', 1.4059752), ('hmm', 1.4072024), ('old', 1.4078462), ('pans', 1.4098581), ('bet', 1.4105533), ('tidy', 1.4107), ('leak', 1.4130055), ('not', 1.417245), ('cake', 1.4197917), ('verb', 1.4202048), ('boy', 1.4212185), ('tips', 1.4234704), ('rows', 1.4241749), ('mad', 1.4251906), ('bad', 1.4286944), ('bit', 1.4295666), ('can', 1.4312757), ('him', 1.4317607), ('got', 1.4320785), ('bowl', 1.4322488), ('kid', 1.4328694), ('her', 1.4330897), ('get', 1.4333998), ("mom's", 1.4338274), ('heel', 1.4339162), ('see.', 1.4344134), ('wow', 1.4350996), ('ask', 1.4372411), ('why', 1.4384414), ('rice', 1.4424441), ('shoe', 1.4425446), ('snow', 1.4444325), ('sakes', 1.4487609), ('may', 1.4522071), ('any', 1.4528416), ('job', 1.457339), ('and.', 1.4630787), ('huh', 1.467005), ('slop', 1.4697505), ('try', 1.4697849), ('cups', 1.4701858), ('she', 1.4714313), ('the.', 1.4726169), ('curl', 1.4782932), ('calm', 1.4806172), ('our', 1.4810752), ('toes', 1.4831598), ('pots', 1.4851589), ('its', 1.4865118), ('june', 1.4872496), ('hit', 1.4877356), ('two', 1.4913657), ('far', 1.4969112), ('wipe', 1.4985394), ('fun', 1.4988352), ('who', 1.5012245), ('mom', 1.5071514), ('beep', 1.5080367), ('son', 1.5088391), ('gee', 1.5122807), ('hiss', 1.5154902), ('cut', 1.5159999), ('age', 1.518278), ('huh.', 1.5196658), ('wee', 1.5204164), ('etch', 1.5210958), ('ten', 1.5232589), ('are', 1.5254831), ('ran', 1.5267687), ('hot', 1.5289186), ('eggs', 1.5302289), ('jars', 1.5304959), ('gal', 1.5327386), ('low', 1.5395107), ('god', 1.5405554), ('sue.', 1.5433152), ('air', 1.5476547), ('wove', 1.5484262), ('few', 1.5492876), ('had', 1.5512424), ('bed', 1.5526206), ('dad', 1.5575577), ('eat', 1.5636983), ('his', 1.5638816), ('arm', 1.5693275), ('let', 1.5697681), ('it', 1.5705615), ('sad', 1.5765046), ('so', 1.5776625), ('or', 1.5787324), ('lay', 1.5792184), ('sash', 1.5813016), ('lad', 1.5839901), ('suds', 1.5890942), ('tie', 1.5905809), ('has', 1.5906081), ('deaf', 1.592173), ('dry', 1.5988597), ('tip', 1.6019702), ('ails', 1.6020328), ('leg', 1.6104491), ('tin', 1.6155429), ('mhm', 1.6198845), ('flue', 1.6219282), ('lids', 1.625902), ("'ll", 1.6279109), ('ass', 1.6305535), ('tap', 1.6371415), ('jam', 1.6445755), ('wet', 1.6446666), ('ajar', 1.6449059), ('so.', 1.6456597), ('pan', 1.6523416), ('ate', 1.6554438), ('nor', 1.6650051), ('is', 1.6678337), ('eve', 1.6700798), ('eye', 1.6700921), ('or.', 1.6719289), ('leg.', 1.6751045), ('how', 1.6762238), ('no.', 1.6771946), ('do', 1.6782742), ('sun', 1.7033525), ('jar.', 1.7088056), ('shh', 1.7279088), ('rag', 1.7287887), ('tad', 1.7322173), ('toe', 1.746266), ('lip', 1.7519182), ('tub', 1.7628415), ("'re", 1.7636905), ('ret', 1.7675014), ('dug', 1.7702287), ('lid', 1.77514), ('jar', 1.7798208), ("'ve", 1.7811925), ('tup', 1.7903208), ('gon', 1.7935925), ('tea', 1.7937768), ('hm', 1.7960472), ('go', 1.7996713), ('as', 1.8042847), ('cup', 1.8206236), ('in', 1.8252), ('ah.', 1.8294266), ('ah', 1.833894), ("'s.", 1.836089), ('on', 1.8366226), ("'s", 1.8384761), ('we', 1.8412105), ('if', 1.8413013), ('hm.', 1.8436054), ('mop', 1.8609148), ('he', 1.8638787), ('me', 1.8663857), ('to', 1.8710617), ('oh', 1.8737997), ('ago', 1.8755449), ('of', 1.8949347), ('fry', 1.899475), ('oh.', 1.9113046), ('at', 1.9200758), ('us', 1.9237912), ('th', 1.9261274), ('no', 1.9291383), ('ya', 1.9446107), ('ai', 1.9472628), ('ta', 1.9602516), ('up', 2.0005991), ("n't", 2.0076501), ('nt', 2.0076501), ('wo', 2.0108745), ('by', 2.0155196), ('na', 2.022354), ('oop', 2.0301309), ('oy', 2.030756), ('am', 2.0792549), ('ho', 2.0826664), ('my', 2.1430993), ('be', 2.1704202), ('a.', 2.1740706), ("'d", 2.18034), ('i.', 2.199671), ('ca', 2.2029927), ('tv', 2.205634), ('mm', 2.2475994), ('ssh', 2.3862245), ("'m", 2.4755332), ('an', 2.660492), ('a', 3.108365), ('i', 3.244889), ('+', 3.2779357), ('s', 3.3005848), ('v', 3.3994474)

A.0.1.2. *English CBOW model Common Crawl without subword information*

('that', 2.0908647), ('and', 2.0920684), ('so', 2.2098858), ('just', 2.2289498), ('it', 2.23818), ('though', 2.2558172), ('one', 2.2628033), ('now', 2.267099), ('only', 2.2682054), ('the', 2.2829373), ('but', 2.2859237), ('this', 2.3222492), ('all', 2.3332114), ('even', 2.3580344), ('too', 2.3658123), ('then', 2.3772957), ('both', 2.383317), ('which', 2.4197624), ('like', 2.428278), ('either', 2.4282978), ('way', 2.4341278), ('not', 2.4744313), ('actually', 2.475012), ('here', 2.4762619), ('obviously', 2.47957), ('in', 2.4907897), ('with', 2.4934402), ('really', 2.505556), ('another', 2.5132318), ('do', 2.518693), ('to', 2.52026), ('still', 2.5314708), ('a', 2.537976), ('because', 2.560396), ('for', 2.5615783), ('of', 2.5629427), ('time', 2.5649378), ('anyway', 2.5658479), ('also', 2.5668368), ('if', 2.5871367), ('work', 2.5980725), ('while', 2.6094828), ('again', 2.6137872), ('there', 2.618066), ('think', 2.6274972), ('part', 2.6293938), ('them', 2.6422303), ('apparently', 2.6432862), ('as', 2.6466703), ('first', 2.647639), ('out', 2.6526635), ('some', 2.6587126), ('other', 2.662976), ('where', 2.679025), ('well', 2.6830945), ('thing', 2.6845047), ('place', 2.687946), ('when', 2.6898627), ('thought', 2.6925516), ('little', 2.6949763), ('already', 2.6950986), ('something', 2.6996152), ('know', 2.7020264), ('those', 2.7034833), ('present', 2.7095208), ('probably', 2.7098062), ('what', 2.7105558), ('after', 2.7147725), ('certainly', 2.7177396), ('others', 2.7199643), ('say', 2.7247436), ('or', 2.7311897), ('back', 2.736106), ('fact', 2.7388997), ('is', 2.740929), ('day', 2.7560155), ('more', 2.7567506), ('particular', 2.7575245), ('although', 2.7606544), ('going', 2.7649467), ('good', 2.7651122), ('have', 2.7659864), ('you', 2.766291), ('on', 2.766319), ('kind', 2.7677677), ('before', 2.7714384), ('second', 2.7757912), ('instead', 2.7796817), ('anyhow', 2.7922735), ('point', 2.8005164), ('from', 2.8021886), ('these', 2.8069713), ('always', 2.8076313), ('perhaps', 2.807846), ('whole', 2.8099878), ('me', 2.810173), ('otherwise', 2.8129332), ('see', 2.8134222), ('since', 2.8193219), ('sure', 2.8290064), ('evidently', 2.8319302), ('being', 2.8336318), ('get', 2.8399172), ('means', 2.8419633), ('great', 2.8442402), ('use', 2.8468106), ('go', 2.850238), ('almost', 2.8508031), ('who', 2.8530123), ('definitely', 2.8531537), ('at', 2.854845), ('rather', 2.860018), ('same', 2.868328), ('very', 2.8715897), ('around', 2.8750558), ('own', 2.8761604), ('possibly', 2.8788562), ('end', 2.8840306), ('they', 2.8852036), ('over', 2.8904076), ('new', 2.8915918), ('next', 2.8932533), ('maybe', 2.8977323), ('its', 2.9001927), ('many', 2.9009166), ('about', 2.9016078), ('yes', 2.9018497), ('nothing', 2.9035606), ('reason', 2.9062214), ('never', 2.907142), ('left', 2.9085324), ('need', 2.9109793), ('right', 2.9120297), ('having', 2.915158), ('area', 2.9163342), ('quite', 2.9187486), ('show', 2.9199493), ("'s", 2.9244037), ('up', 2.9248183), ('real', 2.9288328), ('possible', 2.93326), ('come', 2.934883), ('sort', 2.9370923), ('home', 2.9374778), ('course', 2.9423182), ('by', 2.944883), ('often', 2.9454188), ('he', 2.947308), ('plus', 2.9490378), ('much', 2.9495142), ('no', 2.9497252), ('coming', 2.952297), ('unfortunately', 2.9531298), ('people', 2.9562252), ('any', 2.9607744), ('everything', 2.9608276), ('certain', 2.9617262), ('believe', 2.9634056), ('anything', 2.9654343), ('yet', 2.9674125), ('basically', 2.9690592), ('meanwhile', 2.9705017), ('things', 2.9731383), ('look', 2.9748352), ('stuff', 2.9833207), ('want', 2.9847105), ('full', 2.987472), ('hardly', 2.9877806), ('two', 2.9909778), ('rest', 2.9970536), ('beginning', 2.997568), ('enough', 2.9976864), ('ones', 2.9977567), ('thinking', 2.9994128), ('mostly', 3.0013692), ('working', 3.0097733), ('saying', 3.0191774), ('start', 3.020432), ('except', 3.0237904), ('earlier', 3.0238833), ('got', 3.0260477), ('set', 3.0265236), ('past', 3.0269547), ('we', 3.0333648), ('world', 3.0346026), ('finally', 3.0355268), ('my', 3.0357897), ('him', 3.036579), ('doing', 3.0372925), ('help', 3.0382392), ('essentially', 3.0389404), ('did', 3.0454652), ('presumably', 3.047608), ('man', 3.049286), ('better', 3.0502753), ('life', 3.0506012), ('side', 3.0517662), ('guess', 3.0527523), ('hand', 3.0537124), ('naturally', 3.0542016), ('turn', 3.0560353), ('big', 3.0562098), ('suppose', 3.0578194), ('whatever', 3.0583005), ('off', 3.060668), ('getting', 3.061061), ('year', 3.0615044), ('moment', 3.0617278), ('our', 3.0674644), ('pretty', 3.0679789), ('starting', 3.0683982), ('done', 3.0689595), ('why', 3.0730095), ('myself', 3.0733397), ('be', 3.0737357), ('us', 3.0745988), ('said', 3.075075), ('might', 3.0775664), ('his', 3.079766), ('form', 3.0807161), ('inside', 3.08319), ('above', 3.0859907), ('lot', 3.0906591), ('supposedly', 3.092736), ('called', 3.0941396), ('should', 3.0952892), ('through', 3.0975423), ('nice', 3.09769), ('including', 3.0994027), ('put', 3.1046023), ('their', 3.1051955), ('was', 3.1067007), ('supposed', 3.1070633), ('try', 3.107102), ('found', 3.1072242), ('somehow', 3.1076584), ('make', 3.1092358), ('sometimes', 3.1099017), ('type', 3.1133745), ('least', 3.1167278), ('an', 3.1175592), ('house', 3.118458), ('someone', 3.1203136), ('addition', 3.120379), ('three', 3.12386), ('her', 3.1254358), ('different', 3.126318), ('eventually', 3.1387255), ('further', 3.1423547), ('every', 3.1452425), ('hey', 3.1471112), ('came', 3.1503344), ('okay', 3.1513474), ('small', 3.1547325), ('theirs', 3.15475), ('person', 3.1561468), ('presume', 3.1583564), ('similar', 3.1584973), ('each', 3.159425), ('practically', 3.1645038), ('mean', 3.164734), ('elsewhere', 3.1659265), ('how', 3.167207), ('long', 3.1681354), ('leading', 3.168855), ('front', 3.1696615), ('fine', 3.1723907), ('third', 3.1746645), ('down', 3.1757696), ('shows', 3.1758382), ('beautiful', 3.1758912), ('outside', 3.1759212), ('line', 3.1772091), ('close', 3.1782937), ('your', 3.1785), ('remember', 3.1787326), ('making', 3.1789796), ('mentioned', 3.1790547), ('happens', 3.1836402), ('find', 3.1847959), ('large', 3.185249), ('assuming', 3.1933188), ('could', 3.1936858), ('consider', 3.19875), ('ever', 3.2081983), ('hope', 3.209445), ('th', 3.209501), ('looking', 3.209915), ('process', 3.2103555), ('would', 3.2106967), ('run', 3.2117891), ('had', 3.2138264), ('can', 3.2143154), ('bad', 3.2148619), ('ask', 3.222069), ('few', 3.222701), ('went', 3.2227478), ('will', 3.22362), ('until', 3.226253), ('number', 3.2277482), ('understand', 3.2279801), ('family', 3.2282667), ('completely', 3.2286286), ('fortunately', 3.2292826), ('cause', 3.2317305), ('does', 3.2339854), ('important', 3.2348657), ('matter', 3.2354383), ('old', 3.2372725), ('happen', 3.2382765), ('damn', 3.2399466), ('country', 3.2401383), ('unless', 3.2424726), ('bit', 3.2439377), ('gone', 3.2453265), ('let', 3.2466059), ('guy', 3.246865), ('take', 3.2479188), ('made', 3.2498975), ('style', 3.2510421), ('four', 3.252799), ('short', 3.2530675), ('meant', 3.2531042), ('darn', 3.2535326), ('she', 3.2558312), ('clear', 3.2607155), ('himself', 3.2641528), ('seeing', 3.2644374), ('call', 3.264486), ('near', 3.264909), ('main', 3.265016), ('looks', 3.2666464), ('realize', 3.271177), ('says', 3.2740445), ('soon', 3.2742443), ('special', 3.2786748), ('far', 3.2817073), ('goes', 3.2831335), ('heck', 3.28409), ('yep', 3.284306), ('state', 3.2869487), ('times', 3.2892916), ('showing', 3.2932599), ('yeah', 3.293932), ('years', 3.2950137), ('must', 3.2955394), ('comes', 3.298398), ('presuming', 3.3027782), ('somewhat', 3.3067558), ('oh', 3.3085933), ('away', 3.3104565), ('stop', 3.3115084), ('hard', 3.3123705), ('may', 3.312745), ('half', 3.3129396), ('trouble', 3.3141344), ('seems', 3.3170664), ('sense', 3.3173175), ('usual', 3.3200223), ('talk', 3.3205962), ('top', 3.322943), ('talking', 3.3268952), ('mention', 3.3279006), ('reference', 3.3280566), ('seen', 3.3281078), ('act', 3.3292177), ('telling', 3.3298793), ('single', 3.331767), ('has', 3.3318946), ('fun', 3.3370993), ('expect', 3.33765), ('anymore', 3.3377433), ('lovely', 3.3382678), ('stand', 3.338354), ('together', 3.341818), ('problem', 3.34221), ('saw', 3.3425615), ('read', 3.3432379), ('room', 3.344711), ('imagine', 3.346178), ('are', 3.3464093), ('book', 3.3501542), ('suspect', 3.3509293), ('alright', 3.35364), ('force', 3.3541455), ('whom', 3.355484), ('head', 3.3556013), ('appears', 3.3592176), ('beyond', 3.359875), ('extent', 3.363955), ('whereas', 3.3649533), ('beside', 3.3651454), ('building', 3.3660927), ('silly', 3.37136), ('knew', 3.371481), ('lots', 3.3730829), ('running', 3.3734999), ('wrong', 3.375919), ('stay', 3.376531), ('bring', 3.3766901), ('tell', 3.3767662), ('typical', 3.3775148), ('enjoy', 3.3780305), ('wonder', 3.3798769), ('young', 3.383038), ('s', 3.3856623), ('leaving', 3.3883147), ('else', 3.3888311), ('started', 3.3927953), ('give', 3.3928869), ('picture', 3.3941348), ('sometime', 3.396867), ('variety', 3.3982966), ('game', 3.4018207), ('days', 3.4039447), ('quickly', 3.405639), ('told', 3.407374), ('somewhere', 3.4079876), ('crazy', 3.4094079), ('along', 3.4094708), ('base', 3.4118397), ('assume', 3.4124446), ('considered', 3.4134927), ('five', 3.4136438), ('whoever', 3.4144635), ('figure', 3.415683), ('into', 3.417976), ('total', 3.4180136), ('normal', 3.418103), ('personal', 3.421183), ('wanted', 3.4215202), ('girl', 3.421784), ('action', 3.4222505), ('whether', 3.4229953), ('interesting', 3.423264), ('between', 3.430024), ('speak', 3.430313), ('view', 3.4325368), ('underneath', 3.4340942), ('job', 3.4345345), ('project', 3.4393308), ('happy', 3.4406154), ('ours', 3.4408684), ('event', 3.4415262), ('giving', 3.4425447), ('boy', 3.4426494), ('behind', 3.4449632), ('light', 3.445974), ('high', 3.4466796), ('halfway', 3.447796), ('+', 3.4483767), ('straight', 3.451849), ('continue', 3.4544759), ('allow', 3.4575462), ('double', 3.459196), ('appear', 3.4622858), ('terrible', 3.4638214), ('school', 3.4638293), ('were', 3.465636), ('bunch', 3.4704049), ('turned', 3.4707341), ('wanting', 3.4708118), ('alike', 3.474358), ('play', 3.4745388), ('trying', 3.4747922), ('please', 3.4756353), ('asking', 3.4759872), ('middle', 3.4784894), ('kid', 3.4793267), ('below', 3.479533), ('hoping', 3.48057), ('minute', 3.481565), ('writing', 3.483277), ('than', 3.4845586), ('somebody', 3.4852138), ('i', 3.4855745), ('round', 3.48564), ('amount', 3.488349), ('position', 3.4908888), ('ahead', 3.4912875), ('fairly', 3.494906), ('nobody', 3.4970267), ('awful', 3.4971194), ('setting', 3.4980447), ('kids', 3.4997215), ('putting', 3.5006914), ('taking', 3.5042937), ('unusual', 3.5043328), ('moving', 3.505241), ('hold', 3.5065284), ('fully', 3.5066817), ('discovered', 3.5083323), ('couple', 3.509516), ('space', 3.5099099), ('strong', 3.5116682), ('details', 3.5129857), ('open', 3.514757), ('forget', 3.516782), ('ah', 3.5176764), ('mad', 3.5194528), ('lady', 3.5198803), ('correct', 3.5228348), ('hell', 3.5233586), ('aside', 3.523657), ('seem', 3.5260448), ('observed', 3.5267515), ('finished', 3.5285857), ('situation', 3.5305111), ('helping', 3.5317345), ('stick', 3.532232), ('ground', 3.5344522), ('strange', 3.534541), ('section', 3.5365517), ('standing', 3.5376809), ('needs', 3.538665), ('hmm', 3.5415683), ('sight', 3.5421307), ('mother', 3.5515738), ('pass', 3.552348), ('living', 3.5545137), ('implies', 3.555378), ('wrote', 3.5561833), ('friend', 3.5563703), ('plain', 3.5580146), ('face', 3.5591614), ('stopped', 3.5617852), ('deciding', 3.5634458), ('keep', 3.5638862), ('struggle', 3.5659301), ('touch', 3.5673373), ('children', 3.5674138), ('ready', 3.570022), ('written', 3.5712888), ('fellow', 3.5732038), ('backs', 3.5735335), ('been', 3.574127), ('morning', 3.5743194), ('piece', 3.5744438), ('allowing', 3.575448), ('makes', 3.5755994), ('too.', 3.580919), ('moves', 3.5825992), ('mixed', 3.583066), ('portion', 3.5839603), ('looked', 3.5843241), ('woman', 3.5845513), ('anybody', 3.5855172), ('turns', 3.5862665), ('hands', 3.5871437), ('road', 3.593258), ('lost', 3.593861), ('mind', 3.595132), ('longer', 3.5953379), ('notice', 3.5956423), ('counts', 3.5962312), ('break', 3.596299), ('daughter', 3.5970342), ('reading', 3.5986776), ('amounts', 3.5992143), ('seven', 3.600988), ('wow', 3.6030135), ('eight', 3.6055784), ('huh', 3.6061041), ('opposite', 3.6113176), ('doubt', 3.6116626), ('fall', 3.6117837), ('land', 3.612127), ('sound', 3.6121767), ('extended', 3.6142817), ('difficult', 3.6168418), ('slightly', 3.6177332), ('care', 3.6185048), ('runs', 3.618637), ('step', 3.6197343), ('tops', 3.6198463), ('word', 3.620932), ('page', 3.6216292), ('aware', 3.6236598), ('indicating', 3.6239822), ('perspective', 3.626679), ('nine', 3.627988), ('moved', 3.630147), ('center', 3.6304462), ('pick', 3.6306813), ('sister', 3.6311693), ('hers', 3.6313703), ('according', 3.6314068), ('father', 3.6338952), ('strictly', 3.6373653), ('noticed', 3.6380107), ('taken', 3.6382828), ('hit', 3.638698), ('bet', 3.6388125), ('count', 3.6390579), ('quick', 3.639999), ('encouraging', 3.643207), ('took', 3.6441882), ('turning', 3.6459153), ('watch', 3.646843), ('holding', 3.647518), ('playing', 3.6480393), ('shortly', 3.6483896), ('forgetting', 3.6497884), ('against', 3.6513524), ('words', 3.6541743), ('finish', 3.6606567), ('passing', 3.6611812), ('gets', 3.6628308), ('summer', 3.6638095), ('description', 3.6652095), ('knows', 3.667513), ('activity', 3.6677701), ('husband', 3.671415), ('picked', 3.6715758), ('mess', 3.6722898), ('problems', 3.6723645), ('anywhere', 3.6723819), ('indicates', 3.6735547), ('handful', 3.6739705), ('hurry', 3.6740801), ('neither', 3.6742902), ('gosh', 3.674738), ('active', 3.6759994), ('son', 3.6786463), ('flat', 3.6814315), ('bottom', 3.6830208), ('herself', 3.6834512), ('allowed', 3.6840296), ('hm', 3.684847), ('child', 3.6858811), ('expecting', 3.685974), ('average', 3.685976), ('connection', 3.6863062), ('sill', 3.689016), ('mini', 3.69024), ('cover', 3.6923738), ('image', 3.6934438), ('interest', 3.6940062), ('heard', 3.6954396), ('trip', 3.6981537), ('correctly', 3.6985352), ('letting', 3.6988688), ('ten', 3.699159), ('water', 3.699938), ('lucky', 3.7012124), ('till', 3.703762), ('build', 3.7041116), ('drop', 3.7046914), ('hot', 3.7051227), ('missing', 3.7056904), ('drive', 3.705843), ('difference', 3.7064993), ('extension', 3.7079806), ('distinct', 3.7080867), ('scene', 3.7088857), ('partial', 3.7095509), ('lower', 3.7104042), ('low', 3.7129998), ('keeping', 3.7131896), ('wife', 3.7170453), ('bother', 3.718284), ('poor', 3.7194226), ('fast', 3.7201715), ('insisted', 3.7208173), ('ignoring', 3.7209651), ('background', 3.722961), ('indication', 3.7248123), ('permitting', 3.7252903), ('dumb', 3.7277987), ('condition', 3.7288587), ('proceeding', 3.7295942), ('ya', 3.7320065), ('pictured', 3.7324724), ('terribly', 3.7332826), ('slight', 3.7339334), ('starts', 3.734681), ('share', 3.7347295), ('brother', 3.737257), ('write', 3.7387102), ('fancy', 3.7389176), ('ends', 3.7391458), ('growing', 3.739445), ('fixed', 3.7407718), ('under', 3.7424583), ('partially', 3.7426045), ('changing', 3.7499762), ('sorry', 3.750216), ('shape', 3.7509658), ("boy's", 3.751315), ('boys', 3.751315), ('selection', 3.7521245), ('ta', 3.753755), ('kept', 3.7543406), ('nicely', 3.756717), ('spring', 3.7574692), ('sad', 3.7584603), ('pointed', 3.759398), ('unaware', 3.7597666), ('significance', 3.7617357), ('gave', 3.7626333), ('sets', 3.762937), ('badly', 3.7656868), ('corner', 3.7662394), ('wants', 3.7666447), ('paper', 3.7673745), ('hear', 3.769594), ('food', 3.7698538), ('counting', 3.770634), ('remain', 3.7714546), ('quietly', 3.7735546), ('sitting', 3.774897), ('awfully', 3.7749398), ('heading', 3.7783432), ('watching', 3.778371), ('cut', 3.780687), ('excuse', 3.7819717), ('handle', 3.7821467), ('sees', 3.7827666), ('gotten', 3.782958), ('quiet', 3.784346), ('onto', 3.7861059), ('table', 3.7865126), ('darling', 3.786681), ('across', 3.787794), ('walk', 3.7928078), ('seemed', 3.794043), ('older', 3.7981029), ('attempting', 3.799192), ('happening', 3.7998629), ('baby', 3.8019493), ('leaves', 3.8024693), ('girls', 3.8044245), ('indicate', 3.8045266), ('activities', 3.8050976), ("'d", 3.8065581), ('fricking', 3.8069847), ('sole', 3.8103952), ('venture', 3.8106449), ('sign', 3.8108566), ('fresh', 3.8120012), ('wait', 3.8131757), ('stepped', 3.8148282), ('mom', 3.816108), ('removed', 3.817965), ('worrying', 3.8196528), ('lousy', 3.8199184), ('movement', 3.8206575), ('noticing', 3.8216615), ('neighborhood', 3.8229656), ('wee', 3.8231852), ('visible', 3.8249414), ('dinner', 3.8260355), ('oy', 3.8274307), ('missed', 3.8310673), ('dangerous', 3.831792), ('board', 3.833107), ('ho', 3.8346097), ('wondering', 3.8351762), ('mistake', 3.8356209), ('takes', 3.8357527), ('busy', 3.8362045), ('miss', 3.8365428), ("'ll", 3.8373861), ('sudden', 3.837771), ('waiting', 3.8399532), ('gee', 3.8404958), ('pair', 3.8410811), ('conscious', 3.8417323), ('air', 3.8458354), ('adjacent', 3.8464704), ('dropped', 3.8473477), ('square', 3.847544), ('ordinary', 3.8481214), ('pull', 3.8502586), ('tried', 3.8520315), ('sounds', 3.8525596), ('lose', 3.8570154), ('clean', 3.8606925), ('size', 3.8657415), ('sticks', 3.8676922), ('green', 3.868639), ('upset', 3.8710618), ('send', 3.871148), ('age', 3.8712978), ('opposed', 3.872269), ('managed', 3.8745458), ('hurt', 3.8768005), ('causing', 3.8768196), ('enjoying', 3.8768206), ('junior', 3.8772576), ('pushed', 3.882076), ('signal', 3.8824248), ('so.', 3.8827403), ('handed', 3.882779), ('dump', 3.8843074), ('beat', 3.8863478), ('lay', 3.8866425), ('edge', 3.888622), ('rolling', 3.8887975), ('color', 3.8902335), ('ran', 3.8911743), ('pleasant', 3.892462), ('reach', 3.892719), ('hahaha', 3.8971395), ('eat', 3.897277), ('responsible', 3.8982656), ('affected', 3.9005227), ('gotta', 3.901539), ('dear', 3.9036179), ('foot', 3.9037464), ('walking', 3.9046016), ('quit', 3.9051418), ('detail', 3.905578), ('dream', 3.9080577), ('preparing', 3.9091883), ('identify', 3.9092305), ('worst', 3.9096136), ('carrying', 3.910692), ('catch', 3.910927), ('kitchen', 3.9119782), ('direction', 3.9128542), ('dad', 3.9142404), ('eve', 3.9156382), ('meantime', 3.9175804), ('v', 3.9181106), ('machine', 3.9182384), ('gather', 3.918452), ('actions', 3.920833), ('interested', 3.921946), ('whew', 3.924185), ('wo', 3.925811), ('silent', 3.9258928), ('moments', 3.9275968), ('towards', 3.9277444), ('forgot', 3.9293127), ('female', 3.9295566), ('winter', 3.9341266), ('pulled', 3.9342883), ('kill', 3.9345224), ('receive', 3.9361215), ('tasty', 3.936501), ('women', 3.9384997), ('hits', 3.9389238), ('pieces', 3.945396), ('danger', 3.9455342), ('deeply', 3.947647), ('checked', 3.9562364), ('floor', 3.9567215), ('someplace', 3.9581969), ('careful', 3.9583383), ('bed', 3.960045), ('solidly', 3.9611948), ('pattern', 3.9617047), ('broken', 3.9621649), ('throw', 3.9626384), ('lunch', 3.9643643), ('messed', 3.9651823), ('staring', 3.9665833), ('affair', 3.966722), ('drops', 3.96831), ('window', 3.974673), ('laugh', 3.975752), ('begging', 3.9764576), ('adult', 3.977011), ('sharing', 3.9784946), ('touching', 3.9795341), ('ample', 3.9811718), ('upsetting', 3.9813917), ('dropping', 3.9820485), ('counter', 3.9840195), ('healthy', 3.9850378), ('describe', 3.9854755), ('angle', 3.9857564), ('rooms', 3.9873943), ('length', 3.9891517), ('paying', 3.9903657), ('persons', 3.9904137), ('assisting', 3.991098), ('identifying', 3.9915438), ('hidden', 3.9932446), ('younger', 3.9933412), ('eating', 3.9936767), ('apart', 3.994812), ('na', 3.9957232), ('secret', 3.9974518), ('dry', 3.9991329), ('angry', 3.9996204), ('weak', 4.0021143), ('stepping', 4.0031376), ('tie', 4.008112), ('losing', 4.0108047), ('disaster', 4.01171), ('warm', 4.0142126), ('excited', 4.0149503), ('ai', 4.0159564), ('puts', 4.0174255), ('arm', 4.017607), ('killed', 4.017831), ('forth', 4.0210276), ('wall', 4.0212054), ('saved', 4.0219026), ('gives', 4.0219436), ('closed', 4.022283), ('slip', 4.0240054), ('rear', 4.0291862), ('oblivious', 4.031758), ('drawn', 4.031794), ('resting', 4.0326285), ('titled', 4.033606), ('god', 4.034426), ('meal', 4.0362062), ('grab', 4.0369473), ('and.', 4.0380006), ('neglected', 4.038463), ('caused', 4.0403957), ("'ve", 4.0424247), ('youngster', 4.044959), ('disturbed', 4.046521), ('define', 4.0472617), ('backyard', 4.0490294), ('handing', 4.0492687), ('broke', 4.049791), ("'re", 4.051137), ('grandmother', 4.052316), ("n't", 4.0537815), ('nt', 4.0537815), ('hanging', 4.0541706), ('copy', 4.055003), ('pointing', 4.055501), ('caught', 4.0569825), ('sideways', 4.060709), ('filled', 4.06126), ('daft', 4.0615907), ('reaching', 4.061711), ('rolled', 4.061915), ('fell', 4.063679), ('dress', 4.0640163), ('empty', 4.0655236), ('flying', 4.0691195), ('knock', 4.070079), ('draw', 4.0716443), ('handles', 4.0730214), ('medium', 4.0736384), ('receiving', 4.076364), ('garden', 4.0778494), ('steady', 4.0788836), ('summertime', 4.079148), ('bent', 4.0801682), ('ties', 4.081187), ('kick', 4.0820994), ('hiding', 4.0821247), ('door', 4.0831614), ('rocking', 4.083202), ('goodness', 4.0836053), ('catastrophe', 4.0842643), ("'s.", 4.0845904), ('signaling', 4.085538), ('balance', 4.086393), ('dirty', 4.0864964), ('gentle', 4.0889206), ('bang', 4.091568), ('picking', 4.092399), ('path', 4.0935273), ('tired', 4.0962296), ('steps', 4.0970755), ('tad', 4.097292), ("'m", 4.0973816), ('hundred', 4.097553), ('nor', 4.0975666), ('upper', 4.09763), ('obliterated', 4.097707), ('shocked', 4.0988555), ('listening', 4.1027613), ('artistic', 4.1052866), ('smiling', 4.1054), ('pissed', 4.105733), ('male', 4.106488), ('jerk', 4.10917), ('galore', 4.110563), ('blind', 4.116073), ('stretches', 4.116266), ('conflict', 4.116531), ('opened', 4.11854), ('cross', 4.120216), ('removing', 4.1231384), ('backwards', 4.1266065), ('catching', 4.127838), ('ca', 4.1279173), ('cutting', 4.1281514), ('falling', 4.1286807), ('nineteen', 4.130623), ('dreaming', 4.132775), ('shades', 4.136953), ('dish', 4.1387196), ('wing', 4.1405787), ('feet', 4.140913), ('expression', 4.141119), ('weight', 4.141611), ('wet', 4.142423), ('junk', 4.142804), ('tip', 4.1437683), ('gal', 4.1440434), ('dozen', 4.1443543), ('am', 4.145103), ('neighbor', 4.1461687), ('toward', 4.1477537), ('realizes', 4.1504507), ('duck', 4.1514974), ('throwing', 4.152488), ('colors', 4.152564), ('upside', 4.152782), ('see.', 4.154052), ('hitting', 4.155735), ('wreck', 4.156829), ('nephew', 4.1589923), ('falls', 4.159369), ('crying', 4.1596694), ('instructing', 4.160318), ('outer', 4.1604314), ('dance', 4.1613603), ('leg', 4.1617274), ('crack', 4.1675515), ('yard', 4.1682262), ('plant', 4.1700263), ('keeps', 4.171455), ('minutes', 4.172049), ('filling', 4.1743965), ('minds', 4.1759524), ('styles', 4.1770186), ('mirror', 4.1806784), ('cooking', 4.1834936), ('lad', 4.1854377), ('accident', 4.18649), ('plug', 4.1886387), ('culprit', 4.190449), ('salient', 4.190467), ('incomplete', 4.191867), ('directions', 4.1923175), ('outdoor', 4.193623), ('ate', 4.1963496), ('knocked', 4.197228), ('legs', 4.197708), ('perplexed', 4.202625), ('inner', 4.2054124), ('labelled', 4.207408), ('fallen', 4.2090335), ('unconcerned', 4.2102227), ('shining', 4.2108555), ('plate', 4.211374), ('autumn', 4.2116423), ('eye', 4.2121305), ('skin', 4.213891), ('pool', 4.214629), ('breaking', 4.216859), ('overrun', 4.217151), ('tied', 4.2177124), ('leaning', 4.2181287), ('flower', 4.2204614), ('slipping', 4.2251363), ('criticizing', 4.2270103), ('stretching', 4.2276807), ('animals', 4.2338142), ('grandson', 4.2358184), ('twins', 4.2374444), ('wearing', 4.238144), ('roof', 4.2407637), ('distracted', 4.2408657), ('descriptive', 4.2439046), ('race', 4.2444224), ('youngsters', 4.2451262), ('wash', 4.2460127), ('ass', 4.2461157), ('faced', 4.2474647), ('pathway', 4.247994), ('dumping', 4.2482624), ('plum', 4.250413), ('laid', 4.2529864), ('mama', 4.2533755), ('no.', 4.254431), ('breakfast', 4.2552915), ('tap', 4.25558), ('survived', 4.2560315), ('calamity', 4.2579703), ('splash', 4.2584352), ('singing', 4.258488), ('outdoors', 4.262381), ('slower', 4.264852), ('plants', 4.2663803), ('facing', 4.267256), ('tin', 4.26746), ('flowing', 4.2680936), ('height', 4.2694707), ('thank', 4.2721744), ('wind', 4.27237), ('sneaking', 4.274351), ('charged', 4.27499), ('careless', 4.2755837), ('painting', 4.2761364), ('or.', 4.2763686), ('slender', 4.277415), ('laughing', 4.279585), ('sink', 4.280232), ('ball', 4.2814584), ('grabbing', 4.282566), ('shiny', 4.283047), ('chaos', 4.28425), ('cease', 4.2850447), ('mercy', 4.2854943), ('kicked', 4.286797), ('tries', 4.288621), ('slipped', 4.292917), ('glass', 4.306715), ('seventy', 4.3086557), ('preoccupied', 4.310238), ('shelf', 4.3109326), ('shut', 4.3114705), ('sun', 4.312493), ('hammer', 4.316793), ('sloppy', 4.3169675), ('shall', 4.3209457), ('daylight', 4.3222322), ('mommy', 4.322294), ('spells', 4.3248525), ('homely', 4.3248954), ('peaceful', 4.327446), ('flexible', 4.327613), ('flood', 4.3283644), ('drew', 4.3297315), ('windows', 4.330851), ('furiously', 4.3328695), ('distinguish', 4.334089), ('steal', 4.334706), ('momma', 4.334713), ('garage', 4.335647), ('glance', 4.3358755), ('glimpse', 4.3374643), ('shoulder', 4.3377275), ('collapse', 4.338165), ('toe', 4.338891), ('laying', 4.338987), ('skirt', 4.3390083), ('doors', 4.339704), ('hungry', 4.3410463), ('cook', 4.3422356), ('clothes', 4.3424654), ('leaf', 4.3440895), ('rubber', 4.3460627), ('jam', 4.3485), ('coffee', 4.3493977), ('dressed', 4.350118), ('spiral', 4.3521395), ('jumping', 4.353812), ('index', 4.355805), ('instructions', 4.3574276), ('depends', 4.358317), ('wood', 4.3584175), ('reached', 4.364005), ('trap', 4.3647985), ('spread', 4.3648157), ('tv', 4.3652716), ('calmly', 4.3664184), ('reacting', 4.3666015), ('mm', 4.3667808), ('cloth', 4.3685923), ('sizes', 4.369532), ('curved', 4.370069), ('motion', 4.370855), ('trees', 4.371438), ('smile', 4.374542), ('chair', 4.3756046), ('upward', 4.3756275), ('giggling', 4.3780384), ('shirt', 4.3804235), ('rag', 4.3809104), ('damaged', 4.3819976), ('the.', 4.3831863), ('oldest', 4.383904), ('letters', 4.385194), ('lengths', 4.3868856), ('eaten', 4.388549), ('designate', 4.391103), ('hair', 4.391733), ('breeze', 4.392063), ('tidy', 4.3928366), ('billy', 4.3941708), ('cleaning', 4.3960724), ('strap', 4.3961525), ('nose', 4.3978834), ('neglecting', 4.397889), ('tree', 4.4006553), ('immersed', 4.4019537), ('tips', 4.4054174), ('neck', 4.406285), ('silence', 4.406523), ('string', 4.4083886), ('dishes', 4.4086003), ('payed', 4.4114647), ('june', 4.411755), ('pouring', 4.412782), ('goody', 4.413982), ('i.', 4.414285), ('depart', 4.414323), ('want.', 4.417187), ('elbow', 4.4175196), ('okey', 4.4191127), ('ruining', 4.4222317), ('yelling', 4.422496), ('coaxing', 4.423389), ('shoes', 4.4237432), ('crash', 4.4242234), ('blooming', 4.4260645), ('finger', 4.426451), ('weather', 4.4282246), ('johnny', 4.4287024), ('reaches', 4.428959), ('stacked', 4.4291263), ('admonishing', 4.430402), ('pan', 4.4330797), ('seat', 4.4331284), ('flooding', 4.434558), ('dried', 4.436885), ('shush', 4.437306), ('scan', 4.439322), ('raised', 4.4393406), ('airplane', 4.4403334), ('devil', 4.4419713), ('flowers', 4.4447927), ('tipped', 4.446028), ('nondescript', 4.446613), ('distressing', 4.4475636), ('harmed', 4.450473), ('noise', 4.4525666), ('uhhuh', 4.4581375), ('a.', 4.4583607), ('pipe', 4.461003), ('climbing', 4.4617953), ('crashing', 4.4619365), ('loud', 4.46398), ('outfitted', 4.465102), ('horse', 4.4654117), ('overlooking', 4.4676514), ('raising', 4.4687147), ('sentence', 4.4705863), ('whisper', 4.4737663), ('lip', 4.475477), ('soaked', 4.4761643), ('drugs', 4.478272), ('dripping', 4.478668), ('plates', 4.4811444), ('lawn', 4.4846926), ('brisk', 4.485541), ('floors', 4.487543), ('bench', 4.4900727), ('bemused', 4.491552), ('attention', 4.494369), ("mom's", 4.498598), ('stealing', 4.502474), ('pants', 4.5029917), ('bumping', 4.503602), ('tea', 4.507991), ('lips', 4.508635), ('boots', 4.508711), ('shove', 4.509274), ('smashed', 4.511194), ('stumbling', 4.512064), ('cautioning', 4.512163), ('honey', 4.51248), ('frowning', 4.515147), ('dug', 4.517257), ('fingers', 4.517822), ('bloom', 4.519279), ('sunny', 4.5227556), ('basin', 4.524217), ('knee', 4.526504), ('sidewalk', 4.5282364), ('shaking', 4.5306315), ('whispering', 4.5307174), ('mouth', 4.53251), ('cup', 4.5330615), ('button', 4.5344396), ('shoe', 4.536472), ('precarious', 4.537634), ('happenings', 4.540649), ('attentive', 4.541325), ('haste', 4.5421166), ('percent', 4.5454054), ('blown', 4.5484858), ('sleeve', 4.5485625), ('hazardous', 4.551795), ('slid', 4.553012), ('sissy', 4.554933), ('bending', 4.5552444), ('sinks', 4.5571904), ('clock', 4.557524), ('threesome', 4.5598817), ('birds', 4.56227), ('climb', 4.563784), ('daze', 4.5678654), ('tilt', 4.569143), ('parted', 4.572356), ('snickering', 4.5754633), ('grass', 4.5762177), ('scenery', 4.5763392), ('signing', 4.577357), ('uhuh', 4.579785), ('calm', 4.5806484), ('ago', 4.58087), ('shoulda', 4.58119), ('ladder', 4.582034), ('tub', 4.5966988), ('cake', 4.5979567), ('gadgets', 4.599737), ('cream', 4.601808), ('stopper', 4.603212), ('slanted', 4.6042194), ('lopsided', 4.6053615), ('foliage', 4.6060224), ('sprinkling', 4.609122), ('bowl', 4.6095076), ('pour', 4.6120234), ('cabinet', 4.613228), ('jar', 4.6139007), ('crossed', 4.6171594), ('washing', 4.6187167), ('soaking', 4.6189404), ('overflow', 4.619953), ('bitchy', 4.620807), ('hears', 4.621698), ('kiddos', 4.6250954), ('christmas', 4.6260734), ('knees', 4.6296406), ('tup', 4.630423), ('lane', 4.632855), ('nails', 4.6337056), ('walkway', 4.6371603), ('socks', 4.638388), ('tear', 4.640464), ('snowy', 4.641592), ('cascading', 4.6423607), ('thicker', 4.6432214), ('abstractly', 4.645616), ('blowing', 4.6465364), ('crawling', 4.647151), ('baking', 4.64746), ('decorating', 4.648333), ('crooked', 4.6487784), ('mistakes', 4.6496334), ('wobbling', 4.6499066), ('jogging', 4.6554794), ('sugar', 4.656294), ('rice', 4.66332), ('cups', 4.6643543), ('overflowing', 4.669756), ('lid', 4.672102), ('tilted', 4.6805987), ('sneakers', 4.6806784), ('impervious', 4.681244), ('swung', 4.681305), ('washed', 4.6823187), ('drain', 4.6825876), ('ankle', 4.683977), ('shorts', 4.685985), ('oh.', 4.6890597), ('closet', 4.689635), ('pleasures', 4.6910768), ('draining', 4.691832), ('cracks', 4.6936045), ('tumbling', 4.6960735), ('throat', 4.6968703), ('clause', 4.7015834), ('climbs', 4.7027826), ('bush', 4.7037163), ('toes', 4.705204), ('distract', 4.705773), ('rows', 4.7091064), ('mhm', 4.7121406), ('gravity', 4.7125907), ('karen', 4.71482), ('housekeeping', 4.717129), ('drawer', 4.717428), ('cloudy', 4.718716), ('bordered', 4.719698), ('leak', 4.7236915), ('planted', 4.7315497), ('tumble', 4.7340484), ('gon', 4.7347507), ('slop', 4.736054), ('spreading', 4.740329), ('fiddling', 4.741818), ('observant', 4.7430806), ('drying', 4.743263), ('bushes', 4.7470527), ('ret', 4.74777), ('trunk', 4.749313), ('cabinets', 4.751967), ('snow', 4.752071), ('puffy', 4.7564063), ('heel', 4.7592463), ('trousers', 4.760458), ('jelly', 4.763161), ('shrubs', 4.7644258), ('raining', 4.7703333), ('cockeyed', 4.77232), ('torrent', 4.7723846), ('haywire', 4.77336), ('apples', 4.774929), ('sweater', 4.7784076), ('straps', 4.779429), ('uhoh', 4.779751), ('wahoo', 4.782757), ('sleeveless', 4.7833166), ('splashed', 4.788112), ('pineapple', 4.792875), ('mouse', 4.794291), ('robbing', 4.7990465), ('unconscious', 4.8041215), ('cupboard', 4.8078175), ('landscaping', 4.809812), ('daydreaming', 4.8110375), ('driveway', 4.811909), ('turkey', 4.8153305), ('bowls', 4.8161936), ('pots', 4.817796), ('shh', 4.822664), ('slanting', 4.824606), ('weeds', 4.826867), ('wiped', 4.827094), ('accidents', 4.827336), ('pans', 4.8276114), ('strapped', 4.8286543), ('leaking', 4.8301105), ('huh.', 4.831093), ('recessed', 4.8356695), ('cleanup', 4.8361506), ('shrubbery', 4.8369036), ('towel', 4.8369546), ('toying', 4.840874), ('engrossed', 4.8458014), ('petroleum', 4.847221), ('wrist', 4.8542137), ('curtains', 4.856815), ('overrunning', 4.8598375), ('platter', 4.8627033), ('wipes', 4.865682), ('drapes', 4.8667655), ('waving', 4.8670225), ('fry', 4.867712), ('diamonds', 4.8727193), ('washes', 4.873929), ('wipe', 4.875621), ('mouths', 4.8806424), ('legged', 4.8810205), ('grasses', 4.8815746), ('fountain', 4.8820605), ('climbed', 4.8826933), ('tumbles', 4.8845825), ('gushing', 4.886056), ('absentmindedly', 4.886942), ('gloves', 4.8871894), ('gesturing', 4.8892417), ('sleeves', 4.8922706), ('spill', 4.8945537), ('mop', 4.895288), ('tilting', 4.8964734), ('tennis', 4.896984), ('toppled', 4.901345), ('sandals', 4.9167843), ('teetering', 4.9204326), ('housewife', 4.9204764), ('glasses', 4.927168), ('cookie', 4.9298444), ('cocoa', 4.9314795), ('grassy', 4.9331126), ('jars', 4.933141), ('discharged', 4.9337673), ('silencing', 4.9361014), ('spout', 4.937179), ('inattentive', 4.93743), ('donating', 4.939692), ('overturning', 4.9401937), ('blouse', 4.941809), ('attentively', 4.9426126), ('grading', 4.9446535), ('washer', 4.945677), ('tipping', 4.948067), ('pantry', 4.961836), ('splashing', 4.961844), ('curtain', 4.964251), ('puddle', 4.9663835), ('moisture', 4.9675384), ('swiping', 4.968286), ('crawled', 4.9688997), ('hiss', 4.9723043), ('topple', 4.972599), ('eggs', 4.982301), ('curl', 4.985842), ('housekeeper', 4.98859), ('overboard', 4.9937496), ('tilts', 4.9952164), ('beep', 4.995999), ('shrub', 5.0003023), ('mantle', 5.0053897), ('yards', 5.0076604), ('drawers', 5.008067), ('jersey', 5.0081143), ('ruffled', 5.011468), ('slopping', 5.0174017), ('verb', 5.019611), ('apron', 5.045441), ('slippers', 5.0470963), ('ah.', 5.0487328), ('geese', 5.0494833), ('copyright', 5.051449), ('stool', 5.051878), ('drooping', 5.05359), ('fluttering', 5.0539885), ('towels', 5.057377), ('pencils', 5.0604553), ('toppling', 5.073779), ('lids', 5.076266), ('raiding', 5.0795074), ('wiping', 5.0804386), ('hairdo', 5.0843453), ('dries', 5.0872846), ('outstretched', 5.0935264), ('pardon', 5.0952787), ('pudding', 5.0990653), ('billowing', 5.103607), ('evergreens', 5.1117287), ('mamas', 5.1139402), ('cookies', 5.118392), ('sakes', 5.1221013), ('etch', 5.127246), ('squirting', 5.129844), ('splattered', 5.139367), ('spilling', 5.13953), ('hedge', 5.151756), ('memorize', 5.153137), ('unbalances', 5.1580725), ('overflows', 5.169667), ('cupboards', 5.1788874), ('puffed', 5.1803594), ('overran', 5.1822996), ('deaf', 5.1879873), ('falled', 5.1919203), ('dishwasher', 5.193047), ('countertop', 5.193984), ('dryer', 5.195247), ('combed', 5.197404), ('clogged', 5.208178), ('sash', 5.208576), ('huhuh', 5.2115903), ('untied', 5.213554), ('motioning', 5.2142367), ('overfilled', 5.2204466), ('sledge', 5.225486), ('scallop', 5.2296286), ('ails', 5.230563), ('knobs', 5.2365756), ('spills', 5.23731), ('wove', 5.238714), ('saucer', 5.257269), ('groomed', 5.2664165), ('commode', 5.2674937), ('miniskirt', 5.2709956), ('carpentry', 5.2719126), ('snowing', 5.2844315), ('suds', 5.2848577), ('overflowed', 5.29232), ('dishwater', 5.3029737), ('waisted', 5.3146625), ('spilled', 5.3149576), ('hm.', 5.328412), ('stockings', 5.3404317), ('panes', 5.340738), ('upend', 5.341178), ('hedges', 5.3468566), ('aprons', 5.352728), ('spank', 5.353221), ('wetting', 5.3572426), ('wowie', 5.360895), ('fouled', 5.3697295), ('tripod', 5.3840876), ('sleeved', 5.3878083), ('faucet', 5.397679), ('vaseline', 5.414614), ('prong', 5.417558), ('oop', 5.4296575), ('neatness', 5.431798), ('snitching', 5.443614), ('knots', 5.448757), ('spigot', 5.4500294), ('protrude', 5.483966), ('draperies', 5.4984646), ('pompoms', 5.5009203), ('upraised', 5.516057), ('faucets', 5.5228214), ('stools', 5.523981), ('leg.', 5.5611544), ('flue', 5.5798645), ('bandaids', 5.591637), ('pompadour', 5.595873), ('windowsills', 5.6258597), ('tippy', 5.6285467), ('wreath', 5.6634817), ('nuthouse', 5.6695824), ('ajar', 5.689156), ('ruffling', 5.689658), ('shushing', 5.699189), ('overflown', 5.7354226), ('saucers', 5.7524076), ('anklets', 5.7687955), ('crocker', 5.774255), ('spigots', 5.7760806), ('hickey', 5.7964835), ('valance', 5.83507), ('overturns', 5.8810773), ('sue.', 5.907385), ('dokey', 5.93109), ('hoyle', 5.9714737), ('laterals', 5.9986196), ('dishpan', 6.026424), ('casement', 6.110257), ('jar.', 6.2407675), ('apron.', 6.3752275), ('tripodal', 6.396562), ('ssh', 6.440299)

A.0.2. Common Crawl and Wikipedia CBOW models with subword information

See E. Grave et al[**50**] for the implementation details of these models

A.0.2.1. *French model, Picnic corpus full vocabulary*

('propriétaires', 0.35758877), ('d'identification', 0.3944995), ('mademoiselle', 0.39996502), ('grand-chose', 0.42103803), ('personnages', 0.42606032), ('appartement', 0.43203473), ('reproduction', 0.43672895), ('difficulté', 0.44184092), ('impression', 0.44292438), ('demoiselle', 0.45286793), ('aménagement', 0.45827416), ('grand-père', 0.46133018), ('proximité', 0.46580577), ('intérieur', 0.46686688), ('grand-frère', 0.47969562), ('pique-nique', 0.48578495), ('couverture', 0.48727682), ('personnes', 0.487565), ('gendarmerie', 0.4900348), ('exemple', 0.49629745), ('automobile', 0.4994929), ('chaussettes', 0.50248665), ('personne', 0.5026566), ('opération', 0.5059918), ('situation', 0.50601166), ('nord-ouest', 0.50609), ('histoire', 0.5075176), ('propriété', 0.51749414), ('stéréotypes', 0.5209794), ('français', 0.52121997), ('alentours', 0.52170235), ('plaisanciers', 0.5227564), ('nouvelles', 0.52619547), ('chaussures', 0.5268109), ('cerf-volant', 0.5279882), ('conjointe', 0.5320119), ('promenade', 0.5367809), ('résidence', 0.5367885), ('grand-papa', 0.5425854), ('fla-fla', 0.547905), ('pantoufles', 0.55474985), ('espadrilles', 0.5585054), ('instant', 0.5608468), ('monsieur', 0.5617424), ('montagnes', 0.57436496), ('douzaine', 0.57459754), ('bouteille', 0.5768037), ('véhicule', 0.5803223), ('rapport', 0.5838426), ('garnitures', 0.58787227), ('casquette', 0.590991), ('pantalons', 0.592418), ('fillette', 0.5927177), ('traverse', 0.5943681), ('campagne', 0.594678), ('peinture', 0.59644145), ('couverte', 0.59674704), ('famille', 0.5968012), ('feuillage', 0.5969658), ('affaires', 0.5978325), ('étudiant', 0.60701776), ('vacances', 0.61318296), ('sœurette', 0.61481935), ('conjoint', 0.6298223), ('enfants', 0.63032436), ('chaudière', 0.6313666), ('courant', 0.63189524), ('pantalon', 0.63581055), ('affaire', 0.6366765), ('monument', 0.6400068), ('dizaine', 0.6409981), ('serviette', 0.6416634), ('branches', 0.64191437), ('journée', 0.6421393), ('genre-là', 0.6455916), ('arrière', 0.64757025), ('dessous', 0.6486711), ('parents', 0.6511412), ('fenêtres', 0.6544048), ('voiture', 0.65558374), ('lecture', 0.65671766), ('musique', 0.6584947), ('feuilles', 0.6661552), ('terrain', 0.6677856), ('poissons', 0.6699213), ('retrait', 0.672704), ('attache', 0.67473155), ('souliers', 0.67850393), ('dedans', 0.68122166), ('partie', 0.68554455), ('dessus', 0.694506), ('horizon', 0.6982131), ('garçons', 0.6997538), ('maison', 0.70011014), ('fenêtre', 0.7024164), ('besoin', 0.7052333), ('chapeau', 0.70524234), ('sandales', 0.71658623), ('buissons', 0.72244614), ('arbustes', 0.72436154), ('costume', 0.7275846), ('cuillère', 0.7276097), ('dessins', 0.7283293), ('breuvage', 0.7322181), ('château', 0.7329703), ('antenne', 0.73316854), ('châteaux', 0.7337286), ('liquide', 0.73428285), ('lunettes', 0.7348878), ('rivière', 0.73605037), ('patente', 0.7370345), ('chemise', 0.7374616), ('boissons', 0.7377751), ('paysage', 0.73820126), ('poisson', 0.73865294), ('médaille', 0.7422315), ('maisons', 0.74723554), ('châssis', 0.7526334), ('choses', 0.75347847), ('enfant', 0.75377285), ('veille', 0.75386894), ('détail', 0.75497466), ('sentier', 0.75842375), ('entrée', 0.75884897), ('carpette', 0.76469725), ('planche', 0.76491046), ('droite', 0.76678854), ('collier', 0.7695461), ('gauche', 0.76956904), ('phrases', 0.77027637), ('verdure', 0.7703374), ('oiseaux', 0.7705353), ('conifère', 0.7759602), ('ombrage', 0.77909976), ('chandail', 0.7801632), ('garçon', 0.784859), ('espèce', 0.7877551), ('avant', 0.7880719), ('couple', 0.7919321), ('bosquet', 0.7991276), ('t-shirt', 0.80115354), ('pelouse', 0.80128086), ('chemin.', 0.8033738), ('drapeau', 0.8034438), ('volume', 0.80665696), ('chemin', 0.80676115), ('femmes', 0.80739063), ('orchard', 0.80810124), ('années', 0.81015867), ('cheveux', 0.8127854), ('genre', 0.8142699), ('course', 0.8153003), ('voilier', 0.81795084), ('camisole', 0.818431), ('bouches', 0.8217281), ('place', 0.8219508), ('pêcheur', 0.8225377), ('liqueur', 0.8248566), ('couples', 0.82492167), ('sorte', 0.82680887), ('adulte', 0.8283664), ('petits', 0.82913727), ('doute', 0.82989126), ('pousses', 0.82990265), ('culotte', 0.8312904), ('numéro', 0.8320101), ('écoute', 0.8342299), ('genoux', 0.8351762), ('temps', 0.83835953), ('camping', 0.84543204), ('manches', 0.84566766), ('moteur', 0.847146), ('madame', 0.84722155), ('bateau', 0.8539547), ('thermos', 0.85442406), ('bermuda', 0.85624766), ('monde', 0.8620422), ('blonde', 0.86292565), ('plante', 0.8641346), ('sujet', 0.86685646), ('garage', 0.86693347), ('cheval', 0.86851877), ('panier', 0.870872), ('petit', 0.8718097), ('homme', 0.87520015), ('plaque', 0.875474), ('soleil', 0.87859356), ('chose', 0.8822381), ('langue', 0.8840727), ('livre', 0.8850671), ('animal', 0.8856325), ('fille', 0.8868898), ('rivage', 0.88827187), ('camion', 0.889284), ('terme', 0.8893145), ('fleurs', 0.89135855), ('volant', 0.8927614), ('image', 0.89573795), ('arbres', 0.89830536), ('jambes', 0.9029645), ('pigeon', 0.9040267), ('style', 0.90452206), ('anneau', 0.90506035), ('jeune', 0.9073741), ('bulldog', 0.9089023), ('femme', 0.9163706), ('ligne', 0.9182775), ('ouais', 0.919044), ('chalet', 0.92197967), ('scène', 0.9260013), ('indien', 0.9266599), ('peine', 0.9267531), ('poupée', 0.9278891), ('train', 0.9288216), ('frère', 0.9305654), ('navire', 0.93162775), ('porte', 0.9334446), ('gamin', 0.9344431), ('touffes', 0.940175), ('fleuve', 0.94349027), ('souche', 0.9495143), ('canada', 0.95156205), ('quai.', 0.95218366), ('pattes', 0.9524311), ('bobine', 0.9530068), ('fait', 0.95748174), ('texte', 0.9616757), ('hot-dog', 0.96258277), ('terre', 0.9642138), ('photo', 0.9662894), ('nuages', 0.9694819), ('bosse', 0.9742022), ('souper', 0.97696567), ('route', 0.9805876), ('radeau', 0.98105043), ('québec', 0.98188007), ('donc', 0.98641944), ('poteau', 0.98883766), ('ouest', 0.99272), ('large', 1.0036298), ('pelles', 1.0042125), ('maman', 1.0067027), ('truite', 1.0078505), ('ombre', 1.0083616), ('pieds', 1.0266846), ('pichet', 1.0288837), ('mains', 1.034332), ('voiles', 1.0352365), ('roman', 1.0355728), ('allée', 1.0374999), ('chien', 1.03898), ('aviron', 1.044475), ('part', 1.0447158), ('dupont', 1.0552948), ('radio', 1.0747455), ('érable', 1.0756065), ('plage', 1.0879799), ('anges', 1.0906398), ('fois', 1.0925523), ('repas', 1.0959677), ('côté', 1.1060721), ('pelle', 1.10685), ('avion', 1.1157897), ('paire', 1.1190608), ('boite', 1.1194569), ('tapis', 1.1217319), ('verre', 1.1227362), ('tasse', 1.1227474), ('nuage', 1.1244512), ('bout.', 1.1252209), ('boîte', 1.128043), ('arbre', 1.1294967), ('pêche', 1.134014), ('short', 1.1355337), ('étage', 1.1381241), ('sens', 1.1408592), ('coté', 1.1499264), ('tête', 1.1513233), ('fond', 1.1585466), ('pitou', 1.1624216), ('voyer', 1.1641436), ('cooler', 1.1695746), ('gars', 1.1733669), ('herbe', 1.1750085), ('queue', 1.182825), ('bière', 1.1865537), ('sapin', 1.1915406), ('père', 1.1916186), ('sceau', 1.1938003), ('tour', 1.199801), ('face', 1.2067837), ('scout', 1.2068276), ('sable', 1.2075893), ('louis', 1.2086076), ('voile', 1.2134328), ('nappe', 1.2145536), ('fils', 1.2210325), ('aide', 1.2218627), ('soir', 1.2242849), ('corde', 1.2287399), ('gens', 1.2330258), ('gilet', 1.2394174), ('main.', 1.2428163), ('pays', 1.2446994), ('long', 1.2514141), ('lunch', 1.2517012), ('dunes', 1.2547396), ('grève', 1.2569041), ('gazon', 1.259875), ('mère', 1.2651126), ('boisé', 1.2696201), ('mari', 1.2701906), ('papa', 1.2731718), ('haut', 1.2791064), ('plan', 1.2839457), ('auto.', 1.3110886), ('bout', 1.3191693), ('main', 1.3249536), ('bord', 1.3273237), ('dieu', 1.328674), ('cité', 1.3327928), ('sœur', 1.340839), ('mots', 1.3435413), ('bois', 1.364263), ('cèdre', 1.3658913), ('dame', 1.3666223), ('cour', 1.3814017), ('bébé', 1.382994), ('café', 1.3986834), ('port', 1.4116884), ('bain', 1.4142685), ('mimi', 1.4209392), ('auto', 1.435433), ('aise', 1.4544616), ('toit', 1.460658), ('vent', 1.4611012), ('peau', 1.462367), ('pont', 1.4706979), ('rame', 1.50483), ('sexe', 1.5076038), ('lac.', 1.517351), ('rive', 1.5261723), ('noël', 1.5569111), ('quai', 1.5599339), ('eaux', 1.6003847), ('race', 1.6173464), ('ding', 1.6233786), ('cas', 1.6254677), ('bus.', 1.629893), ('foin', 1.6333681), ('son', 1.6718472), ('fido', 1.6807272), ('mal', 1.7342423), ('euh', 1.7460696), ('nom', 1.7846867), ('ben', 1.8307676), ('été', 1.9069505), ('ok.', 1.9123869), ('air', 1.9270046), ('lit', 1.9579239), ('eau', 1.9946096), ('nom.', 1.9976349), ('mer', 2.0143878), ('sol', 2.0391202), ('fil', 2.0869265), ('âge', 2.1733193), ('sac', 2.2186875), ('cou', 2.2348332), ('vin', 2.2404203), ('pot', 2.244819), ('mat', 2.278324), ('bol', 2.2806735), ('lac', 2.3402455), ('thé', 2.3488524), ('mât', 2.4584613), ('dog', 2.5653305), ('du', 3.0930722), ('ok', 3.5996804)

A.0.2.2. *English model, DementiaBank corpus full vocabulary*

('unfortunately', 0.26862106), ('apparently', 0.33851898), ('essentially', 0.34301236), ('particular', 0.34715593), ('presumably', 0.35538542), ('fortunately', 0.36167252), ('obviously', 0.367618), ('supposedly', 0.36787397), ('practically', 0.37794423), ('interesting', 0.37944183), ('everything', 0.38092178), ('definitely', 0.3862852), ('something', 0.38714626), ('eventually', 0.38766038), ('significance', 0.39390844), ('basically', 0.39718598), ('meanwhile', 0.39847457), ('considered', 0.3985513), ('certainly', 0.40186659), ('understand', 0.40344322), ('perspective', 0.40432206), ('neighborhood', 0.4059759), ('actually', 0.40693936), ('attempting', 0.4087332), ('including', 0.41080478), ('otherwise', 0.41137987), ('absentmindedly', 0.41153082), ('encouraging', 0.41229007), ('beginning', 0.41879013), ('description', 0.4202711), ('elsewhere', 0.42317787), ('sometimes', 0.42350325), ('evidently', 0.42929918), ('discovered', 0.4333781), ('instructions', 0.43340808), ('forgetting', 0.43400076), ('although', 0.43444884), ('indication', 0.43679336), ('underneath', 0.43684843), ('probably', 0.43791962), ('distinguish', 0.4379492), ('difference', 0.43820837), ('responsible', 0.44305444), ('instructing', 0.4442265), ('possibly', 0.44537672), ('different', 0.4462385), ('completely', 0.44717723), ('connection', 0.44938582), ('obliterated', 0.45078856), ('presuming', 0.4539943), ('naturally', 0.45421058), ('grandmother', 0.4552605), ('addition', 0.45597777), ('reference', 0.45677462), ('interested', 0.4571511), ('expecting', 0.4608922), ('supposed', 0.46228477), ('important', 0.46271476), ('somewhere', 0.46487194), ('catastrophe', 0.467597), ('proceeding', 0.4678907), ('identifying', 0.4680594), ('possible', 0.46844375), ('background', 0.46905386), ('activities', 0.4700728), ('whatever', 0.47054097), ('descriptive', 0.47132), ('mentioned', 0.4715329), ('criticizing', 0.47173586), ('preoccupied', 0.47205412), ('beautiful', 0.47263947), ('anything', 0.47529647), ('situation', 0.4801009), ('according', 0.48140478), ('distressing', 0.48284462), ('because', 0.4845818), ('starting', 0.48510227), ('wondering', 0.48641995), ('another', 0.4865743), ('unconcerned', 0.4870371), ('permitting', 0.4890893), ('nondescript', 0.492058), ('overflowing', 0.49267295), ('happening', 0.49318004), ('indicating', 0.49691555), ('admonishing', 0.49735588), ('difficult', 0.49756372), ('incomplete', 0.4995848), ('housekeeping', 0.49967864), ('youngsters', 0.4996886), ('assuming', 0.50094676), ('instead', 0.50210196), ('summertime', 0.50225025), ('thinking', 0.5031728), ('indicates', 0.503302), ('expression', 0.5042937), ('already', 0.5046379), ('stretching', 0.50496256), ('extension', 0.50789684), ('remember', 0.5080305), ('daydreaming', 0.51010627), ('consider', 0.51035666), ('neglecting', 0.51247144), ('thought', 0.512504), ('overlooking', 0.5139019), ('suppose', 0.51549727), ('perhaps', 0.52128583), ('condition', 0.5221065), ('outstretched', 0.5232175), ('correctly', 0.5265522), ('directions', 0.5266438), ('believe', 0.52774835), ('direction', 0.5294022), ('present', 0.5310812), ('overturning', 0.53226507), ('somebody', 0.53234684), ('preparing', 0.53272206), ('receiving', 0.53306603), ('either', 0.5338114), ('selection', 0.5343393), ('housekeeper', 0.53694195), ('snickering', 0.5370992), ('upsetting', 0.53749883), ('straight', 0.5379829), ('happenings', 0.5409068), ('youngster', 0.54116535), ('dangerous', 0.54171723), ('standing', 0.54210824), ('partially', 0.54229015), ('someplace', 0.5424245), ('disturbed', 0.54257), ('terrible', 0.5437858), ('landscaping', 0.5456471), ('distracted', 0.5457147), ('building', 0.54587156), ('appears', 0.5459328), ('certain', 0.5460264), ('presume', 0.5466153), ('continue', 0.5469623), ('stretches', 0.5474108), ('position', 0.54857147), ('overrunning', 0.5503296), ('together', 0.550336), ('unconscious', 0.5505524), ('cautioning', 0.55146354), ('assisting', 0.5535954), ('precarious', 0.55616707), ('though', 0.5565249), ('whereas', 0.55689293), ('perplexed', 0.55814517), ('finished', 0.559203), ('extended', 0.5594808), ('nothing', 0.5596743), ('working', 0.56012166), ('neglected', 0.5601956), ('whispering', 0.5617047), ('sprinkling', 0.5619473), ('allowing', 0.56216556), ('conscious', 0.56292653), ('personal', 0.56349933), ('similar', 0.5646362), ('meantime', 0.5651639), ('daughter', 0.5653634), ('struggle', 0.5663993), ('observed', 0.5664307), ('attentively', 0.5674176), ('designate', 0.56926256), ('earlier', 0.56950486), ('finally', 0.56985873), ('anyway', 0.57022387), ('describe', 0.5730473), ('oblivious', 0.5735297), ('impervious', 0.57391566), ('somewhat', 0.57433754), ('inattentive', 0.5764251), ('sometime', 0.5764723), ('anymore', 0.57704026), ('abstractly', 0.57711416), ('listening', 0.5774202), ('children', 0.5777641), ('anywhere', 0.57827264), ('changing', 0.57933027), ('looking', 0.5800683), ('decorating', 0.58011574), ('ordinary', 0.5808281), ('backwards', 0.581306), ('realize', 0.5816036), ('activity', 0.5818155), ('unbalances', 0.58203125), ('somehow', 0.58321255), ('mention', 0.5841312), ('counting', 0.5843838), ('someone', 0.58535033), ('carrying', 0.5855719), ('started', 0.58579886), ('interest', 0.585849), ('problems', 0.58694667), ('discharged', 0.5882043), ('noticing', 0.58857095), ('indicate', 0.5890717), ('happens', 0.5904168), ('imagine', 0.5904768), ('distinct', 0.590787), ('insisted', 0.5915989), ('affected', 0.5920284), ('ignoring', 0.5928206), ('getting', 0.5934955), ('showing', 0.59574854), ('pointing', 0.59778017), ('reaching', 0.5978554), ('typical', 0.5990233), ('through', 0.59904945), ('outfitted', 0.59942925), ('wanting', 0.60083276), ('enjoying', 0.6010037), ('pictured', 0.6022836), ('others', 0.6028448), ('deciding', 0.6036926), ('anyhow', 0.6039587), ('putting', 0.604417), ('country', 0.6045816), ('outside', 0.6049761), ('alright', 0.6051491), ('talking', 0.60522485), ('signaling', 0.6056513), ('identify', 0.60698146), ('suspect', 0.6072305), ('movement', 0.6090453), ('breakfast', 0.6093273), ('opposite', 0.6093587), ('opposed', 0.6118498), ('allowed', 0.6118653), ('dropping', 0.61300105), ('adjacent', 0.61341393), ('leaving', 0.6134193), ('implies', 0.6137714), ('attention', 0.61384773), ('running', 0.6144256), ('always', 0.61483794), ('touching', 0.61501867), ('pleasant', 0.6157562), ('problem', 0.6158876), ('really', 0.6159399), ('trouble', 0.61609054), ('leading', 0.61613995), ('slightly', 0.6162367), ('telling', 0.616296), ('worrying', 0.61694473), ('fluttering', 0.61881906), ('catching', 0.619243), ('correct', 0.62024593), ('watching', 0.62053114), ('overfilled', 0.621405), ('furiously', 0.6229412), ('goodness', 0.6235412), ('stopped', 0.62494963), ('splattered', 0.62525827), ('holding', 0.6257226), ('coming', 0.6261196), ('threesome', 0.62635607), ('throwing', 0.62664974), ('labelled', 0.6270464), ('evergreens', 0.6272613), ('project', 0.62762046), ('special', 0.6281395), ('setting', 0.6286112), ('process', 0.6298232), ('stepping', 0.63032514), ('should', 0.63048714), ('terribly', 0.6308948), ('engrossed', 0.6311489), ('strictly', 0.6311535), ('breaking', 0.63134354), ('spreading', 0.63288814), ('turning', 0.6329852), ('noticed', 0.63308734), ('windowsills', 0.6332099), ('christmas', 0.6367305), ('pointed', 0.6369425), ('stumbling', 0.6375248), ('called', 0.63762355), ('sideways', 0.638184), ('countertop', 0.6382384), ('grandson', 0.6384565), ('slipping', 0.639078), ('letting', 0.63980764), ('passing', 0.6409329), ('overboard', 0.64093953), ('things', 0.64205223), ('picture', 0.6421472), ('having', 0.64230627), ('husband', 0.64292395), ('unusual', 0.6432549), ('variety', 0.6433641), ('quickly', 0.64337283), ('overflowed', 0.643445), ('conflict', 0.64463913), ('dreaming', 0.6460332), ('survived', 0.64613503), ('written', 0.64618933), ('removing', 0.6469314), ('strange', 0.6483693), ('before', 0.6495718), ('whoever', 0.6506239), ('heading', 0.6524001), ('grabbing', 0.6537957), ('cascading', 0.6554319), ('whether', 0.65558565), ('writing', 0.65562075), ('neighbor', 0.65572447), ('anybody', 0.65588516), ('except', 0.6561526), ('silencing', 0.6565788), ('backyard', 0.6571593), ('reacting', 0.65720516), ('missing', 0.65765196), ('people', 0.65853435), ('sleeveless', 0.6589319), ('playing', 0.65925306), ('details', 0.660117), ('dishwasher', 0.6603954), ('managed', 0.6606729), ('shrubbery', 0.661137), ('nineteen', 0.66148084), ('pineapple', 0.6617071), ('further', 0.66208106), ('waiting', 0.66260517), ('fricking', 0.6673939), ('helping', 0.6676419), ('handful', 0.6685442), ('rather', 0.66924393), ('splashing', 0.673252), ('housewife', 0.6732948), ('section', 0.67397267), ('saying', 0.67400557), ('sneaking', 0.67408985), ('morning', 0.67447716), ('distract', 0.67467856), ('second', 0.67528087), ('attentive', 0.6753551), ('unaware', 0.6754779), ('flexible', 0.67594457), ('hazardous', 0.67695266), ('dropped', 0.6769935), ('receive', 0.6771499), ('squirting', 0.67743987), ('sitting', 0.67966837), ('calamity', 0.6797714), ('brother', 0.6807589), ('which', 0.6813222), ('realizes', 0.6818058), ('careless', 0.68244237), ('stepped', 0.6827877), ('partial', 0.68311477), ('removed', 0.6836445), ('keeping', 0.68368244), ('artistic', 0.68421197), ('accident', 0.6843043), ('blooming', 0.6870226), ('outdoors', 0.6872253), ('gesturing', 0.6872765), ('neither', 0.68737525), ('pleasures', 0.68790454), ('darling', 0.68847215), ('shortly', 0.68915564), ('collapse', 0.6893773), ('course', 0.68970025), ('almost', 0.6901057), ('quietly', 0.6901142), ('reason', 0.6901201), ('shoulder', 0.6908723), ('expect', 0.69121575), ('actions', 0.6912251), ('painting', 0.69147867), ('teetering', 0.6915289), ('theirs', 0.69199127), ('disaster', 0.69216204), ('mistake', 0.6921807), ('visible', 0.69231564), ('laughing', 0.6924701), ('accidents', 0.6935258), ('himself', 0.693595), ('reading', 0.6939299), ('making', 0.6951074), ('unless', 0.69557464), ('little', 0.6956534), ('crashing', 0.6964612), ('checked', 0.6968399), ('petroleum', 0.69725394), ('portion', 0.6980844), ('inside', 0.69858766), ('stealing', 0.69907546), ('sharing', 0.69928104), ('lopsided', 0.6995156), ('against', 0.6998302), ('trying', 0.7007367), ('better', 0.7011107), ('causing', 0.70132965), ('turned', 0.70168036), ('sentence', 0.7021572), ('mostly', 0.70218873), ('giggling', 0.702681), ('fiddling', 0.7027688), ('overflows', 0.70280254), ('person', 0.7034856), ('halfway', 0.7036847), ('growing', 0.70369744), ('miniskirt', 0.7044582), ('enough', 0.7048527), ('motioning', 0.70505464), ('matter', 0.7055406), ('tumbling', 0.7058532), ('wanted', 0.7067591), ('draining', 0.7090542), ('observant', 0.709442), ('cleaning', 0.71103215), ('immersed', 0.71136177), ('average', 0.7114587), ('mistakes', 0.7118114), ('around', 0.71310323), ('peaceful', 0.7138026), ('crawling', 0.7162924), ('wonder', 0.71687627), ('hoping', 0.7177545), ('assume', 0.71826196), ('number', 0.7184223), ('moment', 0.7187061), ('hanging', 0.7203335), ('happen', 0.72035235), ('frowning', 0.7215603), ('picking', 0.7224659), ('snitching', 0.7225072), ('wobbling', 0.72279114), ('persons', 0.72305006), ('stockings', 0.7235522), ('climbing', 0.72362924), ('hitting', 0.72398657), ('hardly', 0.7246488), ('while', 0.72514576), ('there', 0.72553396), ('forget', 0.72637385), ('reached', 0.7267917), ('airplane', 0.72726524), ('venture', 0.728104), ('sidewalk', 0.7285336), ('shocked', 0.7289816), ('excited', 0.7290697), ('careful', 0.72917134), ('counter', 0.72947735), ('carpentry', 0.7298063), ('taking', 0.7303706), ('younger', 0.7303731), ('couple', 0.7304367), ('splashed', 0.7314844), ('cupboards', 0.73187375), ('between', 0.7320692), ('pattern', 0.7325687), ('seeing', 0.73286736), ('appear', 0.73301345), ('cockeyed', 0.7331981), ('moments', 0.7339224), ('walking', 0.7339366), ('handing', 0.7339994), ('machine', 0.7343334), ('overflown', 0.7349752), ('pretty', 0.7353467), ('filling', 0.7372582), ('minutes', 0.73783886), ('billowing', 0.73792154), ('rolling', 0.73819107), ('dripping', 0.73884827), ('begging', 0.7420698), ('handles', 0.74240136), ('going', 0.7442476), ('overflow', 0.7443471), ('flooding', 0.7444775), ('knocked', 0.74525803), ('spilling', 0.74600065), ('awfully', 0.7460892), ('trousers', 0.7469972), ('copyright', 0.74737054), ('normal', 0.74755144), ('rocking', 0.74772364), ('think', 0.7479759), ('bother', 0.74813944), ('dishwater', 0.7484363), ('slipped', 0.7489998), ('falling', 0.749201), ('strapped', 0.7500655), ('beside', 0.75007874), ('still', 0.7507225), ('depends', 0.7520853), ('fountain', 0.753154), ('herself', 0.75341463), ('cutting', 0.75370574), ('first', 0.75468785), ('extent', 0.7549764), ('hundred', 0.75661325), ('overrun', 0.7567599), ('picked', 0.7573911), ('family', 0.75783986), ('overturns', 0.7585356), ('driveway', 0.759689), ('slopping', 0.7596903), ('wearing', 0.759742), ('giving', 0.75984883), ('kitchen', 0.7604015), ('sneakers', 0.76044416), ('single', 0.7613259), ('slanting', 0.7619907), ('bordered', 0.76362944), ('figure', 0.7642825), ('culprit', 0.76433694), ('amounts', 0.7643472), ('myself', 0.76440173), ('looked', 0.7649034), ('solidly', 0.7650024), ('moving', 0.76513875), ('resting', 0.76558346), ('mother', 0.7656306), ('asking', 0.7657255), ('recessed', 0.7659631), ('curtains', 0.76636064), ('thing', 0.766824), ('gotten', 0.7669734), ('raising', 0.76698995), ('towards', 0.76814806), ('salient', 0.7690022), ('nobody', 0.76903385), ('father', 0.7698798), ('healthy', 0.7711387), ('handle', 0.7717501), ('notice', 0.7722782), ('longer', 0.7733451), ('pathway', 0.7738129), ('jumping', 0.775444), ('where', 0.77551407), ('again', 0.7762293), ('means', 0.7768182), ('leaning', 0.77727276), ('other', 0.77995795), ('dumping', 0.7804664), ('reaches', 0.78077745), ('balance', 0.7808542), ('ruining', 0.78131026), ('missed', 0.7814838), ('seventy', 0.78164995), ('memorize', 0.7818602), ('animals', 0.782395), ('toppling', 0.78351676), ('might', 0.7843368), ('smiling', 0.7853924), ('lovely', 0.7862376), ('shining', 0.7869375), ('seemed', 0.7870756), ('outdoor', 0.78850704), ('donating', 0.7893477), ('bumping', 0.78950983), ('stopper', 0.78971815), ('silence', 0.78999585), ('action', 0.7911189), ('middle', 0.7917839), ('daylight', 0.79194236), ('windows', 0.7924717), ('glimpse', 0.79286444), ('charged', 0.7929343), ('starts', 0.7930219), ('being', 0.793985), ('shaking', 0.7964711), ('drooping', 0.79681), ('cabinets', 0.7986939), ('friend', 0.79885393), ('signing', 0.7988984), ('percent', 0.7989843), ('cupboard', 0.7992), ('crossed', 0.7995128), ('flowing', 0.79958063), ('pouring', 0.79973155), ('would', 0.80017567), ('sister', 0.80027694), ('fairly', 0.800335), ('counts', 0.80086416), ('coaxing', 0.80200267), ('staring', 0.8020311), ('right', 0.8021094), ('behind', 0.8024206), ('letters', 0.8027958), ('upraised', 0.80316895), ('living', 0.8034376), ('slippers', 0.80367935), ('center', 0.80399394), ('singing', 0.8051289), ('beyond', 0.8051478), ('pompadour', 0.8061817), ('blowing', 0.8063028), ('protrude', 0.8068698), ('clothes', 0.8071946), ('shoulda', 0.80736136), ('stacked', 0.8075852), ('strong', 0.80895096), ('guess', 0.80964226), ('sounds', 0.8114017), ('place', 0.81158715), ('seems', 0.81196344), ('start', 0.81239396), ('could', 0.81255543), ('nuthouse', 0.8128507), ('draperies', 0.8130772), ('bottom', 0.81429), ('ground', 0.8147564), ('active', 0.81513083), ('smashed', 0.8152413), ('slender', 0.8154321), ('pushed', 0.81576145), ('cooking', 0.81612414), ('summer', 0.81634426), ('meant', 0.8169565), ('pulled', 0.8171394), ('found', 0.81748176), ('caught', 0.8174957), ('whole', 0.81780726), ('finish', 0.8179032), ('since', 0.8181986), ('please', 0.81833154), ('after', 0.8197465), ('whisper', 0.8206681), ('messed', 0.8209787), ('minute', 0.8211242), ('dressed', 0.8214402), ('diamonds', 0.8223697), ('school', 0.8243661), ('remain', 0.82523817), ('those', 0.82654685), ('great', 0.8284026), ('tipping', 0.83062875), ('these', 0.830934), ('often', 0.8316484), ('excuse', 0.83211684), ('yelling', 0.8326286), ('caused', 0.8329479), ('moisture', 0.8331564), ('double', 0.83329886), ('define', 0.8334175), ('maybe', 0.8335947), ('weather', 0.8336971), ('paying', 0.8337184), ('forgot', 0.8361632), ('flowers', 0.8367057), ('point', 0.8371376), ('never', 0.837449), ('stuff', 0.8393498), ('bending', 0.8416108), ('doing', 0.84186935), ('comes', 0.8420541), ('leaves', 0.84252), ('gather', 0.84252983), ('across', 0.84312093), ('rolled', 0.8449447), ('least', 0.84518975), ('filled', 0.8453007), ('corner', 0.8465531), ('just', 0.85018325), ('amount', 0.8508597), ('soaking', 0.8518359), ('opened', 0.8522346), ('hahaha', 0.854804), ('damaged', 0.8551588), ('neatness', 0.85572016), ('broken', 0.8562106), ('sudden', 0.8566124), ('bring', 0.8575083), ('fellow', 0.8581688), ('slanted', 0.85842055), ('quite', 0.85881126), ('along', 0.85905796), ('thicker', 0.8593417), ('only', 0.85957086), ('nephew', 0.85961336), ('gadgets', 0.85963273), ('losing', 0.86012006), ('bemused', 0.860131), ('walkway', 0.86015445), ('haywire', 0.86211145), ('spring', 0.86237496), ('pieces', 0.8625739), ('about', 0.86379385), ('shushing', 0.86421055), ('lengths', 0.86443603), ('turns', 0.8646561), ('tripodal', 0.8662366), ('crooked', 0.86643916), ('scenery', 0.8666251), ('detail', 0.8667788), ('gushing', 0.8670348), ('crawled', 0.8682495), ('sweater', 0.86895025), ('tilting', 0.86899805), ('washing', 0.86950004), ('vaseline', 0.8705852), ('planted', 0.87087923), ('fingers', 0.8718977), ('handed', 0.87208736), ('hiding', 0.87230533), ('world', 0.8726886), ('usual', 0.8733439), ('nicely', 0.8740981), ('shows', 0.87508214), ('foliage', 0.8755981), ('danger', 0.8762737), ('gravity', 0.8765189), ('ruffling', 0.8780227), ('robbing', 0.87803924), ('climbed', 0.87845045), ('cabinet', 0.87903726), ('jogging', 0.8796554), ('kicked', 0.87981766), ('affair', 0.880426), ('looks', 0.88119704), ('tumbles', 0.8827895), ('bandaids', 0.88427365), ('pissed', 0.8852065), ('tipped', 0.8860308), ('window', 0.8861233), ('raised', 0.88620883), ('medium', 0.886803), ('killed', 0.8870891), ('sticks', 0.8881177), ('their', 0.8882293), ('laying', 0.888597), ('female', 0.8896272), ('that', 0.89049417), ('secret', 0.892755), ('close', 0.8932322), ('house', 0.8948851), ('raining', 0.89573395), ('platter', 0.8959701), ('leaking', 0.89787257), ('stand', 0.89815664), ('until', 0.8986606), ('times', 0.8990534), ('spilled', 0.90106374), ('cleanup', 0.90109277), ('toppled', 0.9022917), ('length', 0.90273523), ('then', 0.90288395), ('junior', 0.9029518), ('hidden', 0.9036737), ('laterals', 0.9038768), ('winter', 0.90526664), ('three', 0.90535206), ('takes', 0.90583515), ('above', 0.90584016), ('dinner', 0.90596753), ('titled', 0.90821576), ('small', 0.90857345), ('glasses', 0.908878), ('makes', 0.9092846), ('doubt', 0.90988684), ('mirror', 0.9117683), ('slight', 0.91210103), ('torrent', 0.91213495), ('curtain', 0.912349), ('taken', 0.9141538), ('cause', 0.914337), ('overran', 0.91446656), ('closed', 0.9146485), ('clear', 0.9166814), ('wrong', 0.9174687), ('silent', 0.91767776), ('facing', 0.9180902), ('state', 0.91964924), ('front', 0.920335), ('splash', 0.9210926), ('eating', 0.92164916), ('this', 0.92183745), ('spread', 0.92188317), ('every', 0.9225153), ('signal', 0.922729), ('fallen', 0.9247061), ('large', 0.92481834), ('ruffled', 0.9261406), ('garden', 0.926171), ('grading', 0.9282246), ('weight', 0.9305124), ('raiding', 0.9309619), ('also', 0.93210834), ('flying', 0.9322616), ('enjoy', 0.9327525), ('crazy', 0.93533975), ('string', 0.9362299), ('wants', 0.9364944), ('sleeves', 0.93663657), ('groomed', 0.93734777), ('happy', 0.9374441), ('autumn', 0.93766963), ('glance', 0.9380361), ('casement', 0.93816143), ('pudding', 0.9393787), ('crying', 0.93984103), ('years', 0.9401082), ('third', 0.94052464), ('styles', 0.9412641), ('grasses', 0.9470141), ('silly', 0.9474656), ('aware', 0.94799954), ('slower', 0.94931525), ('height', 0.9503008), ('words', 0.9505927), ('upside', 0.95205706), ('even', 0.95310205), ('awful', 0.9535246), ('ready', 0.95356053), ('both', 0.95363164), ('steady', 0.953812), ('hungry', 0.954165), ('toying', 0.9548162), ('harmed', 0.9549204), ('style', 0.9561296), ('sandals', 0.95657927), ('swiping', 0.9567443), ('speak', 0.9568766), ('gentle', 0.957072), ('hands', 0.9613357), ('hammer', 0.9625837), ('oldest', 0.96300703), ('tried', 0.9632603), ('short', 0.96352124), ('needs', 0.96540076), ('plants', 0.96671957), ('sense', 0.96715724), ('break', 0.9674129), ('commode', 0.9681172), ('moved', 0.96878403), ('knows', 0.9687945), ('depart', 0.9691542), ('colors', 0.9695869), ('pencils', 0.96963054), ('hurry', 0.9697546), ('moves', 0.96990746), ('have', 0.97076106), ('soaked', 0.9708395), ('calmly', 0.97200596), ('clogged', 0.97213274), ('build', 0.9725633), ('touch', 0.97268856), ('young', 0.9728785), ('gives', 0.9731542), ('count', 0.9740073), ('stick', 0.97452414), ('square', 0.97479814), ('curved', 0.9751817), ('allow', 0.97574383), ('heard', 0.9760598), ('bunch', 0.9764757), ('flower', 0.97751385), ('sloppy', 0.97828543), ('drawers', 0.97881097), ('round', 0.98025614), ('cover', 0.98157865), ('seven', 0.98165154), ('event', 0.9826082), ('shades', 0.9836249), ('time', 0.9837412), ('eight', 0.9840047), ('write', 0.9840623), ('work', 0.98438096), ('here', 0.98440427), ('share', 0.9845108), ('scallop', 0.984668), ('like', 0.98509127), ('wrote', 0.9872799), ('when', 0.9877276), ('toward', 0.99008274), ('ahead', 0.9927275), ('catch', 0.9929835), ('reach', 0.9929888), ('fixed', 0.9929935), ('come', 0.9934865), ('motion', 0.99358255), ('sleeved', 0.99386436), ('fancy', 0.9949897), ('washed', 0.9953191), ('force', 0.99548984), ('well', 0.99569464), ('snowing', 0.9960557), ('garage', 0.99660903), ('under', 0.9968975), ('homely', 0.9982646), ('deeply', 0.99978274), ('spells', 0.99996734), ('wetting', 1.0004445), ('plates', 1.0035105), ('breeze', 1.0036006), ('piece', 1.0055994), ('part', 1.0089968), ('woman', 1.0092752), ('coffee', 1.0096372), ('tilted', 1.0108216), ('ladder', 1.0109338), ('pompoms', 1.0116153), ('lousy', 1.0121711), ('gotta', 1.0143365), ('kiddos', 1.0152183), ('cracks', 1.0162501), ('know', 1.0163155), ('them', 1.0173092), ('watch', 1.0189024), ('dishes', 1.0194911), ('aside', 1.019987), ('throw', 1.0199875), ('light', 1.0203043), ('sound', 1.0223395), ('space', 1.0234652), ('sight', 1.0244743), ('knock', 1.0246663), ('drawn', 1.025167), ('total', 1.0253181), ('waisted', 1.0255997), ('apart', 1.0259929), ('backs', 1.026585), ('shape', 1.0268086), ('crocker', 1.0275912), ('spiral', 1.0286051), ('finger', 1.0297081), ('below', 1.0303259), ('image', 1.0304712), ('mantle', 1.0316906), ('parted', 1.032316), ('older', 1.0325719), ('with', 1.0331109), ('lucky', 1.0335281), ('closet', 1.034037), ('drive', 1.0350614), ('saved', 1.0382044), ('quick', 1.0387627), ('sleeve', 1.0391412), ('topple', 1.0400763), ('throat', 1.0405666), ('girls', 1.0406996), ('apples', 1.0415279), ('button', 1.0419389), ('sorry', 1.0439987), ('valance', 1.0440978), ('badly', 1.0449815), ('johnny', 1.0459199), ('lower', 1.0459832), ('fact', 1.0467751), ('spigots', 1.0467794), ('drying', 1.0469741), ('anklets', 1.047479), ('upset', 1.0485433), ('bitchy', 1.0488558), ('scene', 1.0494366), ('shorts', 1.0507269), ('bushes', 1.0536723), ('want', 1.0543151), ('washes', 1.0559614), ('wiping', 1.0563833), ('mouths', 1.057672), ('upward', 1.0585623), ('child', 1.0587511), ('clause', 1.0590028), ('some', 1.0603313), ('climbs', 1.0608779), ('mixed', 1.0617647), ('need', 1.0637197), ('cookies', 1.0639797), ('mean', 1.0639975), ('will', 1.0649332), ('and', 1.0651468), ('falled', 1.0651853), ('plain', 1.0667965), ('floors', 1.0671614), ('keeps', 1.0674818), ('they', 1.0675555), ('pardon', 1.0685707), ('same', 1.0688367), ('board', 1.0712137), ('waving', 1.0726757), ('faucets', 1.074211), ('rubber', 1.074622), ('galore', 1.0748304), ('fresh', 1.0748388), ('water', 1.0752349), ('tries', 1.0757647), ('kind', 1.078549), ('alike', 1.0799762), ('table', 1.0818714), ('broke', 1.0837047), ('rest', 1.0855703), ('worst', 1.0859604), ('turn', 1.0871019), ('green', 1.0880264), ('what', 1.0884832), ('tired', 1.0885485), ('blouse', 1.0898739), ('done', 1.0902938), ('drawer', 1.0916225), ('dream', 1.0919781), ('plus', 1.0921108), ('laugh', 1.0921397), ('gloves', 1.0931004), ('baking', 1.0931411), ('forth', 1.0940833), ('shrubs', 1.0954504), ('saucers', 1.0956212), ('paper', 1.0957838), ('good', 1.0958984), ('women', 1.0961367), ('quiet', 1.1007979), ('turkey', 1.1013714), ('adult', 1.1014574), ('steps', 1.1020159), ('straps', 1.1024187), ('show', 1.1024487), ('sort', 1.1025645), ('grassy', 1.1037437), ('drops', 1.1053421), ('came', 1.1062541), ('faced', 1.1067315), ('clean', 1.1095175), ('angry', 1.1109345), ('apron.', 1.1113538), ('hairdo', 1.1117355), ('next', 1.111736), ('cross', 1.111834), ('dozen', 1.1125954), ('make', 1.1144443), ('untied', 1.1150486), ('thank', 1.11608), ('take', 1.1162779), ('falls', 1.1192175), ('spills', 1.1199073), ('legged', 1.1201872), ('ones', 1.1204877), ('back', 1.1206448), ('must', 1.1209099), ('jersey', 1.1209409), ('made', 1.121138), ('seen', 1.121634), ('pantry', 1.1222308), ('fully', 1.1232111), ('look', 1.1234941), ('cloudy', 1.1241804), ('blown', 1.1245651), ('gone', 1.1245935), ('find', 1.1255608), ('drapes', 1.1257235), ('real', 1.1272227), ('color', 1.1286752), ('steal', 1.1291035), ('from', 1.1295588), ('left', 1.1303809), ('type', 1.1305361), ('over', 1.133473), ('goes', 1.133603), ('okay', 1.1355598), ('many', 1.1365258), ('tennis', 1.1370502), ('puddle', 1.137505), ('crack', 1.1382682), ('shall', 1.1386322), ('tell', 1.1413037), ('dirty', 1.1420845), ('home', 1.1424367), ('aprons', 1.143817), ('give', 1.1471653), ('washer', 1.147323), ('minds', 1.1477747), ('payed', 1.1480479), ('sure', 1.1484041), ('form', 1.1500472), ('tasty', 1.1518109), ('floor', 1.1529969), ('ample', 1.154196), ('dishpan', 1.154729), ('much', 1.1561573), ('long', 1.1563258), ('puffed', 1.1564242), ('empty', 1.1574631), ('angle', 1.159362), ('spigot', 1.1605152), ('lunch', 1.1605558), ('said', 1.1613498), ('went', 1.1667509), ('year', 1.1669861), ('been', 1.1680653), ('stop', 1.1693612), ('very', 1.1701301), ('plant', 1.1701782), ('hedges', 1.1711838), ('area', 1.1717143), ('shove', 1.1719785), ('cease', 1.1728531), ('cookie', 1.1738635), ('mommy', 1.1746014), ('rooms', 1.1757368), ('more', 1.1783966), ('mercy', 1.1786952), ('dress', 1.1806159), ('saucer', 1.1810727), ('does', 1.18128), ('sledge', 1.1832863), ('shiny', 1.1848149), ('each', 1.1871471), ('dance', 1.1882502), ('past', 1.1886573), ('heck', 1.189517), ('hand', 1.1895726), ('eaten', 1.1896974), ('momma', 1.1912168), ('tripod', 1.1925868), ('wreath', 1.1925937), ('devil', 1.1958814), ('upper', 1.197524), ('were', 1.1977193), ('hold', 1.1997318), ('play', 1.2007496), ('shush', 1.2017325), ('plate', 1.2027909), ('your', 1.2033933), ('took', 1.2043825), ('towels', 1.2052584), ('life', 1.206218), ('chaos', 1.206294), ('blind', 1.2081069), ('soon', 1.208571), ('twins', 1.2091515), ('nice', 1.209358), ('full', 1.2097952), ('help', 1.2098738), ('ours', 1.2113516), ('mouth', 1.2116858), ('smile', 1.2118728), ('outer', 1.212116), ('combed', 1.2128761), ('hope', 1.2142781), ('tumble', 1.2155564), ('line', 1.2181329), ('read', 1.2197082), ('wreck', 1.2201997), ('shirt', 1.2228339), ('than', 1.2229846), ('climb', 1.2258149), ('hickey', 1.2259189), ('call', 1.2273433), ('crash', 1.2281637), ('head', 1.2303069), ('shelf', 1.2313173), ('chair', 1.2325253), ('glass', 1.232634), ('bench', 1.2333378), ('doors', 1.2333784), ('knees', 1.2334479), ('drugs', 1.2334784), ('days', 1.23386), ('fine', 1.2349252), ('ever', 1.2351248), ('inner', 1.2358181), ('hard', 1.2388296), ('horse', 1.2394861), ('else', 1.2400542), ('trunk', 1.2414201), ('index', 1.2419268), ('haste', 1.2426082), ('karen', 1.2427869), ('four', 1.2434283), ('whom', 1.2451609), ('main', 1.2453912), ('drain', 1.2462806), ('shoes', 1.246303), ('miss', 1.2469507), ('bloom', 1.247939), ('cloth', 1.2489314), ('faucet', 1.2491294), ('near', 1.2507979), ('stools', 1.2512223), ('pants', 1.2522767), ('hears', 1.254945), ('honey', 1.2556839), ('wiped', 1.2562125), ('huhuh', 1.2572336), ('skirt', 1.2581476), ('want.', 1.2587999), ('swung', 1.2593589), ('dried', 1.2593676), ('sinks', 1.2594267), ('side', 1.2600613), ('but', 1.2604231), ('seem', 1.2621613), ('sissy', 1.2626312), ('trees', 1.2638091), ('knew', 1.2640206), ('the', 1.2650335), ('five', 1.2668554), ('yeah', 1.2676666), ('boots', 1.2680777), ('mind', 1.269108), ('half', 1.2698213), ('brisk', 1.2703744), ('spill', 1.2708819), ('gave', 1.2731777), ('pass', 1.2733353), ('wahoo', 1.2743442), ('birds', 1.2744943), ('nails', 1.2747903), ('fouled', 1.2750798), ('elbow', 1.2759895), ('clock', 1.2775657), ('stay', 1.2777092), ('book', 1.277817), ('puts', 1.2779367), ('face', 1.280823), ('keep', 1.2817936), ('flood', 1.2829374), ('says', 1.2834327), ('sunny', 1.2855082), ('lost', 1.2873169), ('room', 1.2887781), ('open', 1.2892236), ('hell', 1.2896752), ('girl', 1.2911905), ('view', 1.291648), ('strap', 1.2929282), ('damn', 1.2934167), ('grass', 1.2961073), ('goody', 1.2980865), ('lots', 1.2987922), ('noise', 1.2992913), ('runs', 1.3002222), ('sizes', 1.3006697), ('cream', 1.3014913), ('base', 1.3015496), ('sees', 1.3027756), ('game', 1.3035519), ('told', 1.3037345), ('snowy', 1.3055581), ('drop', 1.3063928), ('down', 1.3079717), ('pull', 1.310577), ('pick', 1.3123057), ('kept', 1.3130556), ('gets', 1.3140955), ('kids', 1.3158183), ('high', 1.3159584), ('hear', 1.3167173), ('wait', 1.3171011), ('hers', 1.3176367), ('away', 1.3211788), ('lose', 1.321411), ('fall', 1.3239237), ('basin', 1.3239768), ('wipes', 1.3247302), ('sets', 1.3262621), ('word', 1.3276162), ('puffy', 1.3282791), ('wowie', 1.3324226), ('mess', 1.3341122), ('socks', 1.3342315), ('wrist', 1.335773), ('apron', 1.338949), ('page', 1.3405346), ('mouse', 1.3405885), ('nine', 1.3424714), ('tilts', 1.3441603), ('hurt', 1.3452384), ('send', 1.345622), ('wife', 1.3486019), ('beat', 1.349753), ('walk', 1.3524401), ('road', 1.35419), ('one', 1.3547877), ('ends', 1.3549794), ('care', 1.3552101), ('lady', 1.3575876), ('sugar', 1.359949), ('into', 1.3603361), ('now', 1.36283), ('jelly', 1.3653276), ('ankle', 1.3654249), ('bowls', 1.3659949), ("boy's", 1.3695143), ('boys', 1.3695143), ('and.', 1.3698859), ('talk', 1.3714986), ('upend', 1.3740586), ('weeds', 1.3742374), ('step', 1.3764327), ('billy', 1.3777505), ('kill', 1.3783945), ('dries', 1.3798958), ('all', 1.3812569), ('pair', 1.3821466), ('darn', 1.3824525), ('mamas', 1.3826903), ('sign', 1.3865064), ('land', 1.3959776), ('grab', 1.3961594), ('spout', 1.3964747), ('panes', 1.3994803), ('spank', 1.3997593), ('dumb', 1.4055115), ('towel', 1.4060323), ('trip', 1.4091966), ('too.', 1.4103649), ('mini', 1.4123194), ('fast', 1.4128704), ('quit', 1.4131913), ('hedge', 1.4142011), ('busy', 1.4146901), ('draw', 1.4169863), ('foot', 1.4223969), ('tops', 1.4227862), ('dump', 1.4252743), ('stool', 1.4294568), ('flat', 1.4355801), ('sakes', 1.4362255), ('knobs', 1.4388832), ('way', 1.440928), ('cocoa', 1.4436783), ('laid', 1.4461031), ('slip', 1.4480641), ('not', 1.4501693), ('poor', 1.452114), ('onto', 1.4521376), ('tippy', 1.4532231), ('whew', 1.4646919), ('for', 1.4647561), ('edge', 1.4647809), ('weak', 1.4696599), ('food', 1.4712963), ('drew', 1.4720244), ('size', 1.4742932), ('dear', 1.4761401), ('shut', 1.4811373), ('dryer', 1.4816949), ('copy', 1.4866883), ('kick', 1.486962), ('path', 1.487654), ('uhoh', 1.4896686), ('tied', 1.4923241), ('daft', 1.4979383), ('was', 1.4982593), ('hits', 1.4983544), ('geese', 1.5050322), ('baby', 1.5063701), ('door', 1.5067372), ('shrub', 1.508459), ('warm', 1.5099392), ('wall', 1.5099821), ('sill', 1.5182962), ('the.', 1.5199779), ('prong', 1.5244247), ('bent', 1.5257748), ('yards', 1.5258348), ('gosh', 1.5268397), ('yet', 1.5268977), ('fell', 1.5304992), ('mama', 1.5311289), ('see.', 1.5401683), ('meal', 1.540467), ('trap', 1.5412813), ('use', 1.5418706), ('ties', 1.542254), ('knots', 1.5422983), ('say', 1.5432042), ('sink', 1.5472751), ('see', 1.5512257), ('sole', 1.5554934), ('get', 1.5558252), ('race', 1.557737), ('male', 1.559076), ('pool', 1.5598022), ('put', 1.5612332), ('feet', 1.5636744), ('can', 1.5639074), ('roof', 1.5639385), ('too', 1.5734271), ('rear', 1.5744244), ('wing', 1.581654), ('you', 1.5874797), ('legs', 1.5875144), ('day', 1.5906519), ('okey', 1.5921118), ('uhuh', 1.5925218), ('daze', 1.593761), ('cook', 1.600437), ('yard', 1.6034238), ('new', 1.6040224), ('set', 1.6071501), ('wind', 1.6078799), ('bang', 1.609533), ('till', 1.6114714), ('tilt', 1.6170931), ('plug', 1.6171764), ('skin', 1.61745), ('yes', 1.624787), ('dish', 1.6257921), ('hoyle', 1.6267421), ('jerk', 1.6288753), ('has', 1.6293635), ('did', 1.6294812), ('june', 1.6302722), ('got', 1.6326016), ('wash', 1.6338509), ('duck', 1.6369704), ('wood', 1.63816), ('junk', 1.6429703), ('nose', 1.6451824), ('are', 1.6485709), ('slop', 1.65292), ('seat', 1.6530968), ('tips', 1.6554683), ('tree', 1.6584969), ('tidy', 1.6586425), ('tear', 1.6645061), ('slid', 1.6649432), ('had', 1.6658366), ('ball', 1.6670043), ('loud', 1.6675137), ("mom's", 1.6694262), ('scan', 1.6695721), ('pour', 1.6727175), ('may', 1.6765707), ('run', 1.6775252), ('own', 1.6821085), ('plum', 1.686181), ('any', 1.6863061), ('leak', 1.6930658), ('cake', 1.6940329), ('wipe', 1.7044646), ('neck', 1.7068379), ('calm', 1.7118592), ('lips', 1.7177093), ('end', 1.718601), ('cups', 1.7191141), ('huh.', 1.7197118), ('bush', 1.7201662), ('shoe', 1.7247552), ('its', 1.725145), ('leaf', 1.7251807), ('lawn', 1.7265884), ('sue.', 1.7285419), ('out', 1.7287105), ('our', 1.7288264), ('him', 1.730089), ('pipe', 1.7340522), ('hair', 1.7341594), ('rows', 1.7416468), ('toes', 1.749055), ('man', 1.7511206), ('dokey', 1.7530514), ('try', 1.7533469), ('knee', 1.7604184), ('far', 1.7674621), ('why', 1.7677584), ('who', 1.7689615), ('two', 1.7766796), ('saw', 1.7777897), ('rice', 1.7853907), ('ask', 1.7861818), ('bowl', 1.7900509), ('lane', 1.8007497), ('snow', 1.8016422), ('hey', 1.8031138), ('his', 1.8056154), ('yep', 1.8077674), ('beep', 1.8101742), ('big', 1.8109831), ('she', 1.8118273), ('let', 1.812962), ('pots', 1.8178077), ('pans', 1.8257222), ('her', 1.8268137), ('lot', 1.8290079), ('heel', 1.8376583), ('wove', 1.8399072), ('bet', 1.840259), ('verb', 1.8423191), ('act', 1.851707), ("'ll", 1.8551494), ('hiss', 1.8620013), ('guy', 1.8640742), ('few', 1.8685517), ('off', 1.8754896), ('old', 1.880513), ('top', 1.8840092), ('sash', 1.8868515), ('bad', 1.891091), ('how', 1.8912766), ('suds', 1.8928473), ('eggs', 1.893867), ('jars', 1.9016241), ('leg.', 1.9219997), ('lids', 1.9239947), ('curl', 1.9278183), ("'ve", 1.9286153), ("'re", 1.93023), ('ails', 1.9355912), ('hit', 1.9397674), ('deaf', 1.9502012), ('fun', 1.9619684), ('son', 1.9627398), ('etch', 1.9643787), ('ten', 1.9761521), ('kid', 1.9764466), ('hmm', 1.9918756), ('ran', 1.99424), ('boy', 1.9944993), ('mad', 1.9998577), ('cut', 2.0007548), ('lay', 2.0139518), ('mom', 2.015803), ('job', 2.0218894), ('ajar', 2.0235822), ('bit', 2.0285814), ('dad', 2.0424204), ('wow', 2.0470064), ('flue', 2.0652883), ('huh', 2.0741115), ('low', 2.0768962), ('sad', 2.0858855), ('eat', 2.12722), ('wee', 2.1586716), ('ago', 2.1792724), ('eve', 2.1793103), ('gee', 2.1819737), ('tie', 2.1940312), ('age', 2.1982384), ('bed', 2.202106), ('nor', 2.2167706), ('arm', 2.2206051), ('lad', 2.22209), ('or.', 2.227121), ('jar.', 2.2467606), ('hot', 2.2495487), ('ate', 2.2519863), ('air', 2.2553341), ('tup', 2.2603548), ('so.', 2.2604046), ('tip', 2.2625732), ('god', 2.2679138), ('tap', 2.2757578), ('shh', 2.2960072), ('dry', 2.3042972), ('leg', 2.3097982), ('mhm', 2.3207629), ('wet', 2.3300307), ('ass', 2.3353295), ('eye', 2.3577397), ('no.', 2.364199), ('tin', 2.3716648), ("'s.", 2.3737276), ('tad', 2.384532), ('gal', 2.38792), ('dug', 2.392168), ('jam', 2.4315772), ('ret', 2.4366255), ('gon', 2.4559832), ('cup', 2.4610846), ('sun', 2.4858596), ('hm.', 2.4933844), ('toe', 2.5510352), ('tub', 2.5708475), ('pan', 2.5736852), ('oh.', 2.596162), ('lid', 2.6134574), ('tea', 2.6155963), ('jar', 2.6174092), ('rag', 2.6289022), ('lip', 2.686925), ('fry', 2.7105896), ('mop', 2.7310603), ('ah.', 2.7591417), ('so', 2.866878), ('it', 2.8743865), ('or', 2.9529119), ('as', 2.9764423), ('in', 3.0260985), ('to', 3.044186), ('oop', 3.0591154), ('of', 3.0693176), ('is', 3.089255), ('do', 3.1416345), ('if', 3.1905713), ('a', 3.2806578), ('on', 3.3638358), ('th', 3.4101539), ('ssh', 3.4164813), ('go', 3.4416175), ('be', 3.4683225), ('at', 3.4845138), ("'s", 3.4883757), ("n't", 3.565551), ('nt', 3.565551), ('we', 3.582459), ('oy', 3.5900097), ('he', 3.6314945), ('me', 3.6414065), ('no', 3.732416), ('up', 3.7625241), ('my', 3.7905352), ('us', 3.795824), ("'d", 3.7969108), ('by', 3.8022363), ('wo', 3.8452363), ('s', 3.8470917), ('hm', 3.873371), ('i', 3.9522903), ('+', 4.050339), ('ta', 4.1001415), ('am', 4.125909), ('oh', 4.1351843), ('ah', 4.140119), ('ya', 4.2842236), ("'m", 4.289837), ('ho', 4.332503), ('an', 4.3691225), ('ca', 4.587303), ('ai', 4.62603), ('v', 4.6846504), ('na', 4.8349156), ('tv', 5.0770693), ('a.', 5.0978603), ('i.', 5.378265), ('mm', 5.71878)

A.0.3. Wikipedia skip-gram models with subword information

See P. Bojanowski et al[**16**] for the implementation details of these models.

A.0.3.1. *French model, Picnic corpus full vocabulary*

('et', 1.9464904), ('ainsi', 2.0589948), ('même', 2.0763729), ('mais', 2.1223426), ('qui', 2.142369), ('de', 2.1593318), ('alors', 2.165489), ('ailleurs', 2.1976583), ('le', 2.2264657), ('également', 2.2801008), ('que', 2.2897453), ('aussi', 2.2951207), ('encore', 2.3001897), ('-ce', 2.3128672), ('ce', 2.3128672), ('ou', 2.3182867), ('autre', 2.3550937), ('dont', 2.3595033), ('en', 2.381858), ('donc', 2.3872576), ('pour', 2.3901827), ('les', 2.3903227), ('la', 2.3923542), ('toujours', 2.406734), ('seulement', 2.4130855), ('comme', 2.4427743), ('un', 2.451455), ('qu'', 2.4527307), ("qu'", 2.4527307), ('tout', 2.4618266), ('avec', 2.4654016), ('est', 2.4816206), ('à', 2.5028286), ('exemple', 2.5090191), ('l'', 2.5282123), ("l'", 2.5282123), ('plus', 2.5524206), ('une', 2.5594003), ('certainement', 2.5663438), ('autres', 2.5713882), ('bien', 2.5717103), ('il', 2.5729773), ('cette', 2.5934763), ('fait', 2.6083648), ('ces', 2.6200137), ('où', 2.6219847), ('des', 2.6418831), ('dans', 2.6472228), ('apparemment', 2.6486146), ('finalement', 2.6497264), ('là', 2.6638026), ('-là', 2.6638026), ('son', 2.6702745), ('puis', 2.6847482), ('plutôt', 2.7040937), ('évidemment', 2.7052438), ('deux', 2.7104685), ('malheureusement', 2.7117276), ('lui', 2.7134113), ('sorte', 2.7214758), ('quand', 2.726988), ('d'', 2.7346551), ("d'", 2.7346551), ('après', 2.7383204), ('déjà', 2.742083), ('c'', 2.7502604), ("c'", 2.7502604), ('simplement', 2.7504344), ('du', 2.7728686), ('par', 2.7747161), ('autant', 2.7763257), ('a', 2.7806818), ('tandis', 2.7832386), ('pas', 2.7838073), ('ne', 2.7839036), ('peu', 2.7935393), ('puisqu'', 2.7968247), ('soit', 2.798322), ('elle', 2.8023517), ('lequel', 2.813454), ('beaucoup', 2.8300722), ('sa', 2.8362343), ('y', 2.8544896), ('quoique', 2.8614833), ('rappelle', 2.8630095), ('côté', 2.8773465), ('probablement', 2.8783755), ('ses', 2.8826444), ('eux', 2.8875968), ('si', 2.8895211), ('souvent', 2.891275), ('entre', 2.8948417), ('juste', 2.8952646), ('très', 2.8982835), ('suppose', 2.9009507), ('loin', 2.906982), ('certain', 2.9096172), ('avant', 2.9145787), ('parce', 2.9151025), ('leur', 2.9168556), ('quelque', 2.9195518), ('n'', 2.923904), ('autrement', 2.9267123), ('fois', 2.929205), ('vu', 2.9293053), ('sans', 2.9350286), ('appelle', 2.9350703), ('trois', 2.9422457), ('sur', 2.9541054), ('quelqu'un', 2.9597769), ("quelqu'un", 2.9597769), ('quelques', 2.9652772), ('presque', 2.9654522), ('tous', 2.9747295), ('possible', 2.9819505), ('doute', 2.9879882), ('conduit', 2.9884949), ('quant', 2.9921076), ('toute', 3.0043077), ('part', 3.008554), ('longtemps', 3.0191522), ('vient', 3.0203679), ('peut', 3.0223835), ('ceux', 3.0251038), ('parle', 3.0277638), ('moins', 3.0291889), ('être', 3.0302103), ('instant', 3.0308573), ('dire', 3.032967), ('vraiment', 3.0472786), ('partie', 3.052671), ('imagine', 3.0545976), ('pense', 3.0596356), ('était', 3.0657575), ('sait', 3.077216), ('supposé', 3.085674), ('voit', 3.0880466), ('sont', 3.0915534), ('semble', 3.0919597), ('avoir', 3.0931904), ('veut', 3.0953643), ('assez', 3.0964456), ('temps', 3.0992725), ('faire', 3.101806), ('pendant', 3.1018407), ('quoiqu'', 3.1053424), ('jusqu'', 3.1060035), ('ici', 3.1060753), ('arrive', 3.107363), ('sûr', 3.1087809), ('près', 3.1106236), ('va', 3.1163177), ('été', 3.1228688), ('serait', 3.1351304), ('on', 3.1356761), ('alentours', 3.1486645), ('avait', 3.1492097), ('tellement', 3.1494353), ('grand-père', 3.1515064), ('sujet', 3.1622097), ('intéressé', 3.1640327), ('se', 3.1645818), ('toutes', 3.1719508), ('vrai', 3.1743488), ('ensemble', 3.175583), ('retenir', 3.1759894), ('suivi', 3.1780965), ('cas', 3.178435), ('quoi', 3.178786), ('doit', 3.1801324), ('dizaine', 3.185609), ('quatre', 3.1886857), ('principalement', 3.1978703), ('penser', 3.2013943), ('pourrait', 3.2026472), ('certaine', 3.20593), ('complètement', 3.218952), ('voilà', 3.2212956), ('dit', 3.2247627), ('ils', 3.224924), ('-ils', 3.224924), ('montrer', 3.2274256), ('passant', 3.2316544), ('ont', 3.2318025), ('proche', 3.2445757), ('possiblement', 3.246641), ('leurs', 3.2503772), ('cinq', 3.2514844), ('rien', 3.2600815), ('continuer', 3.2681472), ('gardé', 3.2689803), ('faut', 3.2699509), ('écrit', 3.2708318), ('s'', 3.2765512), ('années', 3.2776248), ('continuant', 3.2779472), ('sortir', 3.282483), ('prête', 3.3021986), ('passe', 3.3051784), ('tient', 3.310385), ('mieux', 3.312367), ('pris', 3.3194056), ('dirait', 3.3194516), ('au-dessus', 3.323169), ('cherche', 3.3245327), ('aurait', 3.330182), ('petite', 3.3310168), ('partiellement', 3.334226), ('globalement', 3.3383298), ('veille', 3.3446279), ('continue', 3.3525374), ('parler', 3.3546379), ('répète', 3.3613691), ('jeune', 3.3657982), ('coté', 3.368581), ('douzaine', 3.37462), ('prendre', 3.3785157), ('essayer', 3.3793037), ('oublié', 3.3843772), ('donner', 3.3944404), ('peut-être', 3.397303), ('conjointe', 3.3989606), ('chose', 3.3996105), ('difficulté', 3.404031), ('autour', 3.4115212), ('besoin', 3.4142008), ('manque', 3.414747), ('souvient', 3.4149828), ('devrait', 3.4178946), ('peine', 3.4191093), ('commence', 3.4306803), ('finit', 3.4370914), ('appelles', 3.4402237), ('servir', 3.4429686), ('bon', 3.4464667), ('essaie', 3.446858), ('voir', 3.4531043), ('rassemblée', 3.461274), ('petit', 3.4640465), ('diraient', 3.4680972), ('soutient', 3.472774), ('selon', 3.4728727), ('sérieuse', 3.4736147), ('perdre', 3.4736857), ('remonte', 3.4758823), ('sous', 3.4784892), ('regarde', 3.479206), ('tire', 3.4801881), ('disons', 3.4831), ('personne', 3.488936), ('proximité', 3.4894469), ('vont', 3.4896717), ('laissé', 3.4930248), ('place', 3.493571), ('bout', 3.4943163), ('mis', 3.4951324), ('trop', 3.4952605), ('disparait', 3.4993305), ('mal', 3.500323), ('pouvoir', 3.5066586), ('tantôt', 3.5106926), ('devant', 3.5117748), ('comment', 3.5179632), ('apporté', 3.5187085), ('fallu', 3.5236058), ('profite', 3.5247138), ('vers', 3.5249984), ('voient', 3.528311), ('présumée', 3.5315995), ('suit', 3.5323262), ('aperçoit', 3.5329309), ('grands', 3.5342555), ('gens', 3.5352008), ('ressemble', 3.539061), ('restait', 3.541085), ('lisant', 3.541492), ('prend', 3.5447714), ('font', 3.552294), ('ramené', 3.5526836), ('nouvelles', 3.5578132), ('espère', 3.5608625), ('complété', 3.5612874), ('pique-nique', 3.5652552), ('recommence', 3.5666761), ('gros', 3.5670931), ('père', 3.5678022), ('soir', 3.5692098), ('simples', 3.5702007), ('demande', 3.574039), ('ressembler', 3.5751297), ('empêcher', 3.575398), ('dessus', 3.5756116), ('quelle', 3.5756361), ('mettre', 3.5786908), ('accompagné', 3.5805223), ('non', 3.586371), ('nommer', 3.5867743), ('prépare', 3.5889266), ('intérieur', 3.5895653), ('livre', 3.5954976), ('aide', 3.5957348), ('met', 3.598277), ('semblent', 3.6006978), ('rentrer', 3.603929), ('choses', 3.6051664), ('belle', 3.6067622), ('regarder', 3.6071672), ('situation', 3.6090393), ('bonne', 3.6139467), ('réussir', 3.6175125), ('porte', 3.617647), ('impression', 3.6221306), ('nom', 3.6223717), ('souvenir', 3.624067), ('aider', 3.6246965), ('voyons', 3.6256392), ('caché', 3.627552), ('projette', 3.6303167), ('travaille', 3.6316743), ('sens', 3.63493), ('sert', 3.6370609), ('dépasse', 3.6428006), ('marqué', 3.6436195), ('courant', 3.6441286), ('aller', 3.6532588), ('français', 3.653737), ('paraissent', 3.6557758), ('voyez', 3.661013), ('histoire', 3.662928), ('dedans', 3.6706197), ('remonté', 3.6712027), ('petites', 3.6716495), ('préparer', 3.6740024), ('lit', 3.6778824), ('derrière', 3.6860356), ('étaient', 3.6939008), ('décrive', 3.6979632), ('nous', 3.7030604), ('sort', 3.708195), ('apporter', 3.710136), ('terme', 3.7107208), ('long', 3.7115376), ('pelleter', 3.7119262), ('mari', 3.712659), ('heureux', 3.7168515), ('ouvre', 3.7180958), ('partagent', 3.7196765), ('marcher', 3.7208898), ('tête', 3.7252765), ('grosse', 3.7266138), ('répondent', 3.7281582), ('remplie', 3.7359915), ('connais', 3.7459714), ('petits', 3.746262), ('magnifique', 3.746623), ('âgé', 3.7476156), ('dise', 3.7496357), ('plan', 3.7520106), ('cachée', 3.754532), ('grand', 3.7562654), ('mots', 3.7563841), ('courte', 3.7582042), ('auprès', 3.7607439), ('propriété', 3.7631683), ('fonctionner', 3.7635076), ('correct', 3.767241), ('mère', 3.767629), ('calme', 3.7682135), ('identifier', 3.769425), ('cité', 3.7700765), ('parents', 3.7751193), ('tiré', 3.7786956), ('enlevé', 3.7793617), ('doucement', 3.7794704), ('ombre', 3.7802198), ('retrait', 3.7819753), ('frère', 3.7852657), ('écoutant', 3.7874126), ('doivent', 3.787917), ('entrée', 3.7960916), ('profitent', 3.7974412), ('envoie', 3.7979467), ('construire', 3.8032105), ('prennent', 3.8116872), ('enlevées', 3.8146293), ('dis', 3.8165188), ('fabrique', 3.820042), ('sais', 3.8233712), ('garçon', 3.8296657), ('vaut', 3.829832), ('lecture', 3.8305867), ('jouer', 3.8322818), ('femme', 3.8427658), ('faisait', 3.843319), ('longues', 3.8447578), ('enfant', 3.850816), ('bord', 3.8525624), ('préparent', 3.8546085), ('financièrement', 3.8573596), ('parlent', 3.8581321), ('propriétaires', 3.8601387), ('transporte', 3.865523), ('fillette', 3.8673067), ('soulevée', 3.8760965), ('mains', 3.8777103), ('unifamiliale', 3.8801227), ('vois', 3.8843775), ('large', 3.885283), ('attachés', 3.8860912), ('moi', 3.8918781), ('-moi', 3.8918781), ('camouflent', 3.8925328), ('enfants', 3.8936076), ('face', 3.8943377), ('ça', 3.901796), ('contrôle', 3.9040773), ('personnes', 3.9041743), ('mettons', 3.9049037), ('comprends', 3.9060285), ('fil', 3.9112818), ('monde', 3.9125454), ('je', 3.914881), ('couple', 3.9225252), ('fond', 3.9248402), ('écoute', 3.9258318), ('guider', 3.9260142), ('patente', 3.9291284), ('chez', 3.9373), ('ouvert', 3.939628), ('main', 3.9407191), ('homme', 3.941664), ('nord-ouest', 3.9469936), ('journée', 3.9527042), ('courir', 3.9534905), ('veux', 3.9539397), ('sœur', 3.9557426), ('détendu', 3.9600794), ('attache', 3.9616017), ('installés', 3.9636657), ('campagne', 3.9717586), ('pensais', 3.9734626), ('bougent', 3.9741416), ('voulais', 3.9751527), ('fils', 3.9772089), ('souriants', 3.9773192), ('souviens', 3.9795222), ('haut', 3.9809752), ('terrain', 3.9838817), ('plantant', 3.9852376), ('cherchais', 3.986069), ('vide', 3.9876115), ('pays', 3.9901943), ('m'', 4.001781), ('sourit', 4.0057607), ('riches', 4.008344), ('personnages', 4.0130053), ('écouter', 4.0146704), ('sortie', 4.016506), ('vidait', 4.019826), ('marche', 4.0227003), ('co-propriétaires', 4.036009), ('sœurette', 4.038263), ('familial', 4.0403237), ('traverse', 4.0467324), ('horizon', 4.047137), ('âge', 4.048046), ('chemin', 4.048953), ('vous', 4.04933), ('-vous', 4.04933), ('pieds', 4.0529633), ('train', 4.054253), ('fermé', 4.0562096), ('fille', 4.056588), ('descend', 4.0585613), ('essoufflé', 4.06023), ('maison', 4.061796), ('détail', 4.0649147), ('monsieur', 4.068489), ('appartement', 4.069013), ('payent', 4.069324), ('bête', 4.0701423), ('voler', 4.0703588), ('attends', 4.0707846), ('envoient', 4.073259), ('rapport', 4.0766416), ('écoutent', 4.078094), ('terre', 4.078814), ('construit', 4.0824676), ('ouest', 4.0845437), ('gamin', 4.0885487), ('imite', 4.092131), ('avance', 4.092203), ('spécifier', 4.09721), ('casquette', 4.097385), ('hot-dog', 4.100403), ('secondaire', 4.1012945), ('joue', 4.1023574), ('supporté', 4.1037445), ('couler', 4.103865), ('croisées', 4.1100383), ('sure', 4.115542), ('aise', 4.1181984), ('arrière', 4.11931), ('voiture', 4.124517), ('euh', 4.124994), ('gars', 4.1261744), ('adultes', 4.126498), ('souliers', 4.1277785), ('majestueux', 4.1288605), ('modèle', 4.132399), ('pelles', 4.135927), ('longs', 4.1389747), ('espadrilles', 4.14672), ('mon', 4.1468987), ('boite', 4.1495705), ('attaché', 4.1501546), ('viens', 4.1529326), ('verdure', 4.156232), ('fou', 4.1565194), ('j'', 4.1565576), ("j'", 4.1565576), ('couverture', 4.1622458), ('prêt', 4.162358), ('savez', 4.1632605), ('scène', 4.1661553), ('flotter', 4.1675925), ('lire', 4.1697826), ('maisons', 4.1711874), ('courtes', 4.1726975), ('peinture', 4.1761336), ('debout', 4.176313), ('garçons', 4.178648), ('simple', 4.186824), ('affaire', 4.1929264), ('peux', 4.1962767), ('véhicule', 4.1982865), ('sandales', 4.2004404), ('femmes', 4.203592), ('bateau', 4.205113), ('chalet', 4.205811), ('gauche', 4.2081194), ('entourée', 4.209617), ('plaisanciers', 4.210301), ('cheval', 4.2120533), ('baignait', 4.213673), ('notre', 4.2183523), ('adjacent', 4.219766), ('voguent', 4.220802), ('complètes', 4.2220073), ('court', 4.227949), ('vais', 4.229425), ('verse', 4.235103), ('spéciales', 4.2377067), ('carpette', 4.2384686), ('reproduction', 4.2436333), ('promène', 4.243914), ('dessous', 4.2472725), ('adulte', 4.24806), ('sereins', 4.2485595), ('animal', 4.2527204), ('photo', 4.254191), ('habitent', 4.2554708), ('nationalistes', 4.2600403), ('loué', 4.2638373), ('phrases', 4.2645664), ('habillé', 4.2662354), ('promenade', 4.267413), ('habille', 4.274622), ('paysage', 4.2818346), ('résidence', 4.2873497), ('couverte', 4.2907686), ('stationnée', 4.291485), ('bois', 4.291716), ('sac', 4.293891), ('madame', 4.295532), ('orchard', 4.2962465), ('taquine', 4.2980156), ('vole', 4.2988763), ('bébé', 4.2995925), ('attrapé', 4.2997527), ('jeté', 4.300171), ('voyais', 4.301825), ('droite', 4.3021407), ('verre', 4.3030305), ('souffle', 4.304), ('mouler', 4.3044825), ('planté', 4.304873), ('rivage', 4.3077292), ('serviette', 4.3096232), ('belles', 4.309955), ('louis', 4.310721), ('affolés', 4.311417), ('sol', 4.3123107), ('monte', 4.312682), ('noël', 4.3185425), ('pichet', 4.3240976), ('ai', 4.326432), ('sème', 4.3292694), ('conjoint', 4.3377934), ('volant', 4.343449), ('planche', 4.343487), ('famille', 4.343506), ('me', 4.350053), ('chanceux', 4.350216), ('stéréotypes', 4.3503876), ('manger', 4.3566856), ('ombrage', 4.356925), ('parlé', 4.3586035), ('toit', 4.3604474), ('souche', 4.361786), ('genoux', 4.3629208), ('chemise', 4.367712), ('tasse', 4.372776), ('suis', 4.3744054), ('chapeau', 4.3747716), ('course', 4.3774614), ('voguer', 4.3796153), ('attraper', 4.379936), ('antenne', 4.3804297), ('repas', 4.38134), ('chaussures', 4.381962), ('panier', 4.3821936), ('avais', 4.3823647), ('chien', 4.3839793), ('mer', 4.3872347), ('bosquet', 4.387722), ('bouteille', 4.3918343), ('naviguent', 4.3925366), ('musique', 4.392584), ('boire', 4.398114), ('bosse', 4.4006557), ('hein', 4.4057446), ('branches', 4.407665), ('boit', 4.4109855), ('pogné', 4.4121327), ('cour', 4.4229927), ('pantoufles', 4.4241285), ('eau', 4.424397), ('standard', 4.4247284), ('ainé', 4.4321156), ('perds', 4.4347787), ('paire', 4.4351416), ('dessins', 4.435568), ('boîte', 4.4376607), ('verser', 4.4380646), ('opération', 4.440118), ('auto', 4.4414635), ('assis', 4.446261), ('souper', 4.4470096), ('camion', 4.451823), ('garage', 4.4542933), ('boissons', 4.4544573), ('attrape', 4.45898), ('affaires', 4.4593825), ('chaussettes', 4.4644027), ('patriotes', 4.4691534), ('tour', 4.4724565), ('circulent', 4.473431), ('plaque', 4.474478), ('collier', 4.474517), ('roman', 4.4751444), ('circule', 4.477515), ('breuvage', 4.484625), ('stationné', 4.4890757), ('image', 4.4926434), ('dupont', 4.4957113), ('su', 4.496276), ('demoiselle', 4.4981766), ('oui', 4.501557), ('volume', 4.5020647), ('pot', 4.5025544), ('arbres', 4.50349), ('dieu', 4.5058994), ('pitou', 4.509274), ('maman', 4.513353), ('-tu', 4.5166), ('tu', 4.5166), ('camisole', 4.516671), ('agenouillée', 4.5192094), ('mademoiselle', 4.522387), ('douce', 4.523822), ('pis', 4.5252657), ('vacances', 4.5258417), ('ligne', 4.5263386), ('navire', 4.5331936), ('costume', 4.5334888), ('tapis', 4.536398), ('radeau', 4.5371633), ('courts', 4.5394115), ('soleil', 4.5419383), ('canada', 4.547133), ('bat', 4.5478263), ('élève', 4.5531397), ('route', 4.5581465), ('air', 4.558612), ('cuillère', 4.5624876), ('cheveux', 4.5634074), ('peau', 4.5648093), ('café', 4.564879), ('anxieux', 4.5675273), ('radio', 4.568195), ('jambes', 4.5683975), ('fenêtre', 4.571639), ('pantalons', 4.574142), ('portatif', 4.578861), ('voyer', 4.579309), ('montagnes', 4.584245), ('royale', 4.5843625), ('ok', 4.5847173), ('allée', 4.5866804), ('château', 4.590249), ('garnitures', 4.5904922), ('agenouillés', 4.5969434), ('port', 4.6014214), ('corde', 4.606487), ('foin', 4.6075544), ('texte', 4.6093564), ('pelle', 4.6106606), ('style', 4.6118193), ('pêcheur', 4.614963), ('couples', 4.6168694), ('avion', 4.6203136), ('herbe', 4.622717), ('culotte', 4.6255383), ('ouais', 4.6263046), ('date', 4.626745), ('poupée', 4.6273623), ('feuillage', 4.63695), ('pousses', 4.6370854), ('anges', 4.642132), ('fleurs', 4.6473584), ('old', 4.648707), ('pêche', 4.6503487), ('nu', 4.651185), ('arbre', 4.6569443), ('poteau', 4.657848), ('sentier', 4.665911), ('dog', 4.675396), ('rive', 4.678043), ('tenant', 4.6789336), ('fleuve', 4.678976), ('aménagement', 4.6794167), ('ah', 4.681154), ('plage', 4.686222), ('situé', 4.688499), ('champêtre', 4.689558), ('québec', 4.694557), ('oiseaux', 4.6988907), ('gendarmerie', 4.6996865), ('rivière', 4.7003403), ('voile', 4.7038345), ('liquide', 4.705082), ('langue', 4.7068725), ('cooler', 4.708808), ('bain', 4.709166), ('cou', 4.7112203), ('gilet', 4.7147), ('vent', 4.722688), ('indien', 4.723304), ('eaux', 4.7233777), ('érotique', 4.7250314), ('escarpé', 4.7273726), ('pantalon', 4.7285275), ('flotte', 4.7314353), ('boisé', 4.7320843), ('ouin', 4.7335157), ('bol', 4.733791), ('pelouse', 4.7345037), ('nuages', 4.7426424), ('étudiant', 4.7744503), ('bermuda', 4.7775803), ('short', 4.778206), ('lunettes', 4.782929), ('papa', 4.783176), ('pont', 4.7834196), ('numéro', 4.7901163), ('ben', 4.796252), ('buissons', 4.799389), ('dame', 4.801039), ('grève', 4.8107038), ('sexe', 4.813935), ('lunch', 4.8209095), ('moteur', 4.821486), ('bah', 4.8224006), ('voiles', 4.829788), ('pêcher', 4.830309), ('chandail', 4.832643), ('ding', 4.8331385), ('ensoleillée', 4.838535), ('mat', 4.8417754), ('pigeon', 4.8504014), ('scout', 4.857962), ('poisson', 4.861193), ('vin', 4.880481), ('te', 4.884701), ('fido', 4.8897877), ('fenêtres', 4.8915887), ('feuilles', 4.891959), ('bel', 4.895353), ('nuage', 4.8983493), ('coiffés', 4.9005475), ('thé', 4.90291), ('rame', 4.9213834), ('étage', 4.9222317), ('quai', 4.9267793), ('m'en', 4.929808), ('sexuels', 4.934953), ('anneau', 4.9376006), ('poissons', 4.9476), ('camping', 4.9490633), ('plante', 4.9525747), ('bazou', 4.955598), ('c'est', 4.962208), ('nappe', 4.962387), ('monument', 4.969072), ('garée', 4.9734716), ('sapin', 4.9869785), ('sceau', 4.9936566), ('automobile', 4.9947467), ('voilier', 5.0004396), ('sable', 5.011914), ('jappe', 5.0277658), ('châssis', 5.0309362), ('hey', 5.0372868), ('châteaux', 5.045045), ('blonde', 5.045527), ('t-shirt', 5.063201), ('située', 5.0688353), ('queue', 5.069701), ('sorti', 5.0717044), ('mât', 5.0768595), ('mimi', 5.0945807), ('accoté', 5.099252), ('touffes', 5.1002393), ('bière', 5.11818), ('coiffe', 5.1196985), ('thermos', 5.123279), ('oh', 5.1299324), ('lac', 5.132256), ('pattes', 5.1497235), ('dunes', 5.16221), ('bobine', 5.1793113), ('liqueur', 5.1857166), ('race', 5.2113013), ('drapeau', 5.2182503), ('cèdre', 5.2282434), ('bouches', 5.2531796), ('genre', 5.263944), ('dériveur', 5.2706013), ('chaudière', 5.311918), ('bulldog', 5.3276), ('truite', 5.328149), ('érable', 5.3529177), ('arbustes', 5.4350333), ('espèce', 5.515527), ('gazon', 5.5492473), ('manches', 5.5581746), ('conifère', 5.5642786), ('médaille', 5.938898), ('aviron', 5.944509)

A.0.3.2. *English model, DementiaBank corpus full vocabulary*

('and', 1.8482457), ('the', 2.0135386), ('which', 2.0894387), ('both', 2.1301613), ('but', 2.1601744), ('that', 2.202486), ('although', 2.2350402), ('only', 2.2354772), ('while', 2.2397525), ('though', 2.2447183), ('also', 2.2465353), ('in', 2.2763052), ('even', 2.301794), ('of', 2.3030035), ('one', 2.3092291), ('well', 2.314999), ('a', 2.3324332), ('when', 2.3333845), ('another', 2.3361642), ('addition', 2.340561), ('then', 2.365733), ('finally', 2.3775647), ('actually', 2.3840547), ('apparently', 2.3903255), ('since', 2.3981702), ('presumably', 2.4135563), ('to', 2.4356833), ('as', 2.4421723), ('so', 2.4618144), ('just', 2.4644747), ('instead', 2.4839294), ('unfortunately', 2.4898381), ('they', 2.4935496), ('for', 2.4949763), ('other', 2.4974258), ('some', 2.5044866), ('with', 2.5100958), ('these', 2.5128863), ('it', 2.521663), ('still', 2.5245821), ('being', 2.5344684), ('time', 2.5375586), ("'s", 2.5424023), ('s', 2.5424023), ('because', 2.5469453), ('whereas', 2.5475883), ('essentially', 2.5479548), ('either', 2.550689), ('first', 2.552422), ('again', 2.5550885), ('others', 2.5598664), ('fact', 2.5698013), ('same', 2.5758595), ('particular', 2.577012), ('supposedly', 2.5814316), ('where', 2.5843685), ('evidently', 2.5907838), ('meant', 2.5971506), ('supposed', 2.6051764), ('many', 2.6052938), ('have', 2.6090314), ('always', 2.6105573), ('is', 2.616034), ('now', 2.6232884), ('beginning', 2.6243443), ('called', 2.627306), ('them', 2.6319087), ('already', 2.641857), ('all', 2.641962), ('like', 2.64249), ('yet', 2.6488771), ('basically', 2.6528354), ('earlier', 2.6584287), ('would', 2.6621702), ('after', 2.664131), ('was', 2.6684828), ('fortunately', 2.6842694), ('their', 2.6885107), ('eventually', 2.6886823), ('two', 2.6898663), ('those', 2.6933854), ('before', 2.7108057), ('this', 2.7109737), ('there', 2.717499), ('way', 2.7214496), ('aside', 2.7292664), ('including', 2.7301486), ('neither', 2.7365053), ('theirs', 2.7375355), ('almost', 2.7411034), ('he', 2.742714), ('along', 2.749238), ('probably', 2.753827), ('practically', 2.7551122), ('apart', 2.757689), ('elsewhere', 2.7604437), ('starting', 2.7622917), ('by', 2.762399), ('had', 2.765746), ('rather', 2.7662787), ('having', 2.7701395), ('said', 2.770873), ('may', 2.771123), ('not', 2.7794802), ('from', 2.781884), ('on', 2.7851267), ('thought', 2.7918036), ('his', 2.7920196), ('says', 2.7925823), ('quite', 2.794573), ('sometimes', 2.8011527), ('obviously', 2.801593), ('part', 2.8032043), ('often', 2.8036423), ('three', 2.8064866), ('otherwise', 2.8069746), ('certainly', 2.810303), ('or', 2.8114984), ('usual', 2.814163), ('ones', 2.8149164), ('coming', 2.8156674), ('meantime', 2.817989), ('could', 2.8221893), ('similar', 2.8284771), ('meanwhile', 2.8317256), ('seen', 2.832531), ('who', 2.8344555), ('definitely', 2.8359625), ('seeing', 2.838275), ('hardly', 2.8397424), ('going', 2.8397458), ('turned', 2.841356), ('least', 2.8422618), ('very', 2.8426101), ('few', 2.842957), ('what', 2.843203), ('kind', 2.8512368), ('whatever', 2.8521247), ('turn', 2.8532753), ('until', 2.8542964), ('whole', 2.8548245), ('except', 2.8563263), ('has', 2.8568563), ('come', 2.868381), ('soon', 2.8697722), ('appears', 2.8698094), ('around', 2.8713722), ('its', 2.8783698), ('work', 2.8805206), ('at', 2.883943), ('next', 2.8849971), ('seemed', 2.8855526), ('say', 2.8857827), ('much', 2.8867686), ('might', 2.888838), ('here', 2.8914223), ('presume', 2.8924007), ('never', 2.89353), ('about', 2.8941295), ('anyway', 2.8960319), ('over', 2.8970146), ('maybe', 2.900047), ('more', 2.90101), ('considered', 2.9037766), ('came', 2.904772), ('be', 2.9060571), ('been', 2.9106505), ('allowed', 2.9141018), ('saw', 2.91569), ('will', 2.9181376), ('certain', 2.9182026), ('possibly', 2.9187555), ('leaving', 2.9241655), ('own', 2.9253545), ('rest', 2.9261198), ('something', 2.9273198), ('are', 2.930718), ('really', 2.931019), ('somehow', 2.9332073), ('saying', 2.9385986), ('continue', 2.9485104), ('extent', 2.9485762), ('second', 2.9491222), ('seems', 2.952364), ('wanting', 2.9619129), ('than', 2.9639413), ('looking', 2.968924), ('presuming', 2.9733386), ('expecting', 2.9745748), ('quickly', 2.9769373), ('imagine', 2.977268), ('if', 2.977724), ('why', 2.978985), ('giving', 2.9838753), ('handful', 2.9843943), ('noticed', 2.987642), ('done', 2.9932618), ('possible', 2.9937775), ('out', 2.9950035), ('unusual', 2.9951568), ('sometime', 2.9984293), ('opposed', 2.9999459), ('means', 3.0007899), ('showing', 3.0008287), ('suppose', 3.003174), ('looked', 3.0047698), ('through', 3.0075884), ('clear', 3.009117), ('mostly', 3.015243), ('bring', 3.0174603), ('leading', 3.0176518), ('too', 3.0177062), ('making', 3.0204713), ('him', 3.0212002), ('taking', 3.0219517), ('hand', 3.0223026), ('put', 3.0271895), ('doing', 3.0278506), ('far', 3.0292928), ('beside', 3.031026), ('assuming', 3.0321248), ('were', 3.0352569), ('implies', 3.0361044), ('happens', 3.0395715), ('things', 3.0442805), ('moment', 3.0445802), ('we', 3.0470014), ('working', 3.0481336), ('can', 3.052461), ('better', 3.0559068), ('four', 3.0560148), ('putting', 3.0564775), ('everything', 3.056835), ('knew', 3.0581179), ('doubt', 3.058362), ('believe', 3.0587482), ('ever', 3.058767), ('take', 3.0654573), ('quietly', 3.0673697), ('up', 3.0678024), ('hoping', 3.071871), ('five', 3.0744176), ('makes', 3.077865), ('form', 3.0782335), ('couple', 3.0784862), ('somewhat', 3.0800781), ('got', 3.0812955), ('talking', 3.0813782), ('taken', 3.0831923), ('june', 3.086569), ('see', 3.087945), ('expect', 3.0879693), ('allowing', 3.088679), ('did', 3.0897796), ('anymore', 3.0903592), ('back', 3.091671), ('allow', 3.0934446), ('any', 3.0935845), ('appear', 3.0936778), ('hers', 3.0938215), ('went', 3.0949378), ('make', 3.0957966), ('turning', 3.0973134), ('each', 3.0988986), ('anyhow', 3.0994275), ('get', 3.1076505), ('consider', 3.1107244), ('typical', 3.111024), ('comes', 3.1111193), ('wanted', 3.111309), ('different', 3.1139226), ('nothing', 3.1139846), ('anywhere', 3.1155674), ('partially', 3.1160746), ('permitting', 3.1168256), ('telling', 3.117367), ('realize', 3.117588), ('forgetting', 3.11801), ('completely', 3.1190965), ('outside', 3.1194904), ('kept', 3.1215777), ('started', 3.121911), ('past', 3.1221356), ('seem', 3.1222928), ('think', 3.123051), ('thing', 3.1244204), ('look', 3.1269052), ('inside', 3.127554), ('thinking', 3.1288319), ('stopped', 3.1294131), ('course', 3.1299284), ('large', 3.1318123), ('fairly', 3.1348052), ('sure', 3.1402175), ('assume', 3.1444564), ('whether', 3.149444), ('sort', 3.1497872), ('present', 3.1528053), ('end', 3.153077), ('insisted', 3.153186), ('shortly', 3.1551037), ('happen', 3.1559434), ('between', 3.1640787), ('remain', 3.1644938), ('help', 3.16625), ('plus', 3.1667762), ('read', 3.173202), ('goes', 3.1732635), ('give', 3.175792), ('seven', 3.1819286), ('anything', 3.185492), ('noticing', 3.1895347), ('wrote', 3.189772), ('beyond', 3.191136), ('keeping', 3.192483), ('know', 3.1928637), ('remember', 3.1956825), ('nor', 3.1964715), ('must', 3.1993444), ('need', 3.2002804), ('hope', 3.2031932), ('matter', 3.2039468), ('aware', 3.2044256), ('happening', 3.2062688), ('told', 3.2081707), ('behind', 3.211335), ('eight', 3.2137237), ('according', 3.2167482), ('motioning', 3.2170656), ('himself', 3.2178664), ('third', 3.220433), ('encouraging', 3.2206233), ('go', 3.2221048), ('an', 3.222114), ('gone', 3.2224755), ('together', 3.223819), ('somewhere', 3.2244945), ('letting', 3.226836), ('neglected', 3.2275574), ('how', 3.2280967), ('her', 3.2285907), ('found', 3.2290094), ('good', 3.2299752), ('looks', 3.2301567), ('describe', 3.2322917), ('stay', 3.2368727), ('unaware', 3.2393131), ('half', 3.2419853), ('shows', 3.2428243), ('longer', 3.2429812), ('break', 3.244663), ('knows', 3.247212), ('mean', 3.2479281), ('little', 3.2481458), ('naturally', 3.248285), ('interesting', 3.2486165), ('she', 3.2509732), ('find', 3.252481), ('point', 3.253557), ('standing', 3.2550423), ('trouble', 3.2564754), ('connection', 3.2595608), ('asking', 3.259971), ('day', 3.2600017), ('situation', 3.2625632), ('suspect', 3.2639365), ('ten', 3.264448), ('deciding', 3.2687442), ('whom', 3.269572), ('changing', 3.2698925), ('want', 3.2707329), ('number', 3.2765577), ('getting', 3.2775369), ('lot', 3.2776263), ('down', 3.278427), ('puts', 3.2788646), ('wait', 3.279497), ('nobody', 3.2802992), ('wondering', 3.2813618), ('distressing', 3.2849445), ('indicating', 3.2855442), ('unconcerned', 3.2861867), ('indicates', 3.2865746), ('picked', 3.286841), ('mention', 3.2869096), ('person', 3.2869906), ('moving', 3.289378), ('forget', 3.290359), ('year', 3.2905064), ('stuff', 3.2919004), ('guess', 3.2926307), ('takes', 3.2943904), ('nine', 3.2948484), ('sense', 3.2954495), ('halfway', 3.296959), ('new', 3.2978473), ('difficult', 3.2982984), ('small', 3.2983854), ('else', 3.3010178), ('perspective', 3.305848), ('stand', 3.3064904), ('stumbling', 3.3084362), ('took', 3.3089542), ('interested', 3.309838), ('does', 3.3104255), ('pushed', 3.3115947), ('proceeding', 3.3117223), ('correct', 3.3121927), ('pretty', 3.3135602), ('reference', 3.3136103), ('setting', 3.3136187), ('wrong', 3.3138995), ('mistake', 3.3140373), ('place', 3.3168151), ('nineteen', 3.3220985), ('ordinary', 3.322181), ('set', 3.3224013), ('preparing', 3.3228269), ('long', 3.3237195), ('hurry', 3.3249745), ('important', 3.3253186), ('off', 3.3327107), ('let', 3.334174), ('words', 3.3350532), ('across', 3.3380241), ('i', 3.3419719), ('dozen', 3.3424237), ('into', 3.3443625), ('lots', 3.3443913), ('you', 3.3490381), ('mind', 3.352231), ('hands', 3.354233), ('forth', 3.354931), ('difference', 3.3573682), ('strictly', 3.35921), ('old', 3.3611987), ('counting', 3.3615031), ('danger', 3.3630884), ('awful', 3.3657339), ('labelled', 3.3671176), ('anybody', 3.3683462), ('enough', 3.3720748), ('neglecting', 3.372674), ('use', 3.3731675), ('careful', 3.37497), ('moved', 3.3769732), ('problem', 3.3770502), ('understand', 3.3785765), ('cautioning', 3.379433), ('strong', 3.3821957), ('days', 3.3829181), ('wonder', 3.3841856), ('strange', 3.3850214), ('instructing', 3.3869667), ('broken', 3.3890154), ('ready', 3.390392), ('worrying', 3.3931744), ('start', 3.3936353), ('tried', 3.397013), ('fricking', 3.3978617), ('years', 3.3983374), ('cause', 3.3995693), ('underneath', 3.40119), ('description', 3.4027143), ('gather', 3.4032137), ('assisting', 3.403972), ('piece', 3.4051523), ('cut', 3.405315), ('our', 3.4058704), ('job', 3.406979), ('responsible', 3.4087744), ('face', 3.408851), ('slightly', 3.410266), ('hurt', 3.4107606), ('affected', 3.4127429), ('hold', 3.41376), ('main', 3.415576), ('book', 3.4158888), ('whoever', 3.4162078), ('condition', 3.4169128), ('attempting', 3.4169807), ('keeps', 3.4175436), ('observed', 3.4184117), ('close', 3.4193108), ('directions', 3.4194577), ('coaxing', 3.423112), ('quick', 3.423818), ('managed', 3.4277067), ('snitching', 3.4278545), ('trying', 3.429138), ('pointing', 3.4293287), ('times', 3.429978), ('heard', 3.4301512), ('call', 3.4314203), ('unless', 3.4314594), ('terribly', 3.433), ('distinct', 3.4331748), ('excuse', 3.4333475), ('glance', 3.4336655), ('fully', 3.4338868), ('ignoring', 3.4341412), ('falled', 3.434952), ('gave', 3.435659), ('stepping', 3.4365017), ('life', 3.4377115), ('show', 3.4381857), ('sole', 3.4405801), ('touch', 3.4410954), ('pointed', 3.4444203), ('falling', 3.4448953), ('writing', 3.4453783), ('slight', 3.4454694), ('removed', 3.446036), ('frowning', 3.4487526), ('step', 3.4507506), ('drew', 3.451461), ('somebody', 3.4527645), ('away', 3.4532957), ('normal', 3.453334), ('home', 3.453675), ('sight', 3.4537697), ('vased', 3.4558582), ('gotten', 3.4585874), ('extended', 3.459319), ('tell', 3.4604108), ('dangerous', 3.4611678), ('perhaps', 3.4641147), ('enjoying', 3.4663854), ('attentively', 3.466868), ('holding', 3.4679308), ('lousy', 3.468639), ('pulled', 3.4697118), ('indication', 3.470638), ('near', 3.4710107), ('wants', 3.471858), ('handing', 3.4720483), ('dropping', 3.4745517), ('friend', 3.475296), ('touching', 3.475506), ('saved', 3.4757454), ('lost', 3.4759064), ('quiet', 3.4763074), ('carrying', 3.477336), ('fall', 3.4781702), ('short', 3.4782143), ('share', 3.4792876), ('pull', 3.479584), ("'re", 3.4810743), ('sudden', 3.4815607), ('slopping', 3.482793), ('titled', 3.4847271), ('man', 3.4857032), ('heck', 3.4877598), ('heading', 3.4880092), ('poor', 3.4906156), ('father', 3.4907184), ('word', 3.4926047), ('side', 3.4930243), ('young', 3.494005), ('waiting', 3.4957385), ('details', 3.4962626), ('correctly', 3.4972093), ('lose', 3.497463), ('passing', 3.4974954), ('dropped', 3.4984956), ('admonishing', 3.499439), ('busy', 3.4999444), ('sees', 3.501134), ('starts', 3.5031855), ('catch', 3.5061133), ('watching', 3.5066557), ('build', 3.5070002), ('helping', 3.5090287), ('breaking', 3.5145636), ('ask', 3.5150323), ('mentioned', 3.5166821), ('needs', 3.5189984), ('discovered', 3.5198748), ('someplace', 3.5199502), ('pair', 3.5200503), ('interest', 3.5208251), ('straight', 3.5208726), ('crazy', 3.5245261), ('handle', 3.5262942), ('ample', 3.5264702), ('towards', 3.528566), ('full', 3.5324066), ('harmed', 3.5337594), ('distinguish', 3.5350113), ('area', 3.5361047), ('bother', 3.5376153), ('filled', 3.5377805), ('real', 3.538296), ('fancy', 3.5391312), ('under', 3.5400913), ('precarious', 3.5426514), ('causing', 3.543804), ('broke', 3.544399), ('toying', 3.5453904), ('upsetting', 3.5469868), ('scene', 3.5476122), ('struggle', 3.5484955), ('every', 3.5508933), ('hundred', 3.5512018), ('moves', 3.5531995), ('project', 3.558506), ('reach', 3.5595202), ('ahead', 3.5595365), ('actions', 3.5607362), ('terrible', 3.5612018), ('variety', 3.561873), ('picking', 3.5623386), ('stop', 3.5626214), ('building', 3.5663743), ('haste', 3.567458), ('balance', 3.5684276), ('sharing', 3.5707097), ('ours', 3.5722246), ('run', 3.5724776), ('tired', 3.5733347), ('running', 3.576208), ('silly', 3.5782154), ('amount', 3.5787995), ('shocked', 3.5792437), ('distracted', 3.5821729), ('portion', 3.5831132), ('stepped', 3.583837), ('removing', 3.5839818), ('no', 3.5848117), ('write', 3.5852191), ('happy', 3.5865958), ('paying', 3.5888171), ('fallen', 3.5911934), ('okay', 3.5927513), ('shape', 3.592962), ('receiving', 3.5930579), ('caused', 3.5935524), ('fun', 3.5938256), ('catching', 3.5948815), ('nice', 3.5970004), ('careless', 3.5975957), ('furiously', 3.598953), ('damn', 3.5990553), ('lay', 3.5999997), ('silence', 3.600172), ('hiding', 3.6003053), ('forgot', 3.602185), ('tipping', 3.6025023), ('indicate', 3.6025257), ('enjoy', 3.602753), ('receive', 3.6030266), ('drop', 3.6049511), ('activities', 3.6063116), ('hidden', 3.6064363), ('hickey', 3.607475), ('process', 3.6077783), ('us', 3.6082654), ('single', 3.6094558), ('begging', 3.609966), ('left', 3.6102355), ('turns', 3.6125896), ('happenings', 3.6134496), ('crocker', 3.6142857), ('drops', 3.6146226), ('ends', 3.6155267), ('conflict', 3.6165557), ('cutting', 3.6192415), ('trip', 3.6210983), ('rolled', 3.6212077), ('hear', 3.6232011), ('alike', 3.6274495), ('bumping', 3.6275818), ('youngsters', 3.6304913), ('quit', 3.6360822), ('lovely', 3.6364129), ('mother', 3.6377068), ('right', 3.637855), ('nicely', 3.6387198), ('activity', 3.6406047), ('middle', 3.6407566), ('moments', 3.640875), ('checked', 3.6455383), ('counter', 3.6475356), ('myself', 3.6533577), ('my', 3.6538622), ('raising', 3.6572835), ('hard', 3.6581247), ('daydreaming', 3.6592274), ('stick', 3.6603048), ('yeah', 3.6609285), ('onto', 3.6609693), ('great', 3.6614344), ('big', 3.6616807), ('front', 3.6629505), ('sneaking', 3.662958), ('ruining', 3.6642494), ('growing', 3.6643376), ('stretching', 3.6678934), ('teetering', 3.668323), ('me', 3.6685977), ('slipping', 3.6687267), ('sounds', 3.6691623), ('losing', 3.6700695), ('snickering', 3.6702526), ('adjacent', 3.670547), ('active', 3.6706934), ('identifying', 3.6727097), ('slipped', 3.6747513), ('beautiful', 3.676919), ('awfully', 3.6769764), ('solidly', 3.6783197), ('steady', 3.678353), ('absentmindedly', 3.6788614), ('extension', 3.6805573), ('drawn', 3.6812162), ('fell', 3.6861482), ('ran', 3.689344), ('badly', 3.6910448), ('play', 3.6916237), ('picture', 3.6946037), ('obliterated', 3.6959634), ('view', 3.696277), ('steps', 3.697821), ('watch', 3.698104), ('speak', 3.699185), ('paper', 3.7005763), ('mirror', 3.7022457), ('minds', 3.7032125), ('pictured', 3.704033), ('ago', 3.7053983), ('raised', 3.707353), ('splattered', 3.7087975), ('conscious', 3.7108157), ('herself', 3.71107), ('grab', 3.7111862), ('facing', 3.7127922), ('bunch', 3.7129118), ('mess', 3.7141676), ('filling', 3.714391), ('crack', 3.7148383), ('detail', 3.7153037), ('against', 3.717095), ('position', 3.7195404), ('missed', 3.7199912), ('autumn', 3.7217968), ('room', 3.7219043), ('angry', 3.7227678), ('path', 3.723285), ('woman', 3.7236903), ('dad', 3.7257032), ('problems', 3.726077), ('head', 3.7275968), ('blowing', 3.7276342), ('bit', 3.7288136), ('kiddos', 3.7293415), ('hungry', 3.7327032), ('opened', 3.733373), ('glimpse', 3.7345595), ('event', 3.7348254), ('walking', 3.7352436), ('send', 3.7358778), ('threesome', 3.7363448), ('bemused', 3.7363603), ('eve', 3.7365668), ('youngster', 3.736885), ('sets', 3.7390745), ('direction', 3.7410388), ('shushing', 3.7422664), ('sloppy', 3.7433996), ('payed', 3.743646), ('dumb', 3.7439466), ('world', 3.7442641), ('shaking', 3.7443798), ('fresh', 3.7456937), ('dream', 3.74638), ('rocking', 3.746559), ('upset', 3.7472754), ('raining', 3.7481508), ('child', 3.7510157), ('nondescript', 3.7520027), ('chaos', 3.7521598), ('toward', 3.7524967), ('backs', 3.7525694), ('disturbed', 3.75339), ('reaching', 3.7538657), ('unbalances', 3.7539809), ('mistakes', 3.7543862), ('spring', 3.754521), ('visible', 3.7550752), ('keep', 3.755285), ('backwards', 3.7585301), ('someone', 3.7606518), ('sorry', 3.7619), ('attention', 3.7620149), ('blown', 3.7622643), ('crossed', 3.762301), ('criticizing', 3.7623367), ('cover', 3.762965), ('healthy', 3.7635899), ('cease', 3.7638848), ('lucky', 3.770585), ('walk', 3.7714264), ('stretches', 3.7730396), ('listening', 3.773253), ('seventy', 3.7737968), ('perplexed', 3.777177), ('opposite', 3.7793849), ('cook', 3.7799907), ('sitting', 3.7800846), ("'ll", 3.7810254), ('try', 3.7810915), ('handles', 3.7811282), ('pieces', 3.7811866), ('yep', 3.782469), ('gentle', 3.7843628), ('written', 3.7856438), ('staring', 3.785839), ('alright', 3.7877321), ('smile', 3.7881823), ('whispering', 3.788527), ('playing', 3.792246), ('gesturing', 3.7934144), ('kids', 3.7947114), ('ruffling', 3.7952936), ('overflowing', 3.795382), ('finished', 3.7966275), ('edge', 3.7967799), ('engrossed', 3.7996445), ('instructions', 3.7998216), ('hell', 3.8007588), ('backyard', 3.8013673), ('tops', 3.8065896), ('crying', 3.8072107), ('laugh', 3.8120863), ('high', 3.8124347), ('messed', 3.8127408), ('do', 3.8152227), ('bandaids', 3.8153682), ('dreaming', 3.8158998), ('smiling', 3.817058), ('deeply', 3.8209856), ('haywire', 3.8214774), ('darn', 3.8225179), ('pattern', 3.8225796), ('platter', 3.8227777), ('clean', 3.826972), ('hanging', 3.8271844), ('amounts', 3.8305964), ('depart', 3.8325758), ('top', 3.8332808), ('culprit', 3.8349912), ('made', 3.8368201), ('robbing', 3.8368948), ('above', 3.8376796), ('worst', 3.8382387), ('reaches', 3.8384495), ('guy', 3.8390765), ('fellow', 3.8391547), ('figure', 3.8399699), ('resting', 3.8415294), ('whisper', 3.8461933), ('corner', 3.846931), ('double', 3.8474214), ('tumble', 3.8484728), ('country', 3.8486214), ('children', 3.848887), ('hmm', 3.8491287), ('expression', 3.8503156), ('morning', 3.8534021), ('goodness', 3.8545296), ('grabbing', 3.85625), ('foot', 3.8569267), ('spilling', 3.858446), ('tied', 3.8598857), ('overfilled', 3.8612974), ('weak', 3.861384), ('reacting', 3.864476), ('survived', 3.864478), ('identify', 3.8653808), ('selection', 3.8663359), ('fast', 3.8674755), ('stealing', 3.869917), ('giggling', 3.8703725), ('oblivious', 3.8715396), ('wife', 3.8765376), ('beat', 3.876779), ('persons', 3.8767872), ('hey', 3.878676), ('tumbles', 3.8793068), ('eye', 3.8793142), ('incomplete', 3.8795452), ('should', 3.88018), ('baby', 3.8846543), ('people', 3.8853288), ('secret', 3.8860998), ('neighbor', 3.8881786), ('mini', 3.8911488), ('calamity', 3.8920515), ('care', 3.8920617), ('talk', 3.8932483), ('stacked', 3.8933887), ('flexible', 3.8939536), ('reading', 3.8946939), ('younger', 3.8948689), ('define', 3.8967435), ('descriptive', 3.8981054), ('your', 3.8981898), ('sticks', 3.89853), ('designate', 3.9025197), ('girl', 3.9034603), ('crawling', 3.9039361), ('yelling', 3.9058077), ('bed', 3.9113991), ("'ve", 3.9126658), ('huh', 3.9127307), ('ground', 3.9130168), ('brother', 3.913143), ('shades', 3.9133885), ('shove', 3.9134378), ('gushing', 3.9139504), ('pleasures', 3.9146097), ('calm', 3.9158404), ('hit', 3.9163609), ('personal', 3.9196389), ('bad', 3.9222567), ("'d", 3.9226377), ('ate', 3.923812), ('dripping', 3.925108), ('spilled', 3.9251213), ('light', 3.9260614), ('wiping', 3.9265575), ('hitting', 3.9282238), ('empty', 3.9294224), ('bottom', 3.9300215), ('pouring', 3.931044), ('nuthouse', 3.9330997), ('mad', 3.9334946), ('trap', 3.9341497), ('uhhuh', 3.9341586), ('loud', 3.9348593), ('sprinkling', 3.9358146), ('overflown', 3.9368637), ('mom', 3.9377482), ('knock', 3.9425404), ('venture', 3.9437308), ('section', 3.9463162), ('shut', 3.9471853), ("'m", 3.9472568), ('strapped', 3.94781), ('th', 3.9489276), ('dirty', 3.9503534), ('house', 3.9514961), ('boy', 3.9524884), ('blind', 3.9526856), ('overturns', 3.9529064), ('depends', 3.9529862), ('kid', 3.9575844), ('cockeyed', 3.9586785), ('mercy', 3.9625018), ('fixed', 3.9625626), ('distract', 3.9630656), ('gives', 3.963847), ('wall', 3.9642665), ('board', 3.9654372), ('handed', 3.9659283), ('school', 3.9674857), ('wobbling', 3.9699595), ('billowing', 3.970838), ('sakes', 3.9772081), ('pleasant', 3.979242), ('type', 3.9800086), ('open', 3.9820004), ('laying', 3.982248), ('state', 3.9831805), ('line', 3.9832232), ('son', 3.9835176), ('ruffled', 3.986724), ('catastrophe', 3.9889288), ('splashing', 3.9899683), ('attentive', 3.9917777), ('spreading', 3.9922578), ('throwing', 3.9927688), ('fine', 3.9928133), ('cracks', 3.993195), ('kill', 3.993928), ('collapse', 3.994254), ('excited', 3.9994218), ('fry', 3.999965), ('singing', 4.000396), ('smashed', 4.004068), ('ails', 4.00423), ('crashing', 4.0043554), ('overturning', 4.0044), ('garage', 4.008382), ('cleaning', 4.01141), ('pissed', 4.011507), ('force', 4.0126524), ('laid', 4.0155573), ('affair', 4.0170135), ('accident', 4.0196934), ('sound', 4.0220737), ('sign', 4.0229616), ('clothes', 4.0241437), ('slip', 4.0248685), ('combed', 4.0328493), ('stopper', 4.033521), ('outfitted', 4.034915), ('homely', 4.0349436), ('laughing', 4.0356913), ('dinner', 4.035755), ('wipe', 4.0386033), ('closet', 4.039069), ('shall', 4.0403285), ('slanting', 4.041725), ('upraised', 4.0422873), ('partial', 4.043283), ('gotta', 4.043922), ('tap', 4.045433), ('wow', 4.045522), ('upside', 4.047338), ('flat', 4.0479174), ('crooked', 4.0480795), ('minutes', 4.049349), ('darling', 4.0505276), ('slanted', 4.0512424), ('observant', 4.051912), ('vaseline', 4.0519423), ('tumbling', 4.052038), ('tie', 4.052218), ('steal', 4.0567083), ('dressed', 4.059552), ('splash', 4.060431), ('dear', 4.061443), ('draw', 4.0619674), ('throw', 4.06214), ('kitchen', 4.0622697), ('bet', 4.0626364), ('waving', 4.06392), ('dump', 4.0641155), ('green', 4.0642643), ('wearing', 4.0676575), ('tear', 4.068014), ('daughter', 4.068981), ('table', 4.069172), ('overboard', 4.069367), ('ties', 4.069684), ('plain', 4.0699663), ('immersed', 4.072043), ('preoccupied', 4.0727043), ('spread', 4.074067), ('charged', 4.0744996), ('older', 4.0748377), ('gets', 4.074983), ('swung', 4.0764513), ('overrunning', 4.076715), ('fiddling', 4.077169), ('tub', 4.078515), ('wiped', 4.0798783), ('curtains', 4.081486), ('doors', 4.084332), ('animals', 4.086259), ('washer', 4.087188), ('wood', 4.087724), ('bending', 4.0883355), ('tasty', 4.088921), ('gee', 4.0891056), ('door', 4.093566), ('decorating', 4.0939164), ('tad', 4.095053), ('realizes', 4.0952215), ('lopsided', 4.0957193), ('faced', 4.0965724), ('movement', 4.0973773), ('finish', 4.09756), ('family', 4.098458), ('painting', 4.0986867), ('daft', 4.098778), ('kicked', 4.1018505), ('significance', 4.1027822), ('reason', 4.1031513), ('junk', 4.1035666), ('leaning', 4.1055794), ('summertime', 4.1060987), ('act', 4.106108), ('silent', 4.107707), ('game', 4.1079826), ('overrun', 4.108003), ('sentence', 4.108574), ('nephew', 4.109975), ('upend', 4.1104593), ('oldest', 4.1117005), ('special', 4.1143064), ('salient', 4.119666), ('sad', 4.119673), ('minute', 4.1206756), ("boy's", 4.1211467), ('boys', 4.1211467), ('fingers', 4.121683), ('tipped', 4.1230283), ('slower', 4.1250696), ('donating', 4.1259956), ('sister', 4.127357), ('wash', 4.1276445), ('land', 4.128438), ('food', 4.1293473), ('glass', 4.1310353), ('outdoors', 4.131945), ('groomed', 4.1320925), ('brisk', 4.1323977), ('upward', 4.133157), ('drive', 4.1367536), ('peaceful', 4.1372867), ('water', 4.1377196), ('low', 4.1378064), ('base', 4.140117), ('space', 4.1436214), ('grandmother', 4.1438413), ('curtain', 4.1472387), ('pick', 4.1490493), ('pool', 4.150704), ('notice', 4.152751), ('action', 4.1529937), ('shoe', 4.1535354), ('bent', 4.154024), ('god', 4.1549883), ('bloom', 4.1556683), ('gosh', 4.1558285), ('tidy', 4.1567044), ('wipes', 4.1599245), ('dumping', 4.1603985), ('killed', 4.1633134), ('shining', 4.16338), ('galore', 4.1642714), ('total', 4.164964), ('garden', 4.1650863), ('dug', 4.1690354), ('ass', 4.169248), ('christmas', 4.1694865), ('countertop', 4.1698737), ('leaves', 4.170768), ('neatness', 4.176115), ('shoulda', 4.1817455), ('splashed', 4.1854377), ('copy', 4.1881485), ('percent', 4.1882844), ('washing', 4.188335), ('artistic', 4.189317), ('momma', 4.190303), ('airplane', 4.1916857), ('shoulder', 4.194202), ('sideways', 4.1960936), ('impervious', 4.1962214), ('disaster', 4.1971865), ('snowing', 4.198052), ('sneakers', 4.2011647), ('washed', 4.2045474), ('drain', 4.2076797), ('husband', 4.2126017), ('housewife', 4.2129474), ('slippers', 4.214324), ('lower', 4.2144237), ('damaged', 4.2160664), ('tv', 4.216793), ('abstractly', 4.218448), ('shoes', 4.2195306), ('shrubbery', 4.2197185), ('sizes', 4.2200637), ('devil', 4.2215786), ('daze', 4.221669), ('windowsills', 4.22297), ('housekeeper', 4.223037), ('ball', 4.22314), ('scenery', 4.226067), ('silencing', 4.2273917), ('girls', 4.228734), ('clogged', 4.228974), ('caught', 4.2298017), ('fountain', 4.232495), ('summer', 4.233705), ('am', 4.2337947), ('eating', 4.23439), ('winter', 4.239522), ('spills', 4.239821), ('housekeeping', 4.2415295), ('soaking', 4.2430253), ('honey', 4.243401), ('discharged', 4.243636), ('sidewalk', 4.244932), ('pardon', 4.2454767), ('bushes', 4.245535), ('billy', 4.245605), ('finger', 4.2467365), ('apples', 4.2569575), ('knocked', 4.2570004), ('cake', 4.257551), ('soaked', 4.2580767), ('lane', 4.2596445), ('fluttering', 4.2614627), ('sinks', 4.262023), ('machine', 4.262184), ('window', 4.2627764), ('coffee', 4.2630386), ('slid', 4.2652245), ('sink', 4.265269), ('drapes', 4.26576), ('tries', 4.266837), ('karen', 4.2669015), ('pompoms', 4.2689824), ('drawer', 4.269124), ('nose', 4.2708707), ('yes', 4.2716775), ('rag', 4.2726803), ('dry', 4.2728333), ('breeze', 4.273363), ('skin', 4.273543), ('flood', 4.2778163), ('length', 4.279662), ('tree', 4.281614), ('floor', 4.283583), ('flower', 4.284588), ('inattentive', 4.2852736), ('crash', 4.285831), ('hahaha', 4.2858477), ('eat', 4.287234), ('warm', 4.289058), ('overlooking', 4.2923927), ('spigots', 4.2925777), ('mommy', 4.294608), ('chair', 4.2955513), ('signing', 4.2966456), ('shiny', 4.297245), ('tips', 4.2990503), ('cupboard', 4.3019443), ('dishwater', 4.3021564), ('horse', 4.3038216), ('feet', 4.304394), ('colors', 4.30505), ('slop', 4.309862), ('dance', 4.3113747), ('tilt', 4.3137093), ('reached', 4.314159), ('inner', 4.31485), ('cookie', 4.31521), ('leaking', 4.315488), ('hears', 4.316099), ('evergreens', 4.31676), ('whew', 4.317418), ('lips', 4.3182874), ('mouth', 4.3184915), ('plum', 4.326342), ('counts', 4.3268466), ('draining', 4.329247), ('toppling', 4.32937), ('untied', 4.33018), ('dress', 4.331224), ('adult', 4.331848), ('road', 4.3338666), ('nails', 4.3342752), ('leg', 4.3366156), ('ladder', 4.3389363), ('women', 4.3392134), ('pans', 4.3406734), ('noise', 4.344065), ('snow', 4.3451295), ('daylight', 4.3463364), ('knees', 4.3468194), ('pipe', 4.3484054), ('upper', 4.3492327), ('weather', 4.349243), ('cascading', 4.351083), ('crawled', 4.3528748), ('scan', 4.357235), ('tilts', 4.3595443), ('signal', 4.3639746), ('faucet', 4.364039), ('saucer', 4.3642917), ('clock', 4.3646245), ('sissy', 4.366258), ('pass', 4.3708525), ('tilting', 4.371116), ('further', 4.372876), ('parted', 4.373654), ('bitchy', 4.3739343), ('cupboards', 4.374091), ('driveway', 4.3775363), ('center', 4.3799276), ('motion', 4.380313), ('letters', 4.380342), ('living', 4.3819857), ('unconscious', 4.3833594), ('boots', 4.3854275), ('toe', 4.3856273), ('calmly', 4.3861165), ('skirt', 4.386467), ('thicker', 4.3941684), ('meal', 4.394689), ('flying', 4.395214), ('falls', 4.395608), ('towel', 4.3964334), ('outer', 4.397086), ("mom's", 4.4014874), ('dishwasher', 4.403807), ('rolling', 4.4050636), ('glasses', 4.405699), ('mama', 4.4063997), ('johnny', 4.411679), ('medium', 4.413685), ('draperies', 4.4155188), ('jogging', 4.4157333), ('size', 4.4173183), ('drawers', 4.4188833), ('outdoor', 4.418943), ('hedges', 4.4207196), ('flowers', 4.4230824), ('squirting', 4.4238195), ('hoyle', 4.4238973), ('mouths', 4.4247203), ('tripodal', 4.4300222), ('cooking', 4.4300556), ('grading', 4.430783), ('image', 4.432772), ('planted', 4.435021), ('string', 4.437921), ('dryer', 4.439002), ('runs', 4.4417744), ('washes', 4.443112), ('cross', 4.4458647), ('mixed', 4.4477515), ('jerk', 4.448199), ('grandson', 4.450211), ('junior', 4.4528217), ('hairdo', 4.4529696), ('hair', 4.45397), ('weight', 4.454211), ('bench', 4.4553466), ('gravity', 4.456055), ('dried', 4.4594445), ('uhuh', 4.459834), ('sandals', 4.460699), ('roof', 4.4624276), ('wetting', 4.462677), ('sweater', 4.462871), ('jumping', 4.462946), ('tilted', 4.463642), ('grassy', 4.463862), ('landscaping', 4.4640355), ('breakfast', 4.467213), ('plug', 4.467827), ('wet', 4.471926), ('twins', 4.4750814), ('plant', 4.475238), ('suds', 4.476908), ('spiral', 4.4784975), ('bush', 4.4788165), ('overflow', 4.478899), ('swiping', 4.4792943), ('missing', 4.4798403), ('neighborhood', 4.4814863), ('oh', 4.4835215), ('hammer', 4.4873133), ('wreck', 4.4883738), ('climbed', 4.4893355), ('lady', 4.491229), ('topple', 4.494546), ('blooming', 4.4949985), ('goody', 4.4969616), ('deaf', 4.499531), ('wind', 4.501381), ('trees', 4.505153), ('flowing', 4.5051827), ('neck', 4.510868), ('carpentry', 4.511838), ('tea', 4.513094), ('styles', 4.513746), ('tip', 4.5150146), ('thank', 4.5180893), ('beep', 4.5190573), ('button', 4.5200076), ('legs', 4.52101), ('grass', 4.5242405), ('heel', 4.525114), ('pants', 4.5253487), ('drugs', 4.526934), ('age', 4.530541), ('height', 4.5313334), ('mamas', 4.532871), ('round', 4.539821), ('wove', 4.5415754), ('rubber', 4.541673), ('sunny', 4.543202), ('v', 4.543478), ('cloth', 4.543812), ('lad', 4.543913), ('raiding', 4.546108), ('valance', 4.5461984), ('tippy', 4.553064), ('lawn', 4.5568027), ('rows', 4.5587206), ('air', 4.560506), ('faucets', 4.5622663), ('spank', 4.5628552), ('spout', 4.5665197), ('memorize', 4.57778), ('torrent', 4.5786815), ('rooms', 4.5799828), ('plants', 4.5806904), ('sleeve', 4.5809393), ('lunch', 4.5821323), ('rice', 4.5827794), ('throat', 4.5841494), ('pots', 4.5851445), ('saucers', 4.5856814), ('lids', 4.5869923), ('baking', 4.590902), ('hazardous', 4.5928483), ('spells', 4.5971413), ('shirt', 4.597541), ('panes', 4.597768), ('overflowed', 4.5994453), ('puddle', 4.6003337), ('climbing', 4.601047), ('arm', 4.605592), ('gloves', 4.6077824), ('sugar', 4.6092296), ('outstretched', 4.6097727), ('diamonds', 4.6125507), ('lengths', 4.6132493), ('signaling', 4.616641), ('female', 4.6195617), ('leak', 4.621178), ('cream', 4.622875), ('wrist', 4.630713), ('style', 4.634036), ('race', 4.634865), ('square', 4.6389265), ('eaten', 4.639399), ('floors', 4.6453037), ('walkway', 4.6454206), ('cookies', 4.6468563), ('lip', 4.6471453), ('towels', 4.647989), ('sleeves', 4.6529703), ('prong', 4.653622), ('hedge', 4.6587243), ('hits', 4.6619463), ('angle', 4.664148), ('clause', 4.666953), ('jar', 4.6778936), ('jam', 4.6791463), ('turkey', 4.681677), ('plates', 4.6870604), ('gadgets', 4.688497), ('overflows', 4.690435), ('pathway', 4.6920023), ('ajar', 4.694743), ('dokey', 4.6974955), ('miniskirt', 4.69839), ('drying', 4.701185), ('pantry', 4.710137), ('hm', 4.713637), ('ah', 4.715925), ('jars', 4.7195625), ('weeds', 4.7198825), ('wee', 4.72183), ('cabinet', 4.725003), ('strap', 4.725405), ('shelf', 4.72856), ('waisted', 4.73682), ('pineapple', 4.73698), ('trunk', 4.7398524), ('stool', 4.7422743), ('pudding', 4.7435017), ('uhoh', 4.748276), ('lid', 4.7492275), ('anklets', 4.7492943), ('hot', 4.7510524), ('trousers', 4.7514706), ('duck', 4.752462), ('toppled', 4.7531896), ('miss', 4.75374), ('bang', 4.755206), ('ca', 4.758693), ('plate', 4.758915), ('puffed', 4.7636824), ('mantle', 4.7649875), ('jelly', 4.766054), ('jersey', 4.7690687), ('elbow', 4.7700863), ('laterals', 4.7710886), ('ya', 4.7744045), ('climbs', 4.774405), ('sill', 4.7776337), ('count', 4.780213), ('overran', 4.783031), ('flooding', 4.7853346), ('ai', 4.786286), ('straps', 4.7891107), ('wing', 4.791035), ('accidents', 4.7922583), ('climb', 4.797994), ('page', 4.798277), ('seat', 4.7985125), ('commode', 4.813454), ('leaf', 4.813803), ('sledge', 4.8169494), ('pompadour', 4.819623), ('curved', 4.8229218), ('birds', 4.8237715), ('spill', 4.8257294), ('gal', 4.8366065), ('rear', 4.8376093), ('mop', 4.8377743), ('knee', 4.838706), ('puffy', 4.8451324), ('stools', 4.84865), ('curl', 4.852433), ('sun', 4.8530383), ('drooping', 4.8537507), ('basin', 4.8603144), ('apron', 4.860616), ('geese', 4.8670025), ('recessed', 4.877163), ('background', 4.889021), ('ho', 4.890053), ('okey', 4.8943624), ('snowy', 4.9197116), ('sleeveless', 4.929659), ('etch', 4.934397), ('windows', 4.9382057), ('knobs', 4.945689), ('closed', 4.945769), ('wahoo', 4.952069), ('hiss', 4.953059), ('tin', 4.953976), ('scallop', 4.9570036), ('casement', 4.957245), ('+', 4.9692903), ('cloudy', 4.9760203), ('pan', 4.9838605), ('protrude', 4.984482), ('till', 4.986344), ('dries', 4.990792), ('stockings', 4.996527), ('petroleum', 4.9966326), ('fouled', 5.0008426), ('mhm', 5.006081), ('pencils', 5.009256), ('aprons', 5.0114274), ("n't", 5.0246935), ('nt', 5.0246935), ('dish', 5.0256863), ('foliage', 5.0326486), ('cleanup', 5.0479712), ('slender', 5.0555167), ('eggs', 5.0590696), ('dishes', 5.0619373), ('ankle', 5.070829), ('bordered', 5.0978026), ('spigot', 5.11479), ('mouse', 5.1163554), ('wreath', 5.1217165), ('toes', 5.122279), ('oop', 5.1449604), ('ta', 5.1454964), ('cups', 5.155212), ('pour', 5.188912), ('knots', 5.190997), ('moisture', 5.197071), ('cabinets', 5.2146893), ('cocoa', 5.2242455), ('blouse', 5.241955), ('kick', 5.2538176), ('legged', 5.256646), ('wowie', 5.259757), ('tripod', 5.2626514), ('male', 5.267125), ('tennis', 5.2688675), ('wo', 5.274727), ('verb', 5.2780385), ('index', 5.3006215), ('grasses', 5.303882), ('color', 5.3102274), ('oy', 5.3431573), ('flue', 5.3516254), ('shrubs', 5.3541956), ('bowls', 5.3643394), ('average', 5.368314), ('below', 5.403756), ('sash', 5.429716), ('cup', 5.432533), ('gon', 5.436862), ('tup', 5.4521737), ('sleeved', 5.452342), ('bowl', 5.4694076), ('shorts', 5.480028), ('mm', 5.5263977), ('shush', 5.5741053), ('socks', 5.6289372), ('shh', 5.7330294), ('na', 5.739), ('dishpan', 5.741518), ('copyright', 5.7478456), ('yard', 5.810993), ('please', 5.8262305), ('ret', 5.9001803), ('yards', 6.0255), ('ssh', 6.227252), ('shrub', 6.2334075)