Université de Montréal

Learning Discrete Word Embeddings to Achieve Better Interpretability and Processing efficiency

par

Samuel Beland-Leblanc

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

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

Orientation Intelligence artificielle

December 24, 2020

Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

Learning Discrete Word Embeddings to Achieve Better Interpretability and Processing efficiency

présenté par

Samuel Beland-Leblanc

a été évalué par un jury composé des personnes suivantes :

Pierre-Louis Bellec

(président-rapporteur)

Alain Tapp

(directeur de recherche)

Jian-Yun Nie

(membre du jury)

Résumé

L'omniprésente utilisation des plongements de mot dans le traitement des langues naturelles est la preuve de leur utilité et de leur capacité d'adaptation a une multitude de tâches. Cependant, leur nature continue est une importante limite en terme de calculs, de stockage en mémoire et d'interprétation. Dans ce travail de recherche, nous proposons une méthode pour apprendre directement des plongements de mot discrets. Notre modèle est une adaptation d'une nouvelle méthode de recherche pour base de données avec des techniques dernier cri en traitement des langues naturelles comme les Transformers et les LSTM. En plus d'obtenir des plongements nécessitant une fraction des ressources informatiques nécéssaire à leur stockage et leur traitement, nos expérimentations suggèrent fortement que nos représentations apprennent des unités de bases pour le sens dans l'espace latent qui sont analogues à des morphèmes. Nous appelons ces unités des **sememes**, qui, de l'anglais *semantic morphemes*, veut dire morphèmes sémantiques. Nous montrons que notre modèle a un grand potentiel de généralisation et qu'il produit des représentations latentes montrant de fortes relations sémantiques et conceptuelles entre les mots apparentés.
Mots clés: plongements, discret, binaire, Transformer, LSTM, sémantique, morphème, Search Data Structure Learning, Multi-Bernouilli Regression Search, généralisation, interprétabilité

Abstract

The ubiquitous use of word embeddings in Natural Language Processing is proof of their usefulness and adaptivity to a multitude of tasks. However, their continuous nature is prohibitive in terms of computation, storage and interpretation. In this work, we propose a method of learning discrete word embeddings directly. The model is an adaptation of a novel database searching method using state of the art natural language processing techniques like Transformers and LSTM. On top of obtaining embeddings requiring a fraction of the resources to store and process, our experiments strongly suggest that our representations learn basic units of meaning in latent space akin to lexical morphemes. We call these units **sememes**, i.e., semantic morphemes. We demonstrate that our model has a great generalization potential and outputs representation showing strong semantic and conceptual relations between related words.
Keywords: embeddings, discrete, binary, Transformer, LSTM, semantic, morpheme, Search Data Structure Learning, Multi-Bernouilli Regression Search, generalization, interpretability

Contents

List of tables

List of figures

Liste des sigles et des abréviations

DNN	Deep Neural Network
NMT	Neural Machine Translation
GEC	Grammatical Error Correction
NLP	Natural Language Processing
NLM	Neural Language Model
BiLM	Bidirectional Language Model
LM	Language Model
MBRS	Multi-Bernouilli Regression Search
BiLSTM	Bidirectional Long Short-Term Memory
H.D.	Hamming Distance

POS	Part of Speech
XOR	Exclusive OR (bit operation)
k-NN	k Nearest Neighbors

Introduction

Natural language processing has evolved considerably in recent years. It went from using complex statistical methods to using various deep neural network (DNN) architectures. This methodology shift introduced an increasing need for training data to a point where the minimum amount required for acceptable results became hardly achievable in a lot of downstream tasks. One way to mitigate that need was to add a pre-training step allowing the use of abundant raw text in a given language to get adequate prior knowledge about how the words in the vocabulary relate to each other. The result of this pre-training step are word embeddings that can be directly used in tasks such as neural machine translation (NMT) and grammatical error correction (GEC). Further ways to lessen this need for data were also used depending on the task to solve. For example, NMT and GEC use synthetic corpora generation such as using Back-Translation [**11, 43, 20**] or direct corpora generation from versioned text [**23**]. Although these proved to help performance, they are not useful for other tasks and there is still a limit to the amount of useful synthetic samples that can be generated.

Word representations were first being learned implicitly as part of a DNN aimed at solving an NLP task (e.g. Neural Language Model [**2**]). Then, they started to be considered as an independent task [**27, 31**]. The resulting embeddings proving to be useful, other works moved from the single vector representation to propose deeper combinations of vectors which could then be fine-tuned to the required task afterwards [**32, 33, 34, 9**]. All these techniques commonly use continuous vectors for the embeddings. This allows for greater precision, but the usual large vocabularies that are used consenquently make it so that significantly more compute resources are required to store and process them. The continuous nature also prohibits the interpretability of an embedding in isolation, limiting their use to potential downstream tasks only. In this work, we propose a technique to learn binary (i.e. discrete) word embeddings, which we argue can encode what we call *semantic morphemes* representing language independent latent discrete units of meaning analogous to lexical morphology. We also argue that this can provide similar performance while achieving greater interpretability and computation efficiency. This achieved level of interpretability also suggests the resulting embeddings are not only useful as intermediate representations. Our contribution is the adaptation of the Multi-Bernouilli Regression Search [**10**] technique to learn binary word embeddings directly by training our model end-to-end. The learned embeddings are 40 bits wide, greatly reducing the required resources, while still encoding rich semantic information. Our experimentation also shows our architecture having great generalization potential.

In chapter 1, we first present more detailed insight on the motivation that led us to want discrete representations. We then present related work in chapter 2 to provide a better background. In chapter 3, we define our technique and describe our training methodology. Finally, in chapter 4, we present the results of our experimentations.

Chapter 1

Motivation

1.1. Towards an Intelligent Vocabulary

As much as NLP models evolved, a common problem remained present with each different iteration and still is to this day. This problem is tokenization, or how should the text be presented. For example, should the word *don't* be kept as is? Should it be split as *do* · *n't* or *don* · *'t* to explicitly present to a model that this is a contraction of *do* and *not*. The former would lead to a more complex problem as the vocabulary would grow quickly. On the other hand, the latter seems to be more intuitive and also allows for a smaller vocabulary because other contractions like *won't* would share a common token. Then, if splitting common word contractions can lead to a more compact and intelligent vocabulary, why not use an even finer granularity for the tokenization by exploiting the compositionality of the english language. The different morphemes and the different parts of compound words could be considered individually. Table 1.1 shows how prefixes and suffixes can be extracted as independent tokens. Such a granularity would allow models to learn more specific representations for each token. In other words, if we consider the tokens *cat* and *cats*, two different representations would need to be learned even though each token represents the same entity with a different quantity. By using the split *cat* · *s*, this unnecessary overlapping is avoided.

Word	Tokenization
cats	*cat* · *s*
canadian	*canad* · *ian*
countdown	*count* · *down*
considered	*consider* · *ed*
liberate	*liber* · *ate*
devalue	*de* · *value*

Table 1.1. Examples of tokenization based on morphemes

Word	Tokenization	Meaning
cats	*cat · s*	More than one cat
dogs	*dog · s*	More than one dog
canadian	*canada · ian*	Belonging to Canada
italian	*italy · ian*	Belonging to Italy
french	*france · ian*	Belonging to France

Table 1.2. Examples of ideal tokenization

Morphology brings a better separation of concerns, yet the english grammar has a lot of irregularities that make this unusable in practice. A great amount of cases are not trivial. Should the word *ate* be split into *eat · ed* or kept as is? Should bound morphemes (i.e. morpheme without meaning by themselves) like *canad* in table 1.1 really be considered as a token? Even under the assumption that every word we encounter is written correctly, there is no clear good way to properly split raw text for optimal processing. Multiple data-driven techniques [**21, 34, 35, 42**] were explored, but again, they are based on the underlying characters. So, we decided to move our focus from finding an optimal way to tokenize words and instead tried to exploit latent semantic features since we believed they were more consistent.

1.2. Semantic Morpheme

The many inconsistencies with regular languages previously described led us to hypothesize that a word could be translated to an intermediate, language independent, representation based on what we call **semantic morphemes** (sememes). Analogous to lexical morphemes, we intended to represent words by the composition of discrete latent semantic units. Table 1.2 illustrates how the meaning could be split in sememes (e.g. *Belonging to* and *More than one* could be sememes) and how it could be achieve with what we consider an ideal tokenization. This reiterates the complexities of using text while showing how the meaning does not depend on the actual spelling. This also shows why we think of semantics as a more consistent basis. To emphasize this idea, there is another way of visualizing our intuition. If we take for example the words : *cat*, *dog*, *bite* and *claw*. They have no grammatical relation (i.e. they do not share any morpheme), but these four words clearly share common sememes, i.e., they have a common latent relation based on their meaning.

1.3. Embeddings Interpretability

The use of real valued vector for word embeddings is almost ubiquitous in state of the art models for NLP. One problem is that they embed information in a way that is not usable as is. Each dimension of the vectors having an infinite number of possible values makes it that no features or patterns can really be observed directly. Like it was presented in [**28**] and [**31**], there exist interesting relations between these vectors such as the linear subtructures that arise with similar concepts (e.g. difference between *king* and *queen* is similar to the difference between *man* and *woman*), but the interpretation can only come from the comparison with other vectors and cannot be extracted by simply observing their values independently.

One of our inspirations for this work are the chinese hanzis (i.e. characters). They can be seen as a unit between morphemes and our sememes. For example, the word *communism* is translated as **共产主义** in mandarin. A nice property of hanzis is that this word can be understood without even knowing what it is. By examining each characters separately, we can extract the following meaning : shared(共) + production(产) + main(主) + righteousness(义). One could argue that the same could be achieved by looking at the etymology of the word, which is : common (belonging equally to) + ism (a distinctive doctrine)[1]. Yet, as we previously explained in section 1.1 and illustrated in table 1.2, we would have to extract *common* from the morpheme *commun* which is not necessarily trivial as there are a lot of exceptions in the english language. So, we believed that by learning discrete (i.e. binary) word embeddings, we would be able to observe a similar phenomenon by looking at the sememes learned, which could make them interpretable as is and could be used in other areas of research like linguistics.

[1]As defined in *WordReference Random House Unabridged Dictionary of American English © 2020.*

Chapter 2

Related Work

2.1. Continuous Word Embeddings

Pre-trained continuous word embeddings predominate the natural language processing literature. Various methods to learn them have been proposed, all stemming from Firth's statement : "You shall know a word by the company it keeps" [**13**]. Referred as *distributional semantics*, this concept is based on the co-occurences of words. One of the first mentions of this idea used for NLP was in [**2**]. In their work, they indirectly learn word embeddings by training a Neural Language Model (NLM). The word sequences probabilities are jointly learned with what they call the distributed word feature vectors (i.e. the embeddings) which are simply vectors of free parameters used as the input for the NLM instead of one-hot encoded vectors. This allowed for a great dimensionality reduction as the input didn't have to be as wide as the vocabulary. Also, the NLM could now encode word distribution information separately for each word.

Following this, further work was done to explicitly learn the embeddings before using them in a downstream task. With this new pre-training step, the models for these tasks could directly use the distributional information instead of learning it jointly, hence reducing the complexity of the problem. Two popular shallow learning methods stood out in the literature : Word2Vec [**27**] and GloVe [**31**]. Word2Vec proposed two local context methods : Skip-Gram, where the model has to predict the surrounding word window using the middle word ;and CBOW, where the model needs to predict the middle word using the surrounding window. Both of these are simple, but they have one major drawback which is their inherent locality preventing them from using the word co-occurence statistics efficiently. GloVe tries to circumvent this by using the co-occurences data of the whole corpus. In their work, they observed that comparing the ratio of the co-occurence probabilities of words in a common context helps to distinguish relevant and irrelevant relations. This idea can be interpreted in vector space as the difference between the vectors. So, the embeddings are learned by minimizing a least squares regression model based on the pre-computed co-occurences and weighted on the frequencies of the words to mitigate the noise from under-represented words.

Even though these shallow models allowed for downstream tasks to perform better, a single vector representation contrains the learning potential. Words can have multiple meanings and different semantic and syntactic contexts that all must be compressed into a single vector. Deeper representations hence came to be used, such as ELMo [**32**]. Similar to [**2**], they pre-trained an L-layer bidirectional LSTM network as a language model (BiLM). To allow for more accurate representations, embeddings are not static as opposed the the previous methods presented. They actually are a function of the sentence in which they are. Input sentences are processed by the pre-trained BiLM and the final word embedding is a learned weighted combination of the L hidden layers. The next big leap in deep word representations came with the Transformer [**41**] architecture. GPT [**33**] and GPT-2 [**34**] use a Transformer decoder to pre-train a fowards-only language model and then finetune on downstream tasks. Another popular pre-training method is BERT [**9**] which, on the other hand, does not train as a LM. It pre-trains bidirectional representations by using a Transformer encoder to predict masked words and the next sentences, which, similarly to ELMo, incorporates the whole sentence in the representation of a single word. Finetuning layers can then be added on top of the pre-trained model to use it with downstream tasks.

This progress managed to achieve better performance at the expense of greatly increased complexity. Consequently, the interpretation of the learnt features and of the resulting represenions is not trivial, which contrains their generalization potential and their degree of reusability.

2.2. Non-distributional Word Embeddings

The problem with the lack of interpretability in word embeddings has also been raised in [**12**]. Instead of learning latent representations, they craft sparse binary vectors from the different linguistic aspects of a word : part of speech, hypernyms, hyponyms, holonymes, semantic categories, lexical and predicate-argument semantics, emotion, sentiments, colors, synonyms and antonyms. These properties are taken from existing resources like WordNet [**29**][1]. Competitive results on many NLP tasks were obtained when compared with distributional embeddings like Word2Vec and GloVe. This method is highly dependent on availability of linguistic data and manual adaptation of this data, which greatly limits the degree to which it can generalize to unseen words. Although there are a lot of downsides, it still hints towards the fact that discrete and interpretable representations can be learned and that they can be a viable alternative.

[1] Other sources omitted to be more concise.

2.3. Binarization as Compression

In other works, the discretization of word embeddings was considered as a lossless compression problem. One of the motivations for this is that continuous representations account for a significant amount of parameters in a model, which can be a limiting factor if the required resources are not available (e.g. embedded systems). Similar to the VQ-VAE architecture [**30**] used for image processing, a fixed length codebook of continuous vectors can be learned and then indexed appropriately. This is what [**38**] did in their work. They actually use M codebooks of K continuous vectors and then learn how to combine a vector from each codebook together to get a final embedding. An auto-encoder architecture is used with the Gumbel-softmax trick [**25**] to have dfferentiable indexing. Even though downstream models will use the combined continuous embeddings, the discrete indexes can be used to understand what the model learns as related word will index similar vectors.

Another way to achieve compression is by applying a hard threshold on the latent representations to obtain binary representations. As opposed to the previous technique, these discrete features are not indexes. They are analogous to their continuous counterpart. In order to use gradient based learning with this method, [**40**] uses an auto-encoder where the decoder is simply the transposed weight matrix of the encoder. The binarized output of the encoder is used as the decoder's input. To counteract the loss of semantic information from the thresholding operation in latent space, they also added an orthogonality constraint on the shared encoder/decoder weights in order to discourage correlation between binary features. Also using an auto-encoder, [**37**] applies either deterministic or stochastic binarization on the encoder's output. The encoder and the decoder have separate parameters and the gradient flows through using Straight-throught estimation [**3**]. Instead of using orthogonality regularization, they add a semantic preservation loss. This means that the relative distance between a triple of words must be similar in the input space and in the latent space.

These approaches all achieve a reduction in memory usage and better interpretability without sacrificing a significant amount of performance. They also share the same caveat, which is that they depend on pre-trained continuous embeddings. This bounds their results on the quality of the source vectors used.

Chapter 3

Learning Binary Embeddings

As opposed to what was presented in the related work, we do not compress pre-trained embeddings, rather we learn binary word embeddings directly. To do so, we adapted the Multi-Bernouilli Regression Search with the proposed F-beta loss in [**10**]. In section 3.1, we will present the formalism of this adaptation and in section 3.2 the neural models used around this technique.

3.1. Multi-Bernouilli Regression Search with F-Beta optimization

The Multi-Bernouilli Regression Search (MBRS) is a family of models for the Search Data Structure Learning (SDSL) proposed in [**10**]. SDSL is meant as a generalization of database searching where the query entities, the database entities and their relations are not constrained (e.g. using an image patch as a *query* to search for related images in a *database* like CIFAR-10). What is meant by unconstrained relations is that the relevance of a query to an entity (e.g. k-NN or subset) is not explicitly defined but rather implicitly defined by the data.

Let q be a *query* and d be a *document*[1]. Let $f^{Q}_{\theta_1}(q) = \boldsymbol{\pi}^q$ and $f^{D}_{\theta_2}(d) = \boldsymbol{\pi}^d$ be parametric functionals for the *query* and *document*. Here, both functionals are implemented as independent neural networks (see 3.2). Both $\boldsymbol{\pi}^q$ and $\boldsymbol{\pi}^d$ are the parameters for a Multi-Bernouilli[2] distribution such that $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_n) \in [0,1]^n$. Also, assume there is a relation R function (i.e. matching function) such that $F_R : \Omega_Q \times \Omega_D \to [0,1]$ predicts the probability of a relation between the queries domain Ω_Q and the documents domain Ω_D. In the case of MBRS, we need a matching function to determine if a match between both domains has

[1]The terms *query* and *document* are used to be comparable to the database searching literature, but in this context, they could represent any two domains to be matched.

[2]A Multi-Bernoulli random variable is a vector of independent but not identically distributed Bernoulli; the parameters of a Multi-Bernoulli is consequently a vector of probabilities.

occured. We define it as,

$$M(f_{\theta_1}^Q(q), f_{\theta_2}^D(d)) = \prod_{i=1}^{n} \pi_i^q \pi_i^d + (1-\pi_i^q)(1-\pi_i^d). \tag{3.1.1}$$

With this formulation, each Bernouilli parameter represents a different bit. This means $f_{\theta_1}^Q(q)$ and $f_{\theta_2}^D(d)$ only match if the bits at each position are equal. This definition also allows the gradient to flow to both f_{θ_1} and f_{θ_2}, so it can be trained end-to-end. The explicit binarization is obtained by using a hard threshold on $\boldsymbol{\pi}$ after training. As opposed to the hard-thresholding used in the compressing algorithms presented in 2.3, the multi-bernouillis are implicitly binarized during training by pulling each parameter π_i towards 1 or 0.

To train this matching function, we use an approximation of the F-Beta score as our loss function. We first define the precision as

$$p = P(R|M) = \frac{P(M|R)P(R)}{P(M)} = \frac{r \cdot q_R}{m}, \tag{3.1.2}$$

where M represents a match, r is the recall, q_R is the relation marginal probability and m is the matching marginal probability. We then obtain the following F-Beta formulation:

$$F_\beta = \frac{(1+\beta^2)pr}{\beta^2 p + r} = \frac{(1+\beta^2)\frac{rq_R}{m}r}{\beta^2 \frac{rq_R}{m} + r} = r \cdot g(m), \text{ where } g(x) = \frac{(1+\beta^2)q_R}{\beta^2 q_R + x}, \tag{3.1.3}$$

where β is a hyper-parameter. As explained in the original work, unbiased estimates can be obtained for r and m. A naive estimator for p can be obtained, but it would be biased. This is why this formulation was favored. To improve the sampling and implementation efficiency, we reformulate m as

$$m = P(M) = P(R)P(M|R) + P(\neg R)P(M|\neg R) = q_R \cdot r + (1-q_R) \cdot s, \tag{3.1.4}$$

where s is the fallout. Assuming we can sample from $P(Q,D|R)$ and from $P(Q,D|\neg R)$, estimators can be obtained with empirical averages. Let X be samples from $P(Q,D|R)$ and Y samples from $P(Q,D|\neg R)$, then our estimators are :

$$\hat{r}_\theta = \frac{1}{|X|} \sum_{q_x, d_x \in X} M_\theta(f_{\theta_1}^Q(q_x), f_{\theta_2}^D(d_x)) \tag{3.1.5}$$

$$\hat{s}_\theta = \frac{1}{|Y|} \sum_{q_y, d_y \in Y} M_\theta(f_{\theta_1}^Q(q_y), f_{\theta_2}^D(d_y)) \tag{3.1.6}$$

$$\hat{m}_\theta = q_R \cdot \hat{r}_\theta + (1-q_R) \cdot \hat{s}_\theta. \tag{3.1.7}$$

The final cost function we are trying to maximize is (the log is taken to provide numerical stability)

$$J_\theta = \log F_{\beta_\theta} = \log((1+\beta^2)) + \log(\hat{r}_\theta) - \log(q_R\beta^2 + \hat{m}_\theta). \tag{3.1.8}$$

In their work, Duchesneau et al. propose a redifinition of $\log(\hat{r}_\theta)$ where the log is moved inside the sum (see 3.1.5). The gradient of the initial formulation will be near zero for samples that are matching. Since we are trying to match between two domains, it is important that the model receives a good training signal for that case. By using this alternate estimator, a bigger gradient flows through the model when a match should occur. The adjusted cost function then becomes :

$$J_\theta = \log F_{\beta_\theta} = \log((1+\beta^2)) + \rho_\theta - \log(q_R\beta^2 + \hat{m}_\theta), \tag{3.1.9}$$

where $\rho_\theta = \frac{1}{|X|}\sum_{q_x,d_x\in X} \log M_\theta(f^Q_{\theta_1}(q_x), f^D_{\theta_2}(d_x))$. In section 3.2, we explain how we adapted our problem to this formulation.

3.2. Neural Models Architectures

Similar to what we presented in 2.1, we want to extract our binary representations from the word co-occurence statistics. One way to approximate them while easily integrating the required sampling is to consider local context windows of odd length. For a certain word, we consider that both the sequence of words before and after are as important. The odd length makes it so that we can have an equal context surrounding a particular word. We can then define the domain for Q as the middle words and the domain for D as this surrounding context. This formulation makes it easy to see how the data implicitly defines the relation R between both domains. It is also trivial to sample from $P(Q,D|R)$ and $P(Q,D|\neg R)$. We first sample n middle words q_i from the unigram probabilites of the vocabulary. Then, for each q_i, we uniformly sample an associated window d_i. This constitutes our samples $(q_i, d_i) \sim P(Q, D|R), i \in \{1, 2, ..., n\}$. By reusing the middle words with the other contexts we can obtain our $P(Q,D|\neg R)$ samples at the same time. Formally, this means :

$$(q_i, d_j) \sim \begin{cases} P(Q, D|R), & \text{if } i = j \\ P(Q, D|\neg R), & \text{if } i \neq j. \end{cases} \tag{3.2.1}$$

The implementation of $f^Q_{\theta_1}$ and $f^D_{\theta_2}$ could be resource intensive, so doing it like this avoids unecessary processing since the computed $\boldsymbol{\pi}^Q_i$ and $\boldsymbol{\pi}^D_i$ can be reused. Figure 3.1 shows how we implemented both neural networks for f_{θ_1} and f_{θ_2}. Since Q only represents the middle words, the implementation is a $|V| \times k$ continuous embedding matrix, where $|V|$ is the vocabulary size and k is the discrete embedding size. Even though the data is continuous during training, we refer to k as the *discrete* size since at the end, a hard threshold is applied directly on that matrix. In this case, because we are using a sigmoid before the matching function, the elementwise threshold function is

$$\mathrm{bin}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.2.2}$$

We use a combination of bidirectional LSTM and Transformer to process the contexts in D. The rationale behind this architecture is that we wanted to add self-attention on top of the LSTM to provide better focus on important words to the model (e.g. weighting nouns more than determinants). We first tried implementing an architecture like in [**24**], which uses additive attention similar to [**1**], but the results were not impressive. Inspired by [**26**], we then tried replacing the attention layer by a single layer Transformer encoder since this new architecture was proving to boost performance in a variety of NLP tasks. We also thought the multi-head attention would allow the model to capture more complex relations (e.g. different meaning for different context). As opposed to the original paper, we do not use the positional embeddings at the input of the Transformer. As input, we are using the LSTM's output which should provide implicit positional information in its latent representation. The input embeddings for the LSTM are the same as the one used for f_{θ_1} (i.e. $\theta_1 \subset \theta_2$). Our intuition was that whether a word appears in the context before or after has a significant impact. This is why we process the before and after context separately (using the same LTSM and Transformer layer) and then combine them before trying to match them with MBRS. We believe that with this methodology, the self-attention will focus more on words that would impact significantly the meaning of the context.

Our cost function also requires the marginal probability for a relation (i.e. q_R). Since we defined our queries domain Ω_Q as the middle words in the windows, this allowed us to reformulate this term as :

$$q_R = P(R) = \sum_{q \in \Omega_Q} P(R|Q = q)P(Q = q) = \sum_{q \in \Omega_Q} P(Q = q)^2. \tag{3.2.3}$$

This uses the fact that, in our case, a word always has a context, so the following equality holds :

$$P(R|Q = q) = P(Q = q).$$

This term being constant across training, we computed it once using the unigram probabilities of the training vocabulary.

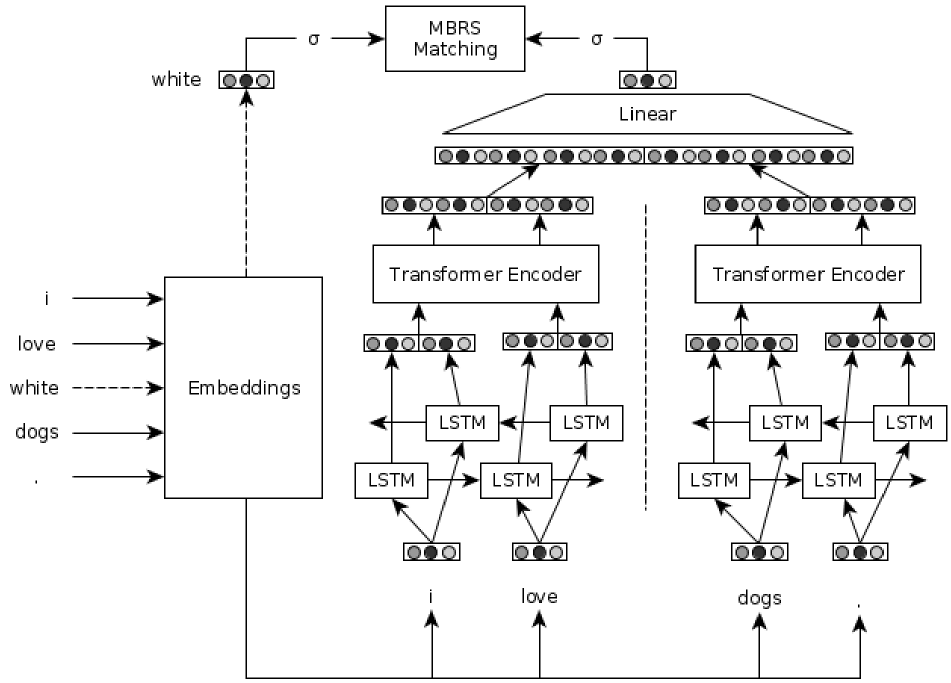

Fig. 3.1. Our implementation of MBRS. f_θ^Q is simply the embeddings. It receives the index for the middle word of a window. f_θ^D is a BiLSTM with a single Transformer layer on top which receives the before and after window context separately (i.e. the after context can't attend to the before context and vice versa). The final representations of the before and after context are then concatenated and projected to the desired discrete embedding size. Both the result of f_θ^Q and f_θ^D are passed through a sigmoid to obtain $\boldsymbol{\pi}^Q$ and $\boldsymbol{\pi}^D$ which are used to optimize the cost function defined in 3.1.9. The embeddings are shared between both sub-models.

3.3. Experimental Setup

3.3.1. Dataset

We built our corpus by combining BookCorpus [**44**] and the whole english Wikipedia[3]. The original BookCorpus contained compilations, which means that certain books were duplicated as their stand-alone version was also present. To avoid a bias in the word probabilities, we removed these stand-alone versions so that no sentences appeared twice unnaturally.

We padded the beginning and end of each document to avoid windows overlapping between them. Documents refer to either a full book or a Wikipedia article. We tokenized the whole corpus with SpaCy [**17**] to standardize the input text. We did not apply any encoding like Byte Pair Encoding (BPE) [**36**] or SentencePiece [**21**]. Even though these techniques are useful for other NLP tasks like Language Understanding [**9, 33, 34**], Machine Translation

[3]The 2020/02/10 dump.

[**11, 39, 22**] and Grammatical Error Correction [**20, 5, 16, 14, 7, 8**], they split words based on their characters, which we argued might not be optimal and is not the focus of this work. Also, their data-driven nature leads to word splits which are not necessarily proper morphemes (e.g. *banana* could split in *ba* · *nana*) and this defeats our goal of interpretability. We finally converted everything in lowercase.

In table 3.1, we summarize the final corpus properties. To limit the memory footprint[4] of the embeddings and to allow us to analyze the quality of each embedding more efficiently, we limit our vocabulary to the 50,000 most frequent words with at least a frequency of 2 (this boundary is not reached in our case). All other words are replaced by an *<unk>* token.

Corpus Information	
Num. of words	3,213,438,722
Num. of documents	6,012,332
Doc. Train/Valid split	20%
Original vocab. size	8,375,078
Restricted vocab. size	50,000

Table 3.1. Final properties after combining both raw corpora and preprocessing them.

3.3.2. Model Configuration and Training

We set the embedding dimensionality to 40 and used windows of 51 words. The LSTM's hidden size and the feed-forward layer in the Transformer encoder were both set to 100. We used a single Transformer layer with 8 attention heads, a ReLU activation and the default dropout rate of 0.1. For MBRS, we used a β of 2^{-15} with a group size (gs) of 200. We used the Adam optimizer [**19**] with a learning rate of 0.0001 and no weight decay. We used a batch size (bs) of 16. Following [**10**], we exponentially increased β to avoid bit saturation by linearly increasing $\log_2 \beta$ from -100 to our desired -15 in 1000 steps. We also use their regularization on the outputs of f_θ^Q and f_θ^D, which is

$$\gamma = \frac{\sum_i^k (z^i_{f_\theta^Q})^2}{k} + \frac{\sum_i^k (z^i_{f_\theta^D})^2}{k} \tag{3.3.1}$$

$$\lambda = \tau * \frac{J_\theta}{\gamma}, \tag{3.3.2}$$

where k is the embedding dimensionality and τ is a hyper-parameter (0.001 in our case). The complete cost function is then

$$J = J_\theta + \lambda * \gamma. \tag{3.3.3}$$

[4]We designed a compact way to store the corpus so that it could be loaded in memory completely for faster sampling during training. It takes about 16GB. If we loaded from persistent memory (e.g. hard-disk drive), we would have more memory headroom, thus allowing more embeddings to be trained.

The model was trained for 455,000 steps and the hard threshold (see 3.2.2) was applied to the embedding weights to obtain the final representations.

The group size refers to the $P(Q, D|R)$ and $P(Q, D|\neg R)$ samples used in equation 3.2.1. Since we are reusing samples for efficiency, the group size is equal to the number of $P(Q, D|R)$ samples. This means that, with our configuration, a single batch has 200^2 estimators for the MBRS matching. This is achieved with only gs $*$ bs windows that needs to be in memory (i.e. 3200). The model was trained on a single GTX1080 and 3200 windows was the maximum that would fit in memory. From our experimentation, a bigger group size compared to the batch size was favorable, so this is why gs $>>$ bs. This also follows from the fact that the recall and fallout are empirical estimates and that having more samples allows the model to have a better estimation of the real word distributions.

The β was chosen to be this low to weight precision way more than recall. If the value was too high, the model always ended up collapsing every embeddings to the same value. If the value was too low, than the precision constraint was too important for the model to learn any meaningful representations. The same goes for the restrained capacity of the model, which is really small compared to the language understanding state of the art BERT for example. If the model had too much capacity, it would also collapse the embeddings to a single representation. If it had too little, the learning potential was too limited. With the current configuration, a good precision/recall balance was achieved.

Chapter 4

Results

4.1. Embedding Neighborhoods

To establish the quality of our embeddings, we explored the neighborhoods of each words in the binary latent space. In 4.1.1, we look at the closest words in terms of hamming distance and compare them with continuous and binary GloVe embeddings. In 4.1.2, we use unsupervised clustering to try to reveal more complex hidden relationships.

4.1.1. Nearest Neighbors

The first qualitative evaluation we did was to manually look at the nearest neighbors for common words. This allowed us to use common knowledge as an informal baseline (e.g. it is assumed to be common knowledge that *computer* belongs to the family of electronics). The hamming distance was used as our distance measure. It is analogous to the cosine similarity which is often used with continuous vectors. It also has a performance advantage since it is adapted for binary representations and the computation is inherently simple. To have a baseline, we wanted embeddings with known good results, so we used GloVe. We trained our own vectors using their public implementation[1] and our training corpus. The vector dimension was set to 256 and distance weighting was disabled. Since we do not explicitly weight co-occurences by their distance, we assumed it would make the available context for each word more akin to ours. We used the same window size of 51. All other parameters were left at their default value and we trained the model for 100 iterations. As presented in the related work, [**40**] proposes a simple model to compress continuous embedding to binary representations. To have a more comparable baseline, we used this technique to compress our trained GloVe embeddings. We set the compressed dimensionality to 40 and trained until convergence using a λ_{reg} of 1.

[1] `https://github.com/stanfordnlp/GloVe`.

We present in table 4.1 some neighborhoods we analyzed. One characteristic that is quickly observable is that our model seems to group based on semantic similarities, which is comparable to both the compressed and continuous GloVe embeddings. The basis of our methodology being distributional semantics, this was the expected behaviour. One novel property that we observed is that our groupings also include grammatical similarities (e.g. same POS). This behaviour is apparent with the words *computer* and *computers*. Both representations were still grouped with other related words (e.g. *laptop*, *workstation*, *programming*). However, most of the closest words to *computer* are singular and most of the closest words to *computers* are plural, which is not true when looking at GloVe. Similar inflectional relations are exposed when looking at the verbs *eat*, *ate*, *eats* and *eating*. In section 4.2, we explore the interpretability of this behaviour by looking at the bits used by the model for different grammatically related words.

Table 4.1 lists the words in ascending order of hamming distance (cosine similarity for the continuous embeddings). Based only on this distance measure, the embeddings from our work appear to put more emphasis on grammatical properties when compared with both GloVe counterparts. To get a more accurate interpretation, we evaluated our results on two similarity datasets. Similar to [**40**], we used : MEN [**4**], for word similarity; and SimVerb [**15**], for verb similarity. We compiled the spearman rank-order correlation coefficients in table 4.2. According to this, continuous GloVe vectors are grouped by their meaning significantly more than the other methods. One caveat is that they require ≈ 200 times more memory (assuming 4 bytes floating point numbers) than our work. To see if this performance was due to their higher precision, we compressed them to 256 bits using the same method previously explained. The results were really close to the originals[2], meaning the training methodology has more importance. At 40 bits, our work is better than compressed GloVe. A quick qualitative evaluation of the neighbors of the verb *eat* and its inflections in table 4.1 seems to indicate that the opposite was expected. From common knowledge, *eaten*, *meal*, *lunch* and *eating* are more related to the verb *eat* than *fly*, *choose*, *explode* and *steal*. Although, this only takes into account the word itself. If we consider the context in which the word might appear, these grammatical properties speak to the semantics of that context. Whether I eat or I am eating has different meaning when considering the sentences in which each would be present. At no point is character-level information provided to our model, meaning this apparent correlation with the morphology of nearest neighbors is the manifestation of a common semantic marker. In consequence, our view of what a sememe might represent might have been too narrow. Another factor we considered is that the low dimensionality of our embeddings forces the model to learn features as a combination of multiple bits. By only looking at the hamming distance, the closest neighbors will have most of the same bits, hence most of the same features. Assuming the model makes a distinction between

[2]The results presented in [**40**] are similar.

	computer	computers	eat	ate	eats	eating
This Work	laptop	hardware	fly	bathed	stirs	watering
	mainframe	printers	choose	whistled	sucks	chewing
	competency	laptops	explode	chattered	dissolves	hoarding
	workstation	gadgets	steal	hid	treats	feeding
	graphical	processors	enjoy	woke	clears	drinking
	webcam	workstations	skim	simmered	withdraws	smoking
	pc	plugins	smelled	palmed	fills	tagging
	space	cpus	mew	craved	fakes	pruning
	computing	calculators	dat	paled	ignores	packing
	microphone	apps	birthdays	imagined	stings	roasting
	smartphone	notebooks	lili	consumed	lowers	cooking
	competency	drinks	crave	searched	begs	healthy
	macro	ipad	breathe	toasted	assembles	nesting
	encryption	modules	infect	clapped	distinguishes	washing
	programming	machines	consume	retorts	shrugs	hatching
Compressed GloVe	laboratory	database	eaten	drink	mashed	meal
	lab	modules	meal	dressed	tart	chasing
	labs	software	lunch	hang	jerky	cooking
	web	computing	eating	fancy	spiced	eat
	handled	labs	meat	drank	sweets	meat
	laboratories	simulation	bread	quietly	garlic	believe
	telephone	electronics	potatoes	smiled	gravy	eats
	console	automated	soup	meal	iced	soup
	software	input	biscuits	stuck	spaghetti	dishes
	computing	user	happened	eyed	rooster	eaten
	processing	micro	drink	drinks	biscuits	chewing
	devices	simpler	beef	mouthed	morgue	sour
	modeling	interface	sauce	glance	sunflower	ate
	controlling	environments	chewing	sight	rhiannon	food
	electronic	feedback	hungry	clothes	donut	lunch
GloVe	computers	computer	eating	eat	consumes	eat
	software	software	ate	meal	eat	ate
	computing	computing	eaten	eaten	sleeps	eaten
	technology	devices	meal	eating	eaten	meal
	electronic	hardware	hungry	drank	feeds	food
	electronics	systems	food	hungry	bites	meat
	systems	ibm	bite	cooked	eating	hungry
	lab	laptops	meat	delicious	consume	feeding
	digital	machines	eats	sandwiches	forgets	bite
	user	mainframe	bread	breakfast	ate	cooked
	hardware	desktop	cooked	bite	cleans	chicken
	internet	interface	breakfast	bread	tosses	drink
	data	laptop	chicken	chewed	hates	consumed
	information	pc	drink	lunch	drags	diet
	ibm	electronics	dinner	wandered	sucks	drinking

Table 4.1. Five words and their 15 closest neighbors in ascending order of hamming distance for this work and for binary embeddings of the same dimension obtained by compressing pre-trained GloVe continuous embeddings. The 15 closest words based on cosine similarity are also included using our trained GloVe embeddings.

grammatical and semantical properties, then encoding grammatical features might require more information, so the closest words are more likely to have these same features, explaining the apparent emphasis on morphological similarities over semantic ones. It is hard to say from these preleminary results only if the model makes this distinction or if grammatical properties can actually be mapped to semantical units. Either way, our conclusion from this is that this distance measure might not be optimal to properly understand and evaluate the relations that were learned. We tried several similarity and distance measures from [**6**], but the results didn't differ in any meaningful way (see table 4.4), which also supports our multi-bit encoding hypothesis. We also present more in depth experimentations based on this in section 4.2.

Even with a constrained embedding width, our model was able to learn more complex features than simply compressing from a good pre-trained model. Seeing how even compressed at 256 bits GloVe manages to still encode better semantic information, pushing the number of bits in our embeddings higher might greatly improve their quality. It would also still require significantly less memory compared to continuous vectors with the same number of dimensions. Their binary nature would make computations faster, but it would also allow the use of hardware acceleration by using vector instructions[3] [**40**]. This avenue and the use of a potentially better distance measure is left for future work.

Dataset		**This work**	**GloVe - Bin**	**GloVe - Con**	**GloVe - Bin**
Name	Coverage	(40 bits)	(40 bits)	(256 dim.)	(256 bits)
MEN	99.0%	53.9	48.2	**73.8**	71.6
SimVerb	94.0%	12.8	5.6	**16.6**	15.6

Table 4.2. Spearman rank-order correlation coefficient for the word similarity dataset MEN [**4**] and for the verb similarity dataset SimVerb [**15**]. Since we limit our vocabulary to 50k, we include the coverage of the dataset. If either word in a pair is unknown, we ignore it. The 256 bits version of the compressed GloVe embeddings were trained similarly to the 40 bits version.

4.1.2. Unsupervised Clustering

The second qualitative evaluation we did was to cluster our results to have a broader view of how our latent space behaved. To do so, we used k-Modes [**18**], an extension of the popular k-Means algorithm that more adequately clusters categorical values. We used 300 clusters with 12 random initializations[4]. This amount of clusters was chosen to obtain easily

[3]The dimensionality of the vectors must be compatible with the hardware registers' width for maximum performance optimization.

[4]As in k-Means, random centroids are initialized at the beginning of training. To ensure the best results, multiple model are fitted and the best one is kept.

interpretable word lists although it leads to some redundancy. Some are showed in table 4.3. Again, they seem to be predominantly grouped around a specific grammatical property. Cluster 7 contains adverbs; 16 the present participle inflection of verbs; 18 the third person inflection of verbs; 44 the past inflection of verbs; 55 the plural form of nouns; and 108 different adjectives. Semantic similarities are still present. In cluster 55, the words all seem to pertain to the concept of buildings or structures in particular. Similar behaviour can also be observed for names, years, integers and floating point numbers. The same conclusions as in 4.1.1 can be made regarding the interpretation of this behaviour and how this apparent grammatical clustering might relate to semantics.

Clusters									
6	**7**	**16**	**18**	**31**	**37**	**44**	**55**	**119**	**180**
classical	brilliantly	sending	proclaims	revisit	12.8	went	walls	69	90th
representational	moderately	discarding	expects	uncover	15.7	signaled	gates	59	81st
esoteric	somewhat	retrieving	interprets	bestow	11.9	changed	pillars	47	74th
meditative	largely	delivering	observes	clarify	16.7	came	rooftops	43	80th
hindustani	predominantly	relinquishing	relates	discuss	19.8	boomed	panels	54	75th
medieval	uniformly	disregarding	proves	recreate	13.1	reverted	shelves	56	39th
rabbinic	remarkably	interrogating	resembles	criticise	5.2	hacked	façades	105	46th
utilitarian	characteristically	locating	recalls	contemplate	23.3	transferred	headstones	58	87th
symbolist	vastly	executing	ceases	rehearse	20.1	ventured	columns	62	30th
gothic	tremendously	revising	inspires	acknowledge	18.8	continued	beams	159	eighteenth
dialectic	noticeably	holding	resumes	applaud	19.2	proceeded	spheres	103	48th
renaissance	undeniably	exercising	declares	awaken	12.0	ensues	eaves	102	seventeenth
rococo	perfectly	halting	mimics	remember	8.3	offered	hangings	76	55th
symbolic	universally	inciting	foresaw	rearrange	22.2	transitioned	mists	61	84th
egalitarian	hitherto	examining	begs	criticize	19.5	switched	fundamentals	202	49th
cubist	proportionally	boosting	believes	ignore	9.5	flipped	lamps	107	91st
confucian	doubly	protecting	confirms	imitate	14.1	strayed	ramparts	57	67th
contemplative	uniquely	assigning	applies	appreciate	9.6	redirected	doorways	53	25th
alchemical	comparatively	compensating	succeeds	discern	9.8	forwarded	marbles	66	twentieth
literary	virtually	securing	explains	conceal	23.5	beckoned	rings	46	10th
sociological	highly	pursuing	pretends	reunite	12.4	shifted	strata	51	50th
contemporary	less	propelling	dislikes	broaden	12.1	presented	runes	77	41st
pedagogical	strikingly	forging	absorbs	confirm	12.2	agreed	shards	127	70th
postcolonial	wonderfully	investigating	acknowledges	illustrate	0.50	conveyed	panelling	52	fifteenth
pictorial	decidedly	renewing	favours	amuse	17.0	amounted	facies	119	61st
neoclassical	oddly	dialing	attaches	reject	6.8	appealed	molars	452	37th
bureaucratic	distinctly	choosing	arises	avert	23.7	agreeing	shackles	366	59th
patriarchal	finely	removing	contends	confide	17.8	consigned	perils	291	36th
humanistic	densely	extracting	feels	forestall	25.1	returned	bricks	205	60th
linguistic	nicely	hoisting	thinks	offend	22.7	turned	stacks	168	40th
italic	reasonably	spurring	settles	compose	19.3	reacted	papyrus	183	31st
gregorian	mildly	surrendering	resolves	indulge	18.4	vanished	corridors	483	64th
secular	hugely	seizing	emerges	absorb	21.9	confined	moons	492	65th
sentimental	infinitely	resisting	considers	improvise	18.3	declined	torches	471	79th
psychoanalytic	sufficiently	sealing	presents	rethink	8.4	inserted	slabs	419	twelfth
monumental	excessively	installing	merges	recognise	32.8	rang	bulkheads	522	32nd
poetic	mathematically	taking	asserts	explore	8.6	confessed	gables	384	27th
pentecostal	marginally	lifting	renders	forgo	24.5	paused	vents	326	88th
realist	vividly	constructing	precedes	portray	8.7	communicated	partitions	209	22nd

Table 4.3. Subset of 10 word clusters taken from the 300 computed using k-Modes. The chosen 40 words for each are in ascending order of hamming distance from the centroid of their respective cluster.

Jaccard	Dice	3W-Jaccard	Nei&Li	Sokal&Sneath-I	Sokal&Michener	Sokal&Sneath-II	Roger&Tanimoto	Faith	Gower&Legendre	Intersection	Innerproduct	Russel&Rao	Hamming	Euclid	Squared-Euclid
sucks	sucks	sucks	sucks	sucks	stirs	stirs	stirs	fills	stirs	świętokrzyskie	npsl	---	stirs	stirs	stirs
stirs	stirs	stirs	stirs	stirs	sucks	sucks	sucks	stirs	sucks	ihl	ekstraklasa	gel	sucks	sucks	sucks
fills	fills	fills	fills	fills	dissolves	dissolves	dissolves	sucks	dissolves	chl	40-man	—	dissolves	dissolves	dissolves
withdraws	withdraws	withdraws	withdraws	withdraws	treats	treats	treats	dissolves	treats	brampton	26.0	shredding	treats	treats	treats
dissolves	dissolves	dissolves	dissolves	dissolves	clears	clears	clears	heals	clears	guelph	naia	relentlessly	clears	clears	clears
treats	treats	treats	treats	treats	withdraws	withdraws	withdraws	stings	withdraws	yugoslavia	2011–2012	unprepared	withdraws	withdraws	withdraws
fakes	fakes	fakes	fakes	fakes	fills	fills	fills	---	fills	baronetage	chl	melt	fills	fills	fills
stings	stings	stings	stings	stings	fakes	fakes	fakes	treats	fakes	npsl	groningen	â yes	fakes	fakes	fakes
heals	heals	heals	heals	heals	ignores	ignores	ignores	collapses	ignores	diocese	xfl	disappear	ignores	ignores	ignores
collapses	collapses	collapses	collapses	collapses	stings	stings	stings	withdraws	stings	cambridgeshire	pittsfield	¦	stings	stings	stings
clears	clears	clears	clears	clears	lowers	lowers	lowers	fakes	lowers	usfl	44th	flow	lowers	lowers	lowers
melts	melts	melts	melts	melts	begs	begs	begs	melts	begs	cincinnati	saarland	grated	begs	begs	begs
warms	warms	warms	warms	warms	assembles	assembles	assembles	mimics	assembles	household	105th	magnified	assembles	assembles	assembles
dries	dries	dries	dries	dries	distinguishes	distinguishes	distinguishes	paints	distinguishes	flintshire	usl	unharmed	distinguishes	distinguishes	distinguishes
explains	explains	explains	explains	explains	shrugs	shrugs	shrugs	explains	shrugs	hampshire	knvb	contrived	shrugs	shrugs	shrugs

Mean-Manhattan	Vari	SizeDiff	ShapeDiff	PatternDiff	Lance&Williams	Hellinger	Chord	Cosine	Gilbert&Wells	Forbesi	Fossum	Sorgenfrei	Mountford	Mcconnaughey	Tarwid
npsl	npsl	stirs	sucks	---	sucks	stirs	stirs	stirs	clears	clears	stirs	stirs	stirs	stirs	clears
ekstraklasa	ekstraklasa	sucks	stirs	gel	stirs	sucks	sucks	sucks	consumes	consumes	sucks	sucks	sucks	sucks	consumes
40-man	40-man	dissolves	fills	sucks	fills	fills	fills	fills	tosses	tosses	fills	fills	fills	fills	tosses
26.0	26.0	treats	clears	stirs	withdraws	dissolves	dissolves	dissolves	sucks	stirs	withdraws	dissolves	clears	dissolves	stirs
naia	naia	clears	fakes	clears	dissolves	withdraws	withdraws	withdraws	stirs	sucks	dissolves	withdraws	dissolves	withdraws	sucks
2011–2012	2011–2012	withdraws	treats	fills	treats	treats	treats	treats	fakes	fakes	treats	treats	withdraws	treats	fakes
chl	chl	fills	dissolves	collapses	fakes	fakes	fakes	fakes	ignores	extract	stings	fakes	treats	---	extract
groningen	groningen	fakes	withdraws	stings	stings	heals	heals	heals	distinguishes	ignores	heals	heals	fakes	stings	ignores
xfl	xfl	ignores	consumes	heals	heals	stings	stings	stings	extract	shrugs	collapses	stings	stings	heals	shrugs
pittsfield	pittsfield	stings	stings	consumes	collapses	collapses	collapses	collapses	shrugs	distinguishes	fakes	collapses	heals	collapses	distinguishes
44th	44th	lowers	collapses	inquires	clears	clears	clears	clears	pulls	consumed	---	clears	collapses	fakes	consumed
saarland	saarland	begs	heals	magnified	melts	---	---	---	consumed	pulls	clears	---	consumes	clears	pulls
105th	105th	assembles	---	grated	warms	sleeps	sleeps	sleeps	keeps	infects	assembles	sleeps	begs	gel	infects
usl	usl	distinguishes	melts	coughs	dries	communicates	communicates	communicates	smacks	cures	melts	communicates	mimics	communicates	cures
knvb	knvb	shrugs	paints	withdraws	explains	assembles	assembles	assembles	cures	smacks	dries	assembles	---	mimics	smacks

Kulczynski-II	Driver&Kroeber	Johnson	Dennis	Simpson	Braun&Banquet	Fager&McGowan	Forbes-II	Sokal&Sneath-IV	Gower	Pearson-I	Pearson-II	Person-III	Pearson&Heron-I	Pearson&Heron-II	Sokal&Sneath-III
stirs	stirs	stirs	clears	gel	stirs	sucks	gel	clears	stirs	stirs	stirs	sucks	sucks	chl	stirs
sucks	sucks	sucks	sucks	---	sucks	stirs	---	sucks	sucks	sucks	sucks	stirs	stirs	44th	sucks
fills	fills	fills	stirs	subside	treats	fakes	fills	stirs	clears	fills	fills	fills	fills	npsl	dissolves
dissolves	dissolves	dissolves	consumes	em	dissolves	clears	clears	consumes	consumes	clears	clears	clears	clears	groningen	treats
withdraws	withdraws	withdraws	fakes	shatter	withdraws	lowers	collapses	fakes	fakes	withdraws	withdraws	dissolves	dissolves	ekstraklasa	clears
treats	treats	treats	treats	firsthand	undermines	inspires	stings	ignores	fills	treats	treats	withdraws	withdraws	brampton	withdraws
---	---	---	dissolves	unprepared	resists	resists	heals	extract	treats	dissolves	dissolves	treats	treats	naia	fills
stings	stings	stings	withdraws	—	fakes	undermines	consumes	distinguishes	withdraws	fakes	fakes	fakes	fakes	xfl	fakes
heals	heals	heals	extract	harmless	lowers	ignores	freezes	shrugs	dissolves	collapses	collapses	heals	heals	40-man	ignores
collapses	collapses	collapses	distinguishes	unharmed	inspires	distinguishes	magnified	dissolves	tosses	stings	stings	collapses	collapses	pittsfield	stings
fakes	fakes	fakes	shrugs	gone	fills	extract	encapsulated	treats	extract	heals	heals	stings	stings	26.0	lowers
clears	clears	clears	ignores	obeyed	melts	shrugs	grated	withdraws	shrugs	consumes	consumes	consumes	consumes	2011–2012	begs
gel	gel	gel	fills	crave	mimics	consumes	inquires	tosses	distinguishes	explains	explains	warms	warms	saarland	assembles
communicates	communicates	communicates	undermines	magnified	paints	precedes	coughs	pulls	ignores	mimics	mimics	dries	dries	headteacher	distinguishes
mimics	mimics	mimics	resists	coughs	begs	makes	crave	cures	heals	dries	dries	melts	melts	minneapolis	shrugs

Sokal&Sneath-V	Stiles	Yuleq	Yulew	Kulzynski-I	Disperson	Hamann	Michael	Baroni-urbani&buser-I	Baroni-urbani&buser-II	Peirce	Eyraud	Tarantula	Ample	Goodman&Kruskal	anderberg
sucks	sucks	---	---	sucks	clears	stirs	fills	stirs	stirs	tosses	sucks	gel	gel	sucks	sucks
stirs	stirs	gel	gel	stirs	consumes	sucks	sucks	sucks	sucks	consumes	stirs	---	---	stirs	stirs
fills	fills	stirs	stirs	fills	sucks	dissolves	stirs	fills	fills	inherits	clears	fills	fills	treats	fills
treats	clears	sucks	sucks	withdraws	stirs	treats	treats	dissolves	dissolves	'll	fills	heals	heals	fills	treats
withdraws	dissolves	fills	fills	dissolves	tosses	clears	dissolves	withdraws	withdraws	clears	fakes	stings	stings	withdraws	withdraws
dissolves	withdraws	clears	clears	treats	fakes	withdraws	withdraws	treats	treats	invokes	withdraws	collapses	collapses	dissolves	dissolves
clears	treats	heals	heals	fakes	---	fills	heals	fakes	fakes	dismisses	treats	encapsulated	encapsulated	fakes	dries
fakes	fakes	stings	stings	stings	gel	fakes	stings	clears	clears	confronts	dissolves	grated	grated	clears	begs
stings	heals	collapses	collapses	heals	fills	ignores	collapses	collapses	collapses	lets	consumes	freezes	freezes	sleeps	communicates
heals	stings	consumes	consumes	collapses	treats	stings	---	heals	heals	' ll	---	inquires	inquires	mimics	assembles
collapses	collapses	treats	treats	clears	dissolves	lowers	fakes	stings	stings	identifies	stings	magnified	magnified	explains	paints
sleeps	sleeps	dissolves	dissolves	melts	withdraws	begs	communicates	mimics	mimics	revives	heals	coughs	coughs	shoots	sleeps
warms	communicates	withdraws	withdraws	warms	extract	assembles	mimics	paints	paints	differentiates	collapses	melt	melt	dries	melts
assembles	mimics	fakes	fakes	dries	shrugs	distinguishes	sleeps	begs	begs	importers	extract	weep	weep	warms	warms
mimics	shoots	coughs	coughs	explains	ignores	shrugs	paints	warms	warms	persuades	shrugs	â yes	â yes	melts	explains

Table 4.4. Comparison of the 15 nearest words for the word *eats* based on different binary similarities and distance measures from [**6**].

4.2. Relations in Bit Differences

Part of our goal was to achieve interpretability. To see if any concepts could easily be linked to learnt features, we examined how bits differed between embeddings of words with known properties. Semi-automatic word lists were built (see appendix A) around these four inflections : plurals, past, third person and present participle. Another similar experiment was done with a manual list of 27 countries and their associated nationality. Then, for each of these lists, we used a bitwise XOR between the embeddings of each word pair and summed the results to get the probability of each bit to be different. We did this process with our embeddings and with the compressed GloVe embeddings from 4.1.1 as a comparative. In 4.2.1, we explain how certain patterns emerged and in 4.2.2 we explore the interpretability of these patterns.

4.2.1. Emergence of patterns

Our assumption was that if a feature has a high probability to differ between words with a known property (e.g. past inflection), then it must be indicative of this property. Using our vectors, we looked at the probabilities of bit being different for the five word lists. In all five cases, unique sets of bits seem to expose the desired property. Figure 4.1 shows that about 6 bits have a high chance of changing between a singular word to a plural word. From our assumption, this means that the concept of plurality would be encoded using these 6 bits by the model. Similarly for past inflections, figure 4.2 shows that at least 8 bits are strongly involved. For present participle inflections, figure 4.3 also shows at least 8 bits are involved. This phenomenon is not as evident for the third person inflections, but from figure 4.4 we can see at least 4 significant bits. As for the countries and their nationality, figure 4.5 shows that 4 bits are heavily involved and 5 are not involved at all. Bits that always stay the same could aslo be indicative of a property, but we leave this for future work. Our compressed GloVe embeddings do not seem to capture the same relations. In all of the cases, except maybe for the nationalities, the probabilities are very similar. If we look at the standard deviation (see table 4.5), we can see that for the compressed version there is an average of 5.6% deviation from the mean as opposed to an average deviation of 15.2% for our work. In other words, from our representation, we can observe that some bits either have a significant chance to differ for a relation or a significant chance to be the same. This can't be concluded for the compressed GloVe embeddings. This correlates with the nearest neighbors experiments (see 4.1) where grammatical features seemed to be encoded. These sets of bit could be considered as semantic morphemes, i.e., they seem to be indicators of common meaning modifiers. See appendix B for the un-ordered version of the probabilities to see exactly which bits are more involved. These results only indicate that certain patterns are exposed. In the next section, we try to gain insight on potential practical applications for them.

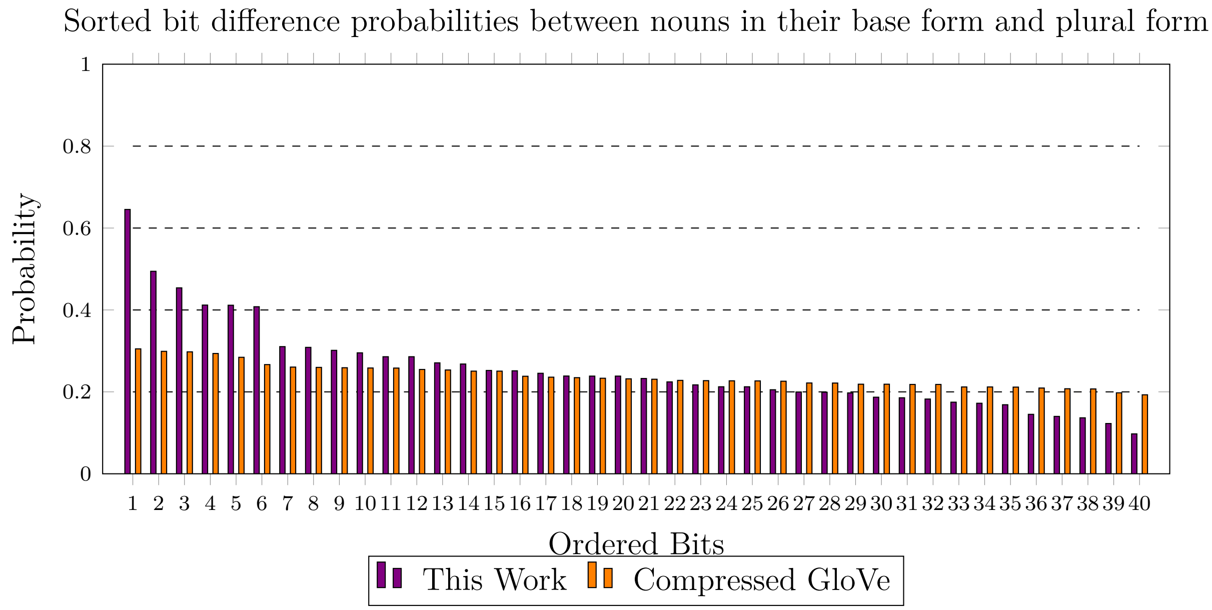

Fig. 4.1. The ordered probabilities of a certain bit being different between the embedding of the base form of a noun and the embedding of its plural form. A semi-automatic list of 4948 nouns present in the vocabulary was used (see appendix A.1.1). Here the ordered bits do not necessarily match between both techniques (see appendix B.1 for an unordered version of this graph).

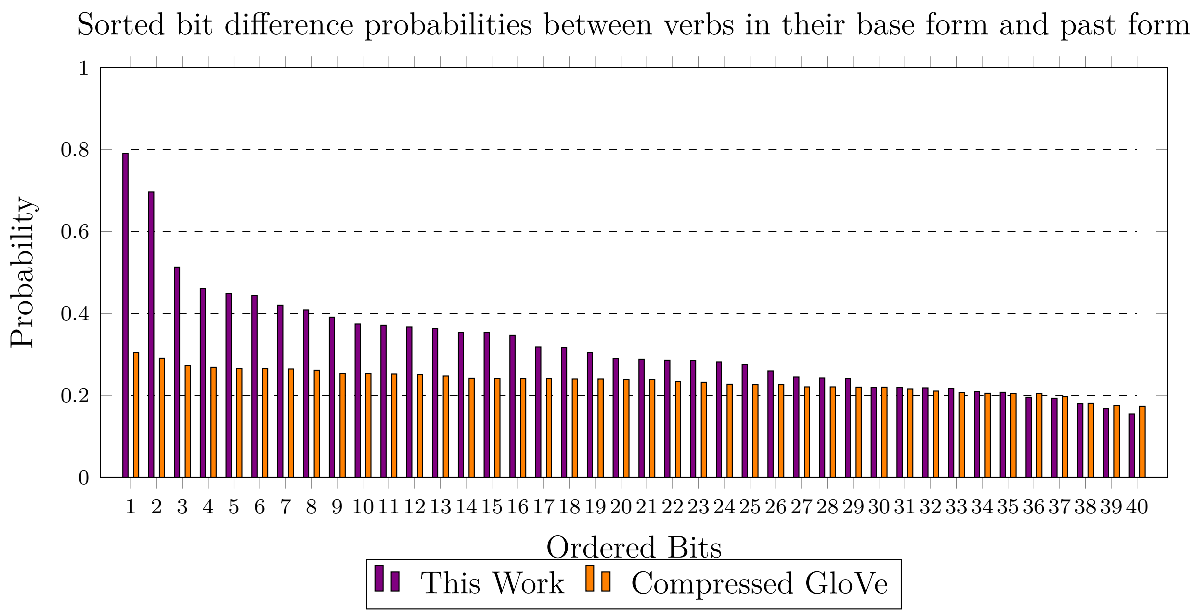

Fig. 4.2. The ordered probabilities of a certain bit being different between the embedding of the base form of a verb and the embedding of its past form. A semi-automatic list of 1638 verbs present in the vocabulary was used (see appendix A.1.2). Here the ordered bits do not necessarily match between both techniques (see appendix B.2 for an unordered version of this graph).

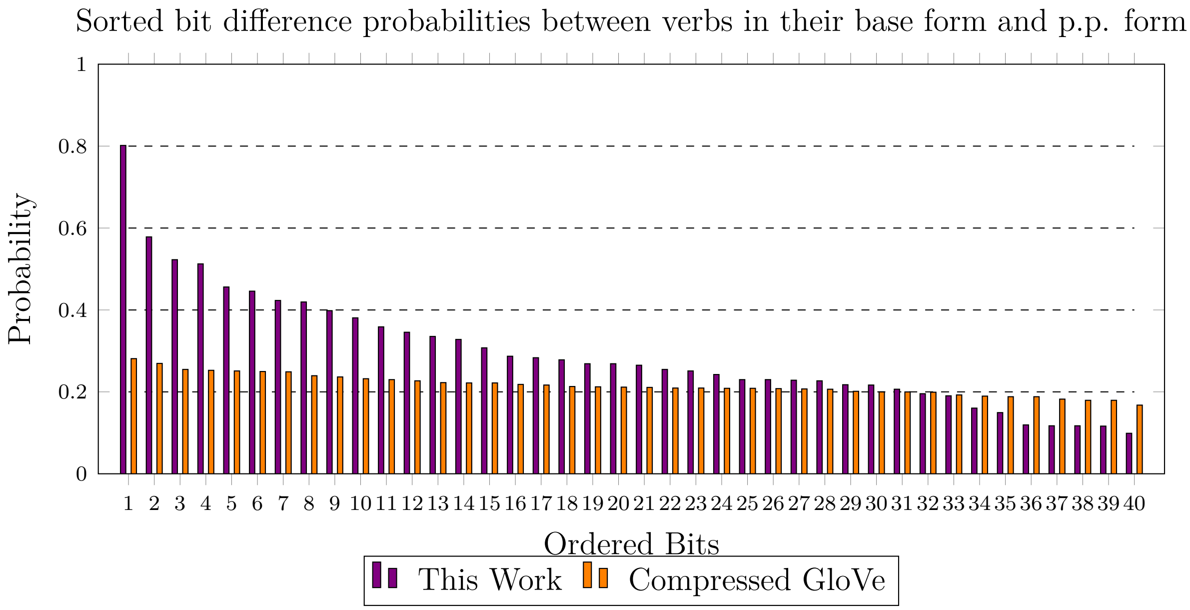

Fig. 4.3. The ordered probabilities of a certain bit being different between the embedding of the base form of a verb and the embedding of its present participle form. A semi-automatic list of 1366 verbs present in the vocabulary was used (see appendix A.1.3). Here the ordered bits do not necessarily match between both techniques (see appendix B.3 for an unordered version of this graph).

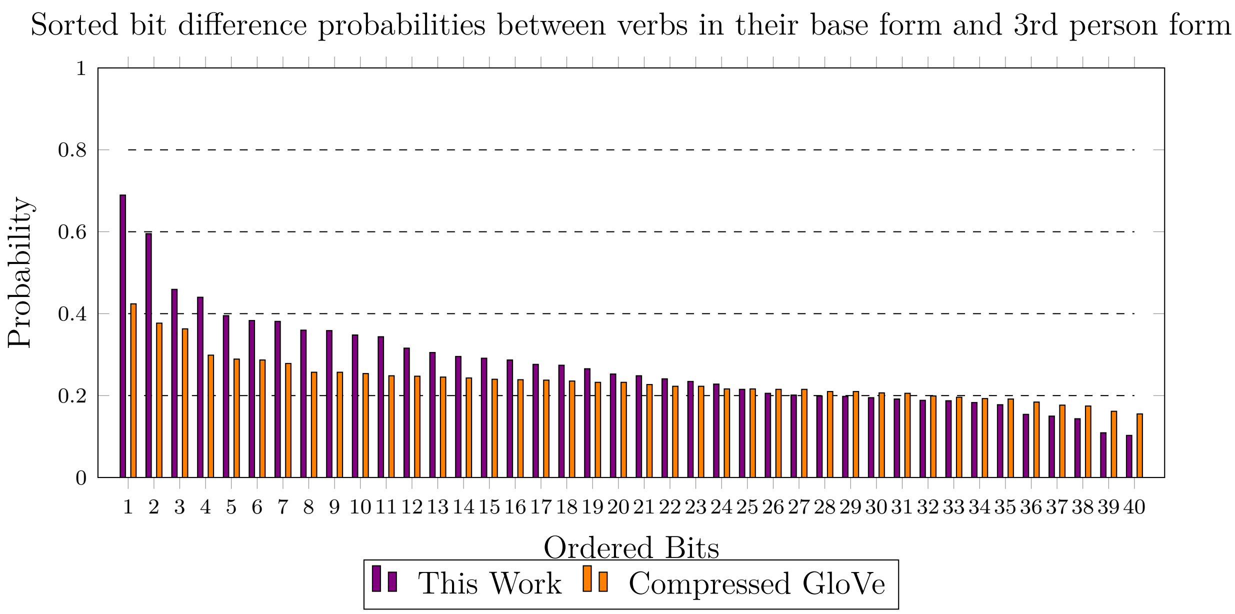

Fig. 4.4. The ordered probabilities of a certain bit being different between the embedding of the base form of a verb and the embedding of its third person form. A semi-automatic list of 934 verbs present in the vocabulary was used (see appendix A.1.4). Here the ordered bits do not necessarily match between both techniques (see appendix B.3 for an unordered version of this graph).

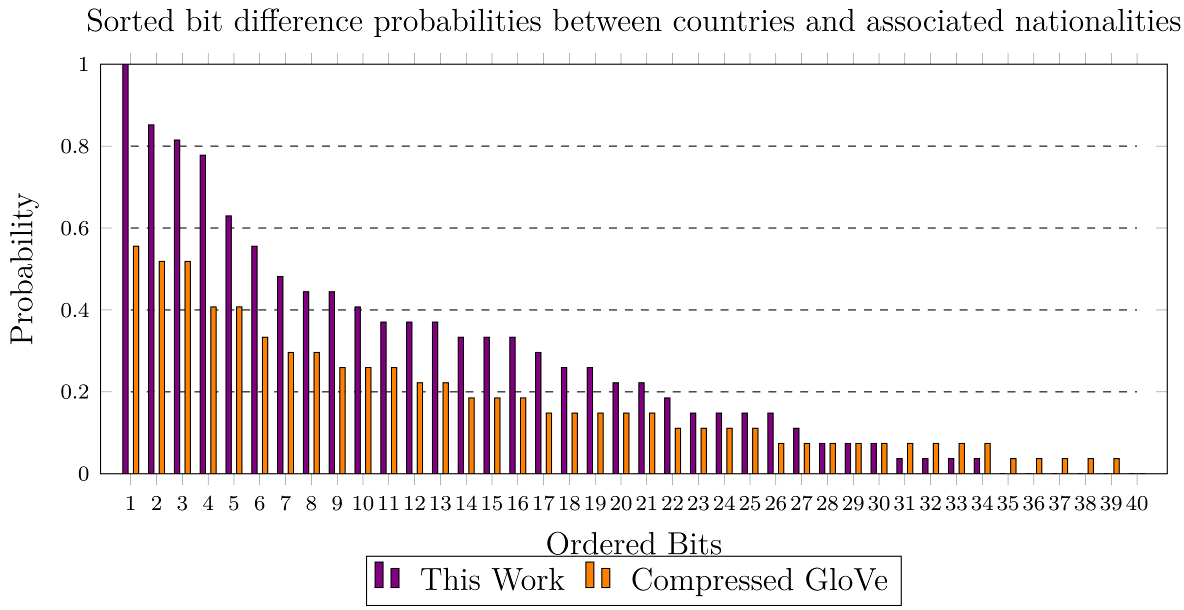

Fig. 4.5. The ordered probabilities of a certain bit being different between the embedding of a country and the embedding of its associated nationality. A manual list of 27 countries present in the vocabulary was used (see appendix A.2.1). Here the ordered bits do not necessarily match between both techniques (see appendix B.5 for an unordered version of this graph).

Bit diff. probabilities	**Standard Deviation**	
	This Work	Comp. GloVe
Base ↔ Plural	0.1084	0.0283
Base ↔ Past	0.1309	0.0287
Base ↔ P.P.	0.1428	0.0253
Base ↔ 3.P.	0.1208	0.0543
Country ↔ Nationality	0.2583	0.1420

Table 4.5. Standard deviation in the bit difference probabilities for our embeddings and the compressed GloVe embeddings from 4.1.1.

4.2.2. Interpretability

In our motivation, we mentionned how we believed the representation of a word should ideally be mostly the same between its root and its inflected forms. We also argued that using morphology is not appropriate to achieve this. It was established that certain bit patterns seem to encode semantical properties, which points to this behaviour having been learned by the model without explicit tokenization (e.g. *cats* split into *cat* · *s*). To try to confirm this conjecture, we first took the 10 bits[5] that were most likely to be different for a

[5]We chose 10 because it empirically gave the best overall results.

specific word list (e.g. for singular to plural, figure B.1 shows these to be: 26, 7, 17, ..., 11). For each of these bits, we computed the probability that it flipped to 1 and the probability that it flipped to 0. If the probability of flipping to a 1 was bigger, we assumed we must force this specific bit to 1 to represent the property (e.g. plurality). The opposite process was applied if the probability of flipping to 0 was bigger. We then took the embedding of the base form of a word and modified it by forcing the 10 bits pattern we just computed and leaving the other 30 bits untouched. We did not simply flip the bits blindly. An example of this is again for singular to plural, when bit 26 is flipped, it is most likely flipped to a one. If the embedding we are modifying already has a 1 for this bit position, it will not be flipped to a 0, it will stay the same.

The closest neighbors for these adjusted embeddings were then computed using the hamming distance. Results are presented for plurality in table 4.6 and for present participle in 4.7. The results are similar for the other inflections. In both cases, the closest words of the adjusted representation expose the desired property, but the neighborhood is not close to the expected neighborhood, that is to say simply adjusting the base embedding does not yield the inflected embedding. Lets assume the bit patterns represent our proposed sememe formulation. Each pattern is considered as a single semantic unit in this asusmption, thus forcing this pattern on an embedding is comparable to altering part of the latent meaning of a word. Both the *car* and the *eat* examples validate this. By analyzing the nearest neighbors of the adjusted embeddings, all the related words expose the desired semantical characteristic. The adjusted vector for *car* seems to be closer to the actual vector of *cars* compared to *eat* and *eating*. This correlates with the neighborhoods in table 4.1 where the nouns *computer* and *computers* were closer to word semantically similar to them individually and where the verb *eat* and its inflected forms seemed to be closer to words sharing more of a contextual semantic similarity. This goes to show that the separation between what we consider to be latent basic units of meaning is not as well-defined as we would have hoped, but they are present. This supports our earlier hypothesis that features are encoded as multiple bits, hence navigating the latent space is not trivial. As in 4.1.1, it also hints that a bigger vector size might have a big impact. Bigger and smaller embeddings were tried, but no results seemed to encode as much information as this configuration (which was chosen arbitrarily). Since tuning the hyper-parameters for our implementation was relatively difficult, we focused on analyzing these good results to understand what the model was learning and we let the exploration of different dimensionalities and the extraction of discrete latent units for future work.

car		car (adjusted)		cars	
Relation	H.D.	Relation	H.D.	Relation	H.D
car	0	envelopes	4	cars	0
truck	3	tires	5	automobiles	1
suv	3	accessories	5	trucks	2
pickup	4	suvs	5	motorcycles	3
brake	4	dials	5	scooters	3
luggage	4	pallets	6	electronics	4
hovercraft	4	handguns	6	vehicles	4
motor	4	cables	6	bikes	4
vehicle	4	machines	6	machines	4
tractor	5	bumpers	6	elevators	4
wagon	5	wheels	6	streetcars	5
drugstore	5	strips	6	cabs	5
motorcycle	5	wires	6	accessories	5
shotgun	5	stickers	6	tires	5
windshield	6	vendors	6	tractors	5

Table 4.6. Lists of closest words based on their hamming distance (H.D.). The middle column represents the base form embedding adjusted with the computed 10 bits pattern representing the plurality feature.

eat		eat (adjusted)		eating	
Relation	H.D.	Relation	H.D.	Relation	H.D
eat	0	modifying	5	eating	0
fly	6	selecting	5	watering	2
choose	6	dispatching	5	chewing	3
explode	7	encountering	5	hoarding	3
steal	7	including	5	feeding	4
enjoy	7	intercepting	5	drinking	4
skim	8	possessing	5	smoking	4
smelled	8	affording	5	tagging	5
mew	8	capturing	6	pruning	5
dat	8	evading	6	packing	5
birthdays	8	destroying	6	roasting	5
lili	8	eliminating	6	cooking	5
crave	8	transporting	6	healthy	5
breathe	8	inspecting	6	nesting	5
infect	8	accessing	6	washing	5

Table 4.7. Lists of closest words based on their hamming distance (H.D.). The middle column represents the base form embedding adjusted with the computed 10 bits pattern representing the present participle feature.

4.3. Generalization

As explained in 3.3.1, we limited our vocabulary to the 50,000 most frequent words. This only represents 0.5% of the whole training corpus vocabulary. This is a limiting factor if we want to use the embeddings in any downstream task. Assuming the model has a good grasp on the relations between words, we should be able to collect windows for unknown words and generate a new embedding. We can take advantage of the fact that we defined our training methodology as a matching problem between the middle words (i.e. the embeddings) and their surrounding context. By processing these new contexts with our trained model, we obtain a latent representation that matches the representation that the new middle word should have. This flexibility gives our work a lot of potential. Embeddings could be generated on new corpora with a lot of emerging words (e.g. *sup* for *what's up*). Technical and rare words could get comparable embeddings by gathering a few context windows to facilitate their understanding. Another application could be the easy identification of unknown named entities by training a classifier based on the binary embeddings. These are currently only conjectures and further reasearch would need to be done, but the results that we will present are favorable to these practical applications.

Formally, like the training phase, we collected windows of 51 words in a new corpus and then we used f_θ^D (fig. 3.1) to get a representation for each window. We averaged all the computed latent vectors for a single word and then applied a hard threshold (see eq. 3.2.2) to get the final binary representation. The source corpus was the complete SimpleWiki dump of September 20th 2020[6] and only the unknown words with at least 10 windows were kept. The same text preprocessing as in 3.3.1 has been used. We processed the words in descending order of their windows count. As soon as an embedding was computed, the word was now considered known, which meant that if it appeared in the context of any subsequent word to process, its new representation was used instead of the *<unk>* token's embedding.

In table 4.8, we can see that the model is able to generate new embeddings that are related to similar words. We observed that new words tend to have a neighborhood composed of mostly other new words. Same thing for the words used in the original vocabulary (e.g. *covid-19*, *evaporates*, *canadian* and *updated*). This is not a problem, since the closest words are still semantically related. In some cases (e.g. *broadcasted* and *tables*), the new words integrate seamlessly into the already existing vocabulary. This makes our idea of translating to a language independent representation really promising and further supports our observations on the seemingly presence of latent discrete units of meaning. The words *evaporates* and *platypus* are examples of predicted representations collapsing to the same vector. This could be explained as the model not being precise enough, hence requiring more capacity, but even

[6]`https://dumps.wikimedia.org/simplewiki/latest/simplewiki-latest-pages-articles-multistream.xml.bz2-rss.xml`.

covid-19	(1293 win.)		**evaporates**	(61 win.)		**broadcasted**	(46 win.)		**platypus**	(74 win.)	
Relation	H.D.	New	Relation	H.D.	New	Relation	H.D.	New	Relation	H.D.	New
leukaemia	3	✓	hardens	0	✓	screened	5		seahorse	0	✓
ebola	3		collides	2	✓	play	5		ibises	0	✓
hcv	4	✓	hibernates	2	✓	performed	5		artiodactyls	0	✓
barbiturates	4	✓	circulates	2	✓	spawned	6		tapir	0	✓
lsd	4		condenses	2	✓	filmed	6		cassowary	0	✓
eas	4		diffuses	3	✓	redone	6		peafowl	1	✓
rubella	4	✓	ejaculates	3	✓	produced	6		marsupial	1	✓
dbs	5	✓	dries	3		choreographed	6		mastiff	1	✓
gsr	5	✓	melts	3		arranged	6		bullfrog	1	✓
ivf	5		germinates	3	✓	commenting	6		orangutan	1	✓
flu	5		corrodes	4	✓	streamed	6		dhole	1	✓
mortality	5		preys	4	✓	sung	6		ornithischia	1	✓
rhinovirus	5	✓	chews	4	✓	sang	6		dugong	1	✓
beeswax	5	✓	vibrates	4	✓	serialization	6		honeyeater	1	✓
suffocation	5	✓	dissolves	4		scripted	6		quokka	1	✓
tables			**canadian**			**updated**			**dog**		
Relation	H.D.	New	Relation	H.D.	New	Relation	H.D.	New	Relation	H.D.	New
staves	3	✓	australian	4		facelifted	2	✓	labradoodle	3	✓
accompaniments	4	✓	norwegian	5		discontinued	5		toddler	3	
slots	4		tasmanian	5		restarted	5		hound	3	
booths	4		gambian	6		overhauled	5		dachshund	3	✓
hangings	4		banff	6		serviced	5		horseman	3	
frets	4	✓	fivb	6		ported	5		sheepdog	3	✓
settings	4		faroese	6		trademarked	5		goat	3	
columns	5		tiwi	6	✓	configured	5		shorthair	3	✓
wristbands	5	✓	estonian	6		reopened	5		whale	4	
files	5		soweto	6		shipped	6		bobtail	4	✓
shelves	5		cbc	6		available	6		lop	4	✓
whiteboards	5	✓	kenyan	6		retrofitted	6		thylacine	4	✓
scopes	5		zimbabwean	6		07	6		elephant	4	
imprints	5		swedish	6		resold	6		sylph	4	✓
polonaises	5	✓	batavian	6		deprecated	6		squirrel	4	

Table 4.8. Lists of closest words based on their hamming distance (H.D.). The top row are words that were computed from new windows from SimpleWiki with the amount of windows used. The bottom row are words that were already in the vocabulary. The *New* column indicates whether the word was added from SimpleWiki or not.

this constrained model does not suffer significantly from this problem. In table 4.9, we also show that the amount of windows used does not need to be important. Even with only 10 windows to generate their embedding, the first three of these examples manage to have a representation that is related to other entities with similar properties. However, a problem appears when we look at the new word *s10*. To put in context, this refers to the *Galaxy S10* phone from Samsung. From the closest neighbors, we can see that the model focuses on both **galaxy** (e.g. *hipparcos*, *extrasolar*, *eon*) and **s10/samsung** (e.g. *s5*, *s6*, *s12*). The small amount of windows makes it hard for the model to make the distinction. Although we mention this as a problem, it is hard to really affirm that it is an unwanted behaviour. It seems our method uses adequate language cues to produce proper embeddings. Also, from our discussion in section 4.1.1, the hamming distance might not fully represent the quality of the learned features. Further research should be done to better understand the relations captured by the model and to see if the generalization is in fact adequate.

dinghies	(10 win.)		**rockhole**	(10 win.)		**memoirist**	(10 win.)		**s10**	(10 win.)	
Relation	H.D.	New	Relation	H.D.	New	Relation	H.D.	New	Relation	H.D.	New
competitors	3		liana	4	✓	sexologist	2	✓	s5	2	✓
belts	5		solanaceae	4	✓	ethnologist	3	✓	hipparcos	3	✓
podiums	5		protea	4	✓	humorist	3		extrasolar	4	✓
snowboarders	5	✓	frog	4		ethnographer	4	✓	1987a	4	✓
floats	5		ruminant	4	✓	logician	4	✓	eon	4	
substitutes	5		goat	5		japanologist	4	✓	s6	4	✓
darts	5		gracilis	5		caricaturist	4	✓	s12	4	✓
favourites	5		cappella	5		moviemaker	4	✓	extragalactic	4	✓
racers	5		mimosa	5	✓	frontiersman	4	✓	f4	4	
openers	5		prostitute	5		neuroscientist	5	✓	quasar	4	
kayaks	6		jacaranda	5	✓	grammarian	5	✓	animated	4	
spots	6		subtribe	5	✓	comedienne	5	✓	hikari	4	✓
sleds	6	✓	toddler	5		neurobiologist	5	✓	pulsar	4	
gunsmiths	6	✓	sophora	5	✓	speechwriter	5	✓	f2	5	
mmorpgs	6	✓	amborella	5	✓	bureaucrat	5		antares	5	✓

Table 4.9. Lists of closest words based on their hamming distance (H.D.) for new words from SimpleWiki with a limited amount of windows. The *New* column indicates whether the word was added from SimpleWiki or not.

4.4. Learned Bit Templates

In 4.1, it was established that similar words are close in embedding space, but no specific bit patterns were found that we could associate to a specific known concept. From our results in 4.2, we got the intuition that maybe some exist, but that they are not as explicit as we would have hoped. To confirm this, in 4.4.1 we try to identify bit templates from manual lists of words we judged similar (e.g. months of the year) and in 4.4.2 we try to identify bit templates by training a classifier on WordNet lexical categories.

4.4.1. From Manual Word Lists

The first technique we used was more systematic. We manually built lists of words we judged related to a common concept (e.g. months of the year). Then, from the embeddings in these lists, we computed the entropy for each bit position. We extracted a bit template by adding bits in ascending order of entropy. Each time we added a feature, we built a list of words matching the template exactly. We repeated this procedure until the group of words was mostly related to the desired concept.

Table 4.10 contains some of these concepts we were able to associate to a specific template. The resulting groups do not necessarily contain all possible related words from the vocabulary. Some interesting phenomenons can be observed. One is that all the months can be identified by a single template. This is important as it means this model has the potential to provide embeddings with humanly recognizable features. Another is that most of these groups represent their associated concept really well. Almost no unrelated words are present as the ones that we consider inadequate is relative to our judgment (e.g. in the *Movie Directors* word group, all the names in bold are not directly linked to the label we gave it but they still had a role in cinema or television). These results prove to be promising although they affirm our doubt about the unexpected complexity of our latent representations as previously stated.

4.4.2. From WordNet Lexical Groups

The second technique we used to learn bit templates is more flexible. We used the fact that WordNet categorizes all the words by their different POS and meaning (see appendix C.2). This allowed us to train a model to predict the possible categories for a word based only on its binary embedding.

We first extracted all the words from our vocabulary that could be associated with at least one category from WordNet (see appendix D). Then we converted our embeddings to contain values in $\{-1, 1\}$ instead of $\{0, 1\}$. This gave an equal importance to the absence of a bit. The inputs were the adjusted embeddings $E \in \{-1, 1\}^k$, where k is the embedding size, and the outputs were Multi-Bernouilli[7] vectors $C \in [0, 1]^l$, where l is the number of categories. Each $C_i, i \in \{1,2,\ldots,l\}$ represents the probability of a word belonging to the i-th WordNet category. In this experiment, we assumed a 100% probability if the word could be matched, otherwise 0%. Consequently, C is equivalent to a binary vector. We trained a single layer feed-forward network with a sigmoid output that we paired with a Binary Cross Entropy loss. Using Adam as the optimizer with an initial learning rate of 0.0001, we trained until convergence, reducing the learning rate by a factor of 10 every 100 steps without improvement. No regularization was used. The results were the weight matrix $W \in \mathbb{R}^{k \times l}$, where each row represented the influence of a bit for predicting a particular cartegory, and the bias terms $b \in \mathbb{R}^l$, which we ignored for interpretation. With more data, we could have empirical estimates for C, which would more accurately weight a bit in relation to each category. This is left for future work.

The adjustment performed on the input allowed us to consider positive values in W to favor the presence of a bit (i.e. translates to a 1) and negative values to favor the absence of a bit (i.e. translates to a 0). For each row, we sorted in descending order each element by their absolute value. Starting by the bit that had the biggest influence, we gradually added bits to our template based on their sign until the amount of words[8] with an embedding matching

[7]See definition in 3.1.
[8]We used the whole vocabulary, not only the matching words used for training.

Concept	Words Following the Template
Months	may march september january june october july august april november december february
Uncountable	water oil coffee wine snow sand sugar rice meat cotton corn powder wheat livestock beef lumber liquor tar malaria foam munitions opium barley mosquito dynamite vodka steak mustard **sausage sugarcane** popcorn groceries cabbage vinegar **biscuit** sunflower hemp firewood fumes kerosene geiger lettuce manure **snowball** mussels tanning molasses harvester moonshine guano tofu veal creams lard nightshade walnuts scurvy slush blackouts
Jobs/Roles	director wife artist writer **firm** teacher agent engineer poet painter actors soldier ceo critic producers merchant reporter expert businessman citizen girlfriend worker physician farmer widow lover employee specialist photographer servant administrator monk instructor bride sculptor researcher novelist entrepreneur explorer employer dealer mistress playwright economist filmmaker educator screenwriter operative commentator sailor curator builder contractor apprentice banker thief tutor illustrator seller psychologist troupe butcher mechanic clown collaborator trader magician practitioner heroine salesman ensign cartoonist prostitute examiner broker heiress gardener millionaire blacksmith geologist navigator anthropologist traveler proprietor industrialist nanny biologist billionaire laird traveller bully goldsmith tailor baroness squire seaman theorist foe classmate aviator sociologist registrar dentist engraver magnate animator turk enthusiast prodigy robber financier adventurer schoolteacher aristocrat essayist pharmacist tycoon dramatist businesswoman publicist entomologist gambler strategist shoemaker overseer socialite rancher coo storyteller midwife informer centurion forester archivist wanderer texan horseman scotsman janitor custodian innovator dutchman sorceress housemate djinn irishman gunslinger philologist statistician laborer assessor playmate grocer innkeeper hairdresser journeyman cfo respondent delinquent postman yeoman draftsman sportscaster retainer pundit prospector draughtsman restaurateur plumber sportswriter seamstress ecologist villager **zealander** bookkeeper whaler realtor machinist pediatrician humorist bureaucrat handyman shipbuilder colonist cto satirist conservator cobbler leaguer watchmaker ophthalmologist newsreader educationist astrophysicist showman newscaster egyptologist joiner
Neg. Results	restraint leverage starvation greed resilience superstition necrosis **cannibalism** moderation **fluency** remission **zorro** toughness stuttering adversity negativity nightshade decorum castration quickening contagion brainwashing
Neg. Adjective	**good** difficult dangerous terrible complicated ugly horrible awkward vague unpleasant pathetic absurd unpredictable hopeless restrictive unbelievable mundane unbearable hysterical unsettling ecstatic nonexistent arduous devious coincidental unimportant aloof inexplicable ludicrous chaste alluring undiscovered otherworldly idiotic disconcerting unimaginable wasteful distasteful tiresome inconsequential hurtful jumpy inconceivable atrocious
Movie Directors	clark allen carson webb kramer hitchcock **manson** whitman chaplin hemingway hays henson downey macgregor **ladd** silverman kubrick griswold lovett redford shakur beckman mcclintock kohn **vonnegut durbin colter** whiteley fincher danko
Adverbs	while although soon today currently probably therefore perhaps unlike apparently meanwhile unless indeed certainly whilst whereas obviously nevertheless hence thereafter furthermore unfortunately whenever suggesting believing moreover nonetheless presumably basically accordingly presently nowadays **fearing prompting** arguably evidently firstly fortunately requesting conversely luckily undoubtedly ironically **sears** alleging lastly theoretically coincidentally notwithstanding anticipating incidentally henceforth alas reputedly suspecting admittedly **desiring** thirdly **archeologists** paradoxically **cyprian** unsurprisingly
Nationalities	german russian japanese african italian korean swiss czech hungarian brazilian austrian ukrainian belgian iranian argentine bulgarian indonesian filipino malaysian yugoslav estonian colombian icelandic ethiopian catalan peruvian bavarian burmese algerian moroccan somali belarusian cypriot mongolian crimean azerbaijani czechoslovak tunisian sicilian cambodian sudanese parisian singaporean kazakh paraguayan guatemalan uzbek yemeni angolan neapolitan rwandan panamanian nicaraguan tanzanian kuwaiti kyrgyz tajik zambian eritrean balinese laotian maldivian bhutanese karelian emirati mauritian czechoslovakian sundanese piedmontese manchurian

Table 4.10. Word lists (ordered by corpus frequency) that contain a manually defined bit pattern associated with known concepts. The words that we deemed unrelated are in bold.

perfectly the current template was really low (<30). This usually led to a very specific list of words related to the desired concept. To account for inaccuracies during training, we also relaxed the matching constraint to allow for a margin of n bits of error (e.g. for a 6 bits template, we would consider embeddings having at least $6 - n$ matched bits).

From the preliminary results in table 4.11, 4.12 and 4.13, the same conclusion as in 4.4.1 can be made, which is that conceptual connections seem to be encoded in our embeddings, especially for nouns. One of our initial hypotheses for this work was that a word could be translated to a series of language independent sememes. Basically, the representation of a word would be based on units of meaning we call sememes rather than grammatical units. For these 3 example categories, different sememes seem to comprise the extracted templates. By allowing a certain margin of error in our matching, one or more of these sememes changes, but most of them stay the same, which in theory should change the meaning slightly while still essentially representing the same thing[9]. We say that one or more sememe changes because we did not experiment enough to affirm how many bits are used to represent each of these latent units (e.g. in section 4.2, it appears multiple bits are involved). Empirical observations follow this application of the theory. For *noun.body*, most of the words can be related to the human body (e.g. chest, lips), but some words refer to the body of an animal (e.g. shell, snout), which is still highly related to the template. For *noun.animal*, multiple aspects of this concept are matched like specific species (e.g. dog, wasp), food (e.g banana, chestnut, leaf) and location (e.g. sea, mariana[10], marine). For *noun.time*, specific years are matched (e.g. 1707, 2010), different months (e.g. september, june, october), seasons (e.g. autumn, summer) and even actions often related to time or a time span (e.g. meal, prom, intermission).

We also did the same experiment with related verbs. The results are shown in table 4.14 and 4.15. In both cases, the definition giving by WordNet is vague when looking at the words they categorize in these lexical groups (see appendices D.4 and D.5), so we will evaluate the templates based on the category name as it seems to correlate better. If we look at the body related verbs, our pattern appears to be made up of sememes representing an action that could be done with our body. Obvious examples are *shrug*, *grin*, *sneeze* or *spasm*. More indirect examples could be : *scare*, by using your arm; *knit*, using your hands; or *abduct*, by carrying someone. Patterns learned for other verb groups were unfortunately not as relevant. One example is for verbs related to consumption. No interesting pattern emerges in the matching words. This could be caused by the ambiguity of the definition or simply because the amount of training words was low compared to other categories.

[9]This is similar to our *cat*, *dog*, *bite* and *claw* example in 1.2 where there obviously is a shared latent representation.

[10]Most likely related to the Mariana Trench, the deepest oceanic trench where sea life was discovered.

Margin	Words fitting the template for ***noun.body***: $* 0 1 * 1 * * 0 1 1 * * 1 * * * * 1 * * * * 0 * 0 0 0 * 0 0 1 * * * * * * * * 1$
0	jaw forearm
1	teeth cheek bone chin costa wrist spine fin thigh tooth abdomen snout torso calf paw nape ni**les thorax pelvis cartilage suture cytoplasm tibia ligaments vertebra ventricle epidermis sternum testicles humerus fibula callus
2	body lips chest shoulder neck finger tongue belt shell tail knee forehead muscle muscles skull hips tissue belly streak fork receptor eyebrow limbs tissues fingertips substrate limb cavity jaws aura posture h*ps muzzle lashes lobe groin streaks colon insides ligament medial strides scalp sheath microscope hardness seam embryo paws c*ck saliva eyelashes spindle tendon nipples vagina forearms biceps intestine collarbone strut forties gash allele uterus underparts n**ples forefinger moustache mandible rosette ovary femur fingertip carapace extremities hemoglobin chromatin cilia tendons acidity gums cheekbone columella atrophy colouration prefrontal ovaries bicep pouches cl*t reticulum fossa aorta ketone mucosa septum lamina pigmentation crustal dentition solute jawline umbilicus lithosphere follicle cuticle reuptake undersides armpit esophageal doublet orifice

Table 4.11. Incremental list of words that fit the template within *margin* bits of error. By WordNet's definition, the template should represent : *nouns denoting body parts.* See appendix D.2 for a list of words used for training.

For both the nouns and the verbs, potential human interpretable bit codes exist. By inspecting certain bit positions, it seems that it can be established whether a word represents a certain concept or not. It is clear that interpretability is better than any continuous counterpart, but again, more research needs to be done to better understand how features are exposed. With greater dimensionality, maybe a 1-to-1 relation could exist between a single bit and our proposed sememe unit.

Margin	**Words fitting the template for *noun.animal*:** $* * * * 1 * 1 0 * * 0 1 1 0 * * 0 1 * * 0 0 * 1 * * 0 * * * * 1 * 1 1 * * 1 0 1$
0	wasp sable gull vesper
1	sea dog y giant beetle snail mare turtle canary leopard moose pigeon elk dolphin camel troll herring mosquito mammal mariana guillermo tuna rook antelope zebra grizzly pelican cyclops padre spiny kiwi kingfisher aster lemur tern algarve darter mako kestrel longbow coos vole cocos tricolor occidentalis amundsen humpback sancti
2	<unk> mountain males marine bear bird sub wolf douglas juan dragon ghost leaf volunteers muscle spider tiger warrior lion snake beast egg deer lance powder hawk frog raven fe cow dwarf snails jade scarlet whale kidney skeleton hector torch insect thief herd lizard shin predator banana chestnut worm verde feather hare capsule julio mole samurai stout ramon squirrel stallion mal axel sparrow myrtle alfredo toad viper wasps mace otter horseshoe flanks buff félix talon lynx scorpion ángel bony parrot heron paw cub yun jaya hen gall ryu mangrove gills trident pea fina eels saber robber stryker rocco newt rodent rodrigues surfer hiro bulldog cuckoo sunflower dung condor breeder tortoise salamander stag sturgeon ob whiting toddler sickle álvaro grasshopper elegans terrier puma dragonfly yew alonzo pygmy lichen pinkish martel rufous raptor woolly pixie pups nom rhinoceros kowalski cheetah keita ngāti falco conus amphibian otters stork hummingbird hoof sedge arrowhead niko starling rattlesnake sitter osprey jethro simplex esperanza canis gazelle grouse phalanx j.r. albatross bunting hawkeye foal dill gunter harrier peralta melo thrush reiner pheasant banshee starfish miki trapper swordsman mueang burg tir ez mchugh bumblebee rhododendron petiole josep appleby chino gros cay weser coughlin grasshoppers toa scavenger walrus dorn corky pizarro grandes anemone bowser kenner mig mussel ficus homeworld gopher trapdoor mackerel trak haller rucker aran bobcat pitcairn posey taiga dingo alfalfa bachman iguana reptilian mallard emitter chia harpoon pumas parra aru palos orca catawba bivalve shredder plover nettle sagebrush neely tarantula ketone tycho tetrahedral testes kanda dimer cormorant ponderosa iceman japonica manatee armadillo feliciano gracilis hummingbirds zook maples último tartar hin merino griffon venter mino gantry yeager stanislas niebla adder hock bonin 100-metre gagarin flamingos karman matos putra a-1 lyase proboscis jäger limpets minnow bruiser roadrunners stingrays opossum arbuthnot lioness dimaggio parakeet camellia vermillion atacama 0.39 brahe lannister cormorants celestino riddick

Table 4.12. Incremental list of words that fit the template within *margin* bits of error. By WordNet's definition, the template should represent : *nouns denoting animals.* See appendix D.1 for a list of words used for training.

Margin	**Words fitting the template for *noun.time*:** $* * * 1 0 * * 0 1 * * * 1 * 0 0 * * * 1 1 1 * * * * * * * 1 0 * * 0 * * * * * *$
0	night summer previous morning wake occasions nights autumn easter daytime supper mornings midday afternoons jeopardy millennia lunchtime jan. summertime 8:30 weeknight
1	day september january june october july august april november december century february days 2018 2019 months week minutes hours weeks annual winter spring minute seconds 1960s quarter mile streets sunday centuries bc 1950s session saturday afternoon weekend 1930s ages friday 1920s holiday occasion pm breakfast eve monday dawn 1940s midnight generations p.m. thursday a.m. tuesday millennium hurry may. 1500 1800 sunset inception cafe holidays 1880s noon meantime spa 1890s sunny midst weekends o'clock vista stereo halloween announce 1900s thanksgiving 1850s mid-1990s ios evenings expo weekday 1600 rainy 90s saturdays 1800s 24-hour kickoff homecoming commemoration 1400 infancy 10:00 zee nw nighttime 9:00 1707 reopening 1690 mondays feb uptown e3 50s midsummer o' clock tallahassee thursdays asheville solstice 1450 5:00 wednesdays multiplex dec. ramadan tuesdays batches hrs aug feb. talladega equinox hundredth 2:00 1700s 9:30 3:00 sep gmt 1490 cet pentecost 1:00 sebring fy 1240 aug. whit 11:30 1470 1770s colspan=9| 7–8 longchamp 12:30 4:30 2005–2006 regionals 9–12 19:00 1390 trimester 1346 16:00 delos boldface docs q1 1:30 safekeeping vernal smashwords.com agm 100-year midweek 7–10 1305 10–11 christianization
2	time do years world place see march next 2010 final 10 2014 late 2016 2017 live big 2000 release ca above middle track 2002 tour stage 1999 1998 1996 1997 1995 fall month drive 40 flight seasons corner normal memory 1980s avenue 1970s visit trial sleep evening kiss 1990s bottom concert christmas cd billboard oxford weather weekly decades dinner 1917 cambridge ad consecutive anniversary 1912 1915 2020 300 1916 tomorrow stages chase moments jump instance parking shopping 33 printed mall indicate 42 jail vegas 55 shower celebration interstate microsoft altitude climb venice closure 600 floors meal nashville google marathon sixty 95 sometime assignment 72 hms duration wednesday cent casino runway elimination tickets adoption download vicinity magnitude horizon 68 renovation speedway jam um 78 76 99 slot 88 salem odds embrace f**k 96 sec tunnels 98 ruin alexandra twelfth 87 lunar 92 downstairs se celebrations 86 theaters 93 yearly 91 97 94 lounge itunes 3000 mar eighteenth kylie hiatus café 2018–19 advent midwest teens sidewalk onset gran 1803 audition renewal waking 1808 semester deadline disappointing 1798 receptions sundays entirety sneak drawer 1860s seventeenth centennial 1776 1799 lowe gig cassette inauguration punt charting wembley elevations shelters deluxe rematch 1780 80s gala mid-1980s preseason primetime continents relocation 1783 parole fireworks meridian dunes listings vhs cinemas rancho 1788 2021 lick termination 1840s 70s rehearsal cafeteria 1784 1781 yorker countdown pasadena jurassic olympia adulthood olds bedrooms ascension elle compton redemption slots remembrance mid-1970s düsseldorf stalls weddings weekdays 2010s rallies biennial commencement spoiled oricon b.c. hottest 4000 emancipation mv nap lastly akron a.d. twitch nokia restart 2022 climates penultimate prom pubs halves kilometer westwood pleistocene vip sixties til 1762 showtime 1761 1771 bash rooftop rejoin 10-year 1745 gma argyle kb civilizations shelton tremble 1715 wrestlemania 1740 disneyland kelsey 1910s eviction 1650 stormy equator miocene 1756 iteration furlongs telecast commotion bam fridays introductions redwood mid-1950s wb darkest seventies promenade fifties 1661 concourse warped sync ascot vhf kraft 6000 greensboro astoria littered hillsborough 1748 gainesville 1751 eighties triples visitation nxt 1708 observance newbury bcs stroll previews 1714 1725 fling 1649 1739 carnage 1702 fortnight 1717 symbolize subchannel 20s walmart pax cafes nightclubs ghz devonian skyscrapers grandstand nineties 1638 fillmore 6:00 1703 lps 5–7 1711 clap 1738 1651 1580 musik nov 1655 11:00 1726 1637 bleachers tierra midpoint 1612 h.r. 1687 disappoint uta msnbc 1684 7,500 1601 1676 750,000 epiphany 1669 delphi 7000 12:00 1691 scoreboard mid-1930s 1694 taping 1588 13-year amazon.com glaciation 1631 1617 gunpoint nuestra feasts gestation 1584 reykjavík campfire rafters mid-2000s res 4–5 clausura 1539 matchday hindustan 1593 nov. vacations bein gutter jag 1–4 delft tenths acropolis 1350 clowns birdie 5–6 reprieve shabbat 1567 2007–2008 lows 1535 1066 3500 vail passover 1575 putt hinckley brixton 2024 777 estrella 4:00 sampler beltway jive 7:30 1790s storefront tryout luncheon 1562 5,500 arca 1549 wgn riverdale 10:30 undress goodwood gw omni lockdown lovemaking carolinas 6:30 2100 ccm 1150 eur heatseekers ubuntu 880 5:30 t2 fenway aquarius cuisines aha apr 1548 nome 2008–2009 bmt 6–8 preakness catwalk pompeii fairmont 1340 1519 1780s 1523 firefight 1480 earths caesars 1420 nite 2009–2010 1260 1050 banishment flor emmerdale nürburgring slaughterhouse 8–10 1290 soflá 1485 18:00 1440 pylon birthdays expiry cbbc scrabble dottie reissues 1962–63 7–9 1320 1080 matlock 1505 1180 wsop 2006–2007 daisies nicobar 1415 1280 reunions awry 1600s 1410 barnstaple 1430 1270 1½ thirtieth reba rmb 1380 cst springtime budweiser 1760s intermission columbine melodifestivalen 1204 1498 q&a amira brunch midseason balloting 2300 canonization 1518 televoting 910 chats hoosier hobo hardback biscayne saros 491 throes pinoy amore donington snows 4d commemorations 1296 cuddle 26-year 2026 lindisfarne aftershocks cupcakes 1348 1955–56 woolworths muchmusic jonesboro deseret leadoff 9–10 pst retrial millionth 434 morgantown remand 1349 1160 ktm safeway diss brac instalment 8–9 14–15 1325 weeknights rupee 1482 1070 20:00 1307 atolls pula 100-meter 1308 tvnz diwali meadowlands 35-year 3200 tydfil vacationing kmart pbr presto backstreet wabc slc 2013–2014 landmass 06:00 annihilate litters gorda redshift 997 fairground pa. ocala olden omg 1225 2009/2010 ano 15–16 1140 buzzards perihelion 1422 1381 1487 quickie skytrain espn2 1937–38 skylab 15:00 925 greensburg 12–13 21:00 bushfires 2009/10 1472 1315 1322 linkedin bolero asti klagenfurt 1932–33 2014–2015 1212 woburn 2012/13 pints

Table 4.13. Incremental list of words that fit the template within *margin* bits of error. By WordNet's definition, the template should represent : *nouns denoting time and temporal relations.* See appendix D.3 for a list of words used for training.

Margin	**Words fitting the template for *verb.body*:** ∗ 0 1 0 ∗ ∗ ∗ 0 ∗ 0 ∗ 0 ∗ 1 ∗ 0 ∗ ∗ ∗ ∗ ∗ ∗ ∗ 1 ∗ ∗ 0 ∗ ∗ ∗ ∗ 0 0 1 ∗ 0 1 ∗ 1 ∗
0	scare shrug absorb decipher inhale
1	gaze handle smell grin burn sigh reveal monitor punch hug cure dig heal glare temper gasp freeze wipe spill shiver expose moan consume chuckle rub smirk wink giggle gag wander summon verify conceal cough tease transplant stain scowl chew massage grunt murmur comprehend jolt vomit bandage boil implant undo confuse arenâ t prick grimace emit sniff conceive soak eradicate exhale yawn squeal p**sy beep relish mourn unravel stun elicit mumble tickle impart underestimate yelp blindfold tempt annoy topple lurch clot scoot bellow meds cleanse divulge tantrum untie croak taint coughs confiscate gulping grumble brighten fasten sneeze stirs snore smother
2	’m feel sound smile phone review kill breath watch kiss ’ ve damn worry softly realize calm knife breathing smoke prove memories expect grip shake admit tape load stare mess bite laughter scream whisper tear sweat infection fix spare suffer smiles eyebrows healing solve wash possess bleeding dante facilitate pump embrace forgive beard stole snap idiot swallow thread squeeze frown whispers tattoo spray cries extract instinct spit adjust ache tasted curb groan depict suppress growl clutch orgasm slap headache lump distract jar invade calculate c**k blush interrupt stab transmit rot calf slit tolerate sob sighs scarf toxicity eats choking grins donkey ecstasy twitch shudder hiss clit compliment furiously shrugs knit coughing caress tremble disclose gulped erase chuckles uncover reassure yank antibody chop i- reset scans flinch weave recreate provoke spoil stitch infiltrate ass**le addict workout pu**y moans gulp snort mutter imitate sling infect fling crank rattling shave restrain exert clutches poke symbolize shatter awaken evoke stubble symbolizes growls caffeine clap puffed magically soothe discern twinkle clench wrench smear pry whimper pant liberate dictate ascertain contemplate delete appease invoke illuminate overload lessen slumber handshake graze wield sneer overwhelm juices disable wail expertly nibbled retort transfusion wailed throb deftly nudge apologise taunt wiggle avert perish flailing cellphone revulsion arse pout injure smirks undress slurred abort fetish replenish despise savor memorize dismantle dimple simmering dissipate rectify conjure edema bustle refill reap stupor spasm stiffen bothers deduce fingernail snag ailment squint banish rinse coax sew dispel humiliate conditioner oblige drool faucet hurl methodically befriend lotion beeped confide compress cramp cl*t blurt secrete implantation churn eject exhaling hasnâ t recycle ointment leer concoction wheezing ulcer crutch lightening snot â youâ ve startle skim abduct placate crushes err germinate cuddle heaviness contraceptives kneading earpiece irritate aundy negate withdraws appreciates know?â fart whoop holler twirl finalize puss supplementation unwind shudders broach squeals scold sharpen dampen reciprocate

Table 4.14. Incremental list of words that fit the template within *margin* bits of error. By WordNet’s definition, the template should represent : *verbs of grooming, dressing and bodily care*. See appendix D.4 for a list of words used for training.

Margin	Words fitting the template for ***verb.consumption***: $* * * 0 * 0 * * * 0 * * * * * * * * * 1 * 0 * 1 * 0 * * * 1 * 0 0 1 1 * 1 * 1 *$
0	afford leap scare consume rub chew conceive scoot compress
1	get take find bring travel carry drop send grow hop smell expect blow load earn hang knock collect permit solve wash dig sip jerk freeze spill crawl borrow grill fetch summon piss attach disclose accumulate chop grit cultivate spoil boil fend exert diminish souvenir lessen carve graze starve overwhelm mourn unravel nudge degrade cede pluck assimilate reload reap mull topple rinse coax recycle germinate brighten fasten enslave
2	have go want care food provide security trade hold try train pass reach pay trust walk vote horse watch share wind safety remain risk traffic save boat protect receive ’ ve coffee trail shop pull express drink bear gain check buy sell maintain reduce poverty spend feed choose throw hide cook hunt waste smoke strip hat kick lift defend stick burn bet shower retreat mess swing deliver pitcher monitor blame bike slide encourage rope retain introduce recover navigation toll acquire rent generate lease steal prefer resist alert tide swim flee feast concentrate possess reward pump f**k refuse ham buck divide wage harvest wrap crane tap submit cart sandwich recruit warn heal click flip boost drift pray grocery tire declare arrange canoe haul adjust dial blockade herd float wipe famine skate expose tub educate slice absorb weigh behave translate glove punish render shove dye bump foam bait evolve tuck pinch splash transmit rot wander donate pasture shortages feeder restrict trolley snack paddle donkey raft drown whisky inherit lifeboat bleed stimulate piracy peanut massage steak batter booster cheat seduce adhere glide timbers inhibit shrink smack flex soaking grind jolt yank disperse divert brew shovel deter await vomit dispose mitigate commute hog cider trickle gulp gully rein drip infect restrain wildcat buffet fuss awaken coldest upkeep bake curfew dedicate sniff inflict discern forfeit weep relinquish booze hover emigrate liberate cling frighten jog upload soak kneel illuminate buggy welcomes deteriorate postpone unload beep disable fleece deflection partake kayak motivate inject grate stow mobilize nurture ration stun duffel barricade rationing hump graces subscribe apologise erode analyse roost elicit regenerate crave tickle vibrate wiggle sow trespassing visualize congratulate pout enrich recharge preside ebb aspire abort underestimate buoy smuggle blindfold despise dismantle displace dissipate conjure bustle heave refill lunge vagrant tern whirl sauna amplify radiate lurch overpower colonize humiliate conditioner stagger oblige browse drool faucet raiser cleanse adorn dredge bugger divulge pellet secrete purify jeopardize snowmobile imprison gape congregate refuel temp stopover untie cob skim cleave taint err backcountry cuddle whiz confiscate corkscrew perk know?â twirl seabird babble brag swish unwind amass catamaran disengage payday smother flatten

Table 4.15. Incremental list of words that fit the template within *margin* bits of error. By WordNet’s definition, the template should represent : *verbs of eating and drinking.* See appendix D.5 for a list of words used for training.

Conclusion

In this work, we propose a novel method for training binary word embeddings inpired by Search Data Structure Learning. Instead of solving a compression problem using pre-trained continuous vectors, we implement a Multi-Bernouilli Regression Search to solve a matching problem between two domains : middle words and surrounding context for given windows of words. Our method has the advantage of learning the discrete representations directly, meaning the model can be trained end-to-end.

Our experiments strongly suggest that the semantical information of each word relative to itself and its surrounding context is encoded in an almost humanly interpretable way. It appears as if our initial idea of sememe, i.e., latent discrete units of meaning, applies to our resulting representations, although not as well-defined as expected. Certain combinations of bits can be associated to known grammatical properties such as different morphological inflections which in turn can be interpreted as contextual semantic markers. Known concepts can also be attributed to manually built or learnt bit templates. Furthermore, our method exhibits promising generalization capabilities by generating binary vectors for unknown words that can integrate seamlessly with the base vocabulary embeddings. Even though continuous GloVe embeddings obtained better similarity scores and it could be argued that their word clusters are more semantically related, our work requires up to 200 times less memory and allows for efficient vector operations to be used such as a XOR operation to compute the hamming distance. The low dimensionality and the discrete nature of our resulting embeddings also display better interpretability that opens up their use in linguistics research and not only as input to another downstream task.

Our extensive qualitative evaluation suggests that our method has a good potential for improvement. Many of our experimentation settings were chosen arbitrarily and much more configurations such as vector dimensionality and model capacity can be tried. Further exploration of these configurations and other potential applications are left for future work.

References

[1] Dzmitry BAHDANAU, Kyunghyun CHO et Yoshua BENGIO : Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2015.

[2] Yoshua BENGIO, Réjean DUCHARME, Pascal VINCENT et Christian JAUVIN : A neural probabilistic language model. *JOURNAL OF MACHINE LEARNING RESEARCH*, 3:1137–1155, 2003.

[3] Yoshua BENGIO, N. LÉONARD et Aaron C. COURVILLE : Estimating or propagating gradients through stochastic neurons for conditional computation. *ArXiv*, abs/1308.3432, 2013.

[4] Elia BRUNI, Nam-Khanh TRAN et M. BARONI : Multimodal distributional semantics. *J. Artif. Intell. Res.*, 49:1–47, 2014.

[5] Yo Joong CHOE, Jiyeon HAM, Kyubyong PARK et Yeoil YOON : A neural grammatical error correction system built on better pre-training and sequential transfer learning. *In Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 213–227, Florence, Italy, août 2019. Association for Computational Linguistics.

[6] Seung-Seok CHOI, S. CHA et C. TAPPERT : A survey of binary similarity and distance measures. *Journal on Systemics, Cybernetics and Informatics*, 8:43–48, 2010.

[7] Shamil CHOLLAMPATT et Hwee Tou NG : A multilayer convolutional encoder-decoder neural network for grammatical error correction. *In AAAI*, 2018.

[8] Shamil CHOLLAMPATT, Weiqi WANG et Hwee Tou NG : Cross-sentence grammatical error correction. *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 435–445, Florence, Italy, juillet 2019. Association for Computational Linguistics.

[9] J. DEVLIN, Ming-Wei CHANG, Kenton LEE et Kristina TOUTANOVA : Bert: Pre-training of deep bidirectional transformers for language understanding. *In NAACL-HLT*, 2019.

[10] Mathieu DUCHESNEAU, Alain TAPP et Hansenclever BASSANI : Search data structure learning. 2021.

[11] Sergey EDUNOV, Myle OTT, Michael AULI et David GRANGIER : Understanding back-translation at scale. *ArXiv*, abs/1808.09381, 2018.

[12] Manaal FARUQUI et Chris DYER : Non-distributional word vector representations. *In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 464–469, Beijing, China, juillet 2015. Association for Computational Linguistics.

[13] J. R. FIRTH : A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*, 1957.

[14] Tao GE, Furu WEI et Ming ZHOU : Reaching human-level performance in automatic grammatical error correction: An empirical study. *ArXiv*, abs/1807.01270, 2018.

[15] Daniela GERZ, Ivan VULIC, Felix HILL, Roi REICHART et A. KORHONEN : Simverb-3500: A large-scale evaluation set of verb similarity. *In EMNLP*, 2016.

[16] Roman GRUNDKIEWICZ, Marcin JUNCZYS-DOWMUNT et Kenneth HEAFIELD : Neural grammatical error correction systems with unsupervised pre-training on synthetic data. *In Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 252–263, Florence, Italy, août 2019. Association for Computational Linguistics.

[17] Matthew HONNIBAL et Ines MONTANI : spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. `https://spacy.io/`, 2017.

[18] J. HUANG : Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2:283–304, 2004.

[19] Diederik P. KINGMA et Jimmy BA : Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.

[20] Shun KIYONO, Jun SUZUKI, Masato MITA, Tomoya MIZUMOTO et Kentaro INUI : An empirical study of incorporating pseudo data into grammatical error correction. *In EMNLP/IJCNLP*, 2019.

[21] Taku KUDO et John RICHARDSON : SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium, novembre 2018. Association for Computational Linguistics.

[22] Mike LEWIS, Yinhan LIU, Naman GOYAL, Marjan GHAZVININEJAD, Abdelrahman MOHAMED, Omer LEVY, Ves STOYANOV et Luke ZETTLEMOYER : Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *ArXiv*, abs/1910.13461, 2020.

[23] Jared LICHTARGE, C. ALBERTI, Shankar KUMAR, Noam SHAZEER, Niki PARMAR et S. TONG : Corpora generation for grammatical error correction. *ArXiv*, abs/1904.05780, 2019.

[24] Zhouhan LIN, Minwei FENG, C. D. SANTOS, Mo YU, B. XIANG, Bowen ZHOU et Yoshua BENGIO : A structured self-attentive sentence embedding. *ArXiv*, abs/1703.03130, 2017.

[25] Chris J. MADDISON, A. MNIH et Y. TEH : The concrete distribution: A continuous relaxation of discrete random variables. *ArXiv*, abs/1611.00712, 2017.

[26] Stephen MERITY : Single headed attention rnn: Stop thinking with your head. *ArXiv*, abs/1911.11423, 2019.

[27] Tomas MIKOLOV, Kai CHEN, G. S. CORRADO et J. DEAN : Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.

[28] Tomas MIKOLOV, Ilya SUTSKEVER, Kai CHEN, Greg S CORRADO et Jeff DEAN : Distributed representations of words and phrases and their compositionality. *In* C. J. C. BURGES, L. BOTTOU, M. WELLING, Z. GHAHRAMANI et K. Q. WEINBERGER, éditeurs : *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013.

[29] G. MILLER : Wordnet: a lexical database for english. *Commun. ACM*, 38:39–41, 1995.

[30] A. OORD, Oriol VINYALS et K. KAVUKCUOGLU : Neural discrete representation learning. *In NIPS*, 2017.

[31] Jeffrey PENNINGTON, Richard SOCHER et Christopher D. MANNING : Glove: Global vectors for word representation. *In Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.

[32] Matthew E. PETERS, Mark NEUMANN, Mohit IYYER, Matt GARDNER, Christopher CLARK, Kenton LEE et Luke ZETTLEMOYER : Deep contextualized word representations. *ArXiv*, abs/1802.05365, 2018.

[33] A. RADFORD : Improving language understanding by generative pre-training. 2018.

[34] A. RADFORD, Jeffrey WU, R. CHILD, David LUAN, Dario AMODEI et Ilya SUTSKEVER : Language models are unsupervised multitask learners. 2019.

[35] Rico Sennrich, Barry Haddow et Alexandra Birch : Neural machine translation of rare words with subword units. *In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany, août 2016. Association for Computational Linguistics.

[36] Rico Sennrich, Barry Haddow et Alexandra Birch : Neural machine translation of rare words with subword units. *ArXiv*, abs/1508.07909, 2016.

[37] Dinghan Shen, Pengyu Cheng, Dhanasekar Sundararaman, X. Zhang, Qian Yang, M. Tang, A. Çelikyilmaz et L. Carin : Learning compressed sentence representations for on-device text processing. *In ACL*, 2019.

[38] Raphael Shu et Hideki Nakayama : Compressing word embeddings via deep compositional code learning. *ArXiv*, abs/1711.01068, 2018.

[39] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu et Tie-Yan Liu : Mass: Masked sequence to sequence pre-training for language generation. *In ICML*, 2019.

[40] Julien Tissier, Amaury Habrard et C. Gravier : Near-lossless binarization of word embeddings. *In AAAI*, 2019.

[41] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, L. Kaiser et Illia Polosukhin : Attention is all you need. *ArXiv*, abs/1706.03762, 2017.

[42] Y. Wu, Mike Schuster, Z. Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, M. Krikun, Yuan Cao, Q. Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, Taku Kudo, H. Kazawa, K. Stevens, G. Kurian, Nishant Patil, W. Wang, C. Young, J. Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, G. S. Corrado, Macduff Hughes et J. Dean : Google's neural machine translation system: Bridging the gap between human and machine translation. *ArXiv*, abs/1609.08144, 2016.

[43] Ziang Xie, Guillaume Genthial, S. Xie, A. Ng et Dan Jurafsky : Noising and denoising natural language: Diverse backtranslation for grammar correction. *In NAACL-HLT*, 2018.

[44] Y. Zhu, Ryan Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba et S. Fidler : Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 19–27, 2015.

Appendix A

Word Lists

A.1. Semi-Automatic

These lists were generated from our 50,000 words training vocabulary. To be more accurate and to take into account irregular inflections (e.g. *eat* and *ate*), we processed about 8 million sentences with spaCy [**17**] to count POS tags for every word (see app. C.1). We used the whole corpus (i.e. not only the train split), which gave us statistics for 43,240 words out of 50,000. We only considered POS tags if their count represented at least 5% of the occurences for the associated word in order to filter out false positives. From this, we looked for every word with inflected forms (i.e. plurals, past, third person and present participle) and extracted their lemma. If both words were in our vocabulary, we added the pair to the appropriate word list. These lists were not validated completely. A brief examination showed they were sufficient and contained very few errors. Hence, it would have no significant impact on our results.

A.1.1. Plurals

Singular form with associated plural modifications
year(s) state(s) eye(s) m**an**(en) wom**an**(en) child(ren) member(s) game(s) time(s) day(s) student(s) hand(s) other(s) work(s)
service(s) thing(s) month(s) area(s) record(s) force(s) team(s) point(s) event(s) song(s) school(s) arm(s) word(s) player(s)
medi**um**(a) f**oot**(eet) right(s) minute(s) book(s) art(s) famil**y**(ies) group(s) stud**y**(ies) hour(s) award(s) friend(s) countr**y**(ies)
sport(s) show(s) film(s) part(s) dat**um**(a) system(s) building(s) parent(s) place(s) championship(s) operation(s) run(s) unit(s)
week(s) play(s) line(s) compan**y**(ies) artist(s) girl(s) issue(s) term(s) star(s) li**fe**(ves) lip(s) name(s) character(s) change(s)
remain(s) activit**y**(ies) match(es) program(s) result(s) goal(s) mile(s) island(s) plan(s) brother(s) house(s) problem(s) track(s)
case(s) stor**y**(ies) finger(s) individual(s) form(s) troop(s) boy(s) cit**y**(ies) season(s) station(s) soldier(s) use(s) single(s)
worker(s) ship(s) election(s) number(s) nation(s) lie(s) product(s) officer(s) male(s) condition(s) level(s) leave(s) element(s)
resident(s) leg(s) tree(s) car(s) rule(s) household(s) appearance(s) project(s) race(s) wall(s) effect(s) look(s) effort(s) side(s)
affair(s) leader(s) club(s) mountain(s) yard(s) source(s) facilit**y**(ies) type(s) second(s) offer(s) language(s) review(s) note(s)
son(s) communit**y**(ies) way(s) part**y**(ies) plant(s) critic(s) episode(s) role(s) relation(s) champion(s) one(s) animal(s) sale(s)
report(s) model(s) seat(s) letter(s) resource(s) class(es) final(s) organization(s) question(s) law(s) field(s) piece(s) title(s)

Singular form with associated plural modifications
position(s) representative(s) claim(s) figure(s) vote(s) site(s) attack(s) power(s) official(s) weapon(s) science(s) district(s)
performance(s) cell(s) train(s) bod**y**(ies) authorit**y**(ies) region(s) thank(s) fan(s) material(s) cover(s) citizen(s) reason(s)
gun(s) order(s) step(s) action(s) street(s) village(s) version(s) set(s) call(s) act(s) human(s) sound(s) attempt(s) turn(s)
tear(s) end(s) article(s) t**ooth**(eeth) vehicle(s) centur**y**(ies) cop**y**(ies) start(s) method(s) town(s) shoulder(s) door(s) bird(s)
metre(s) office(s) kid(s) institution(s) writer(s) picture(s) standard(s) idea(s) head(s) propert**y**(ies) studio(s) road(s)
example(s) detail(s) thought(s) room(s) band(s) fund(s) lyric(s) communication(s) skill(s) structure(s) decade(s) scene(s)
location(s) patient(s) flower(s) function(s) home(s) image(s) church(es) wing(s) pass(es) employee(s) teacher(s) win(s) guy(s)
candidate(s) bank(s) light(s) date(s) inhabitant(s) painting(s) item(s) engine(s) user(s) ground(s) thousand(s) paper(s) lead(s)
application(s) store(s) object(s) cause(s) value(s) good(s) sister(s) course(s) section(s) horse(s) cost(s) guard(s) interest(s)
musician(s) river(s) universit**y**(ies) tr**y**(ies) winner(s) court(s) director(s) technique(s) war(s) owner(s) charge(s) vocal(s)
degree(s) passenger(s) design(s) production(s) business(es) water(s) person(s) count**y**(ies) document(s) scholar(s) return(s)
card(s) operate(s) route(s) publication(s) dut**y**(ies) daughter(s) hundred(s) purpose(s) age(s) stage(s) branch(es) division(s)
actor(s) king(s) visitor(s) factor(s) subject(s) measure(s) recording(s) victim(s) blue(s) wood(s) concern(s) contribution(s)
moment(s) feeling(s) prisoner(s) partner(s) her(s) difference(s) device(s) face(s) movement(s) stair(s) author(s) movie(s)
loss(es) matter(s) aspect(s) meter(s) rock(s) novel(s) test(s) instrument(s) polic**y**(ies) adult(s) practice(s) rate(s) sign(s)
account(s) knee(s) drug(s) forest(s) market(s) industr**y**(ies) couple(s) host(s) job(s) requirement(s) suppl**y**(ies) boat(s)
network(s) competition(s) tool(s) round(s) move(s) agent(s) newspaper(s) component(s) meeting(s) park(s) master(s)
medal(s) experience(s) collection(s) root(s) vessel(s) process(es) territor**y**(ies) center(s) mission(s) list(s) bear(s) customer(s)
injur**y**(ies) color(s) producer(s) dream(s) lake(s) credit(s) circumstance(s) poem(s) farmer(s) ear(s) dollar(s) wave(s)
supporter(s) scientist(s) border(s) share(s) occasion(s) relationship(s) dog(s) page(s) guest(s) engineer(s) style(s) video(s)
exist(s) session(s) mill(s) concert(s) death(s) researcher(s) principle(s) statistic(s) opportunit**y**(ies) participant(s) college(s)
population(s) government(s) saint(s) stone(s) benefit(s) limit(s) agenc**y**(ies) theme(s) rank(s) province(s) all**y**(ies) channel(s)
period(s) decision(s) judge(s) tribe(s) support(s) voice(s) reach(es) athlete(s) railway(s) god(s) shoe(s) driver(s) warrior(s)
release(s) viewer(s) drum(s) trial(s) machine(s) voter(s) platform(s) categor**y**(ies) asset(s) regulation(s) lad**y**(ies) bone(s)
volume(s) link(s) technolog**y**(ies) crime(s) mine(s) settler(s) opponent(s) night(s) pattern(s) slave(s) stop(s) key(s) people(s)
inch(es) enem**y**(ies) immigrant(s) photograph(s) volunteer(s) tie(s) option(s) magazine(s) ranger(s) cheek(s) fighter(s)
festival(s) connection(s) space(s) tradition(s) playoff(s) acre(s) tank(s) municipalit**y**(ies) boundar**y**(ies) origin(s) bar(s)
bridge(s) settlement(s) shop(s) increase(s) coach(es) temperature(s) edition(s) wind(s) price(s) map(s) negotiation(s) battle(s)
minister(s) file(s) tour(s) flight(s) reference(s) grade(s) protest(s) giant(s) lion(s) pupil(s) emotion(s) doctor(s) shot(s)
locomotive(s) speaker(s) demand(s) shadow(s) score(s) topic(s) spot(s) department(s) bas**is**(es) stake(s) muscle(s) talk(s)
interview(s) pant(s) solution(s) coin(s) airline(s) ton(s) height(s) camp(s) column(s) campaign(s) pair(s) visit(s) tax**is**(es)
crosse(s) influence(s) bus(es) symptom(s) restaurant(s) disease(s) bill(s) tower(s) lot(s) challenge(s) leed(s) control(s)
protein(s) target(s) present(s) computer(s) table(s) league(s) angel(s) abilit**y**(ies) comment(s) board(s) experiment(s) enter(s)
tournament(s) difficult**y**(ies) boot(s) quarter(s) answer(s) cut(s) seed(s) travel(s) writing(s) signal(s) walk(s) plate(s) ball(s)
client(s) eagle(s) architect(s) approach(es) descendant(s) hero(es) colon**y**(ies) message(s) theor**y**(ies) veteran(s) hospital(s)
reader(s) improvement(s) contract(s) range(s) assist(s) photo(s) wheel(s) pilot(s) amount(s) rebel(s) generation(s)
adventure(s) kilometer(s) rider(s) reserve(s) exhibition(s) hope(s) pound(s) creature(s) vampire(s) lord(s) post(s) situation(s)
runner(s) colleague(s) floor(s) societ**y**(ies) archive(s) organisation(s) reform(s) ring(s) expert(s) refugee(s) connect(s) rating(s)
direction(s) farm(s) edge(s) circle(s) politician(s) love(s) estate(s) professional(s) programme(s) manufacturer(s) culture(s)
casualt**y**(ies) singer(s) arrangement(s) graphic(s) founder(s) lesson(s) journalist(s) instruction(s) flame(s) kind(s) aim(s)
portion(s) temple(s) sample(s) crop(s) variation(s) concept(s) survivor(s) grant(s) composition(s) plain(s) statement(s) hole(s)
follower(s) association(s) port(s) display(s) centre(s) pirate(s) threat(s) priest(s) broadcast(s) victor**y**(ies) task(s) ruin(s)
arm**y**(ies) development(s) touchdown(s) appeal(s) twin(s) chamber(s) relative(s) manage(s) consequence(s) venue(s) box(es)
gene(s) bishop(s) chapter(s) segment(s) union(s) glass(es) librar**y**(ies) content(s) discover(s) camera(s) audience(s) rival(s)
provision(s) stream(s) nomination(s) tourist(s) criteri**on**(a) restriction(s) conference(s) journal(s) million(s) layer(s) suburb(s)
lecture(s) committee(s) heart(s) sibling(s) deal(s) cardinal(s) fire(s) operator(s) investor(s) hotel(s) corner(s) cloud(s)
procedure(s) panel(s) secret(s) finding(s) associate(s) alumn**us**(i) particle(s) chair(s) insect(s) brand(s) allegation(s) colour(s)
servant(s) count(s) read(s) dragon(s) trustee(s) specimen(s) spirit(s) drawing(s) witness(es) holding(s) performer(s) heel(s)
conflict(s) symbol(s) aire(s) activist(s) world(s) food(s) observation(s) demon(s) wish(es) wol**f**(ves) wound(s) strike(s) string(s)
lawyer(s) workshop(s) cat(s) truck(s) mention(s) museum(s) reaction(s) developer(s) competitor(s) drive(s) publisher(s)

Singular form with associated plural modifications
sheet(s) senator(s) bag(s) portrait(s) rise(s) scale(s) fact(s) estimate(s) deposit(s) smile(s) zone(s) firm(s) bond(s) rover(s) code(s) composer(s) variant(s) motor(s) investigation(s) ticket(s) drink(s) achievement(s) eyebrow(s) fee(s) trail(s) chain(s) environment(s) variet**y**(ies) kilometre(s) fight(s) delegate(s) wicket(s) dav**y**(ies) surface(s) sequence(s) compound(s) highway(s) poet(s) tactic(s) grave(s) repair(s) proposal(s) bed(s) incident(s) argument(s) murder(s) monument(s) capabilit**y**(ies) proceeding(s) plane(s) mechanic(s) enterprise(s) interaction(s) dimension(s) discussion(s) monk(s) legend(s) slope(s) rebound(s) cap(s) kiss(es) short(s) entr**y**(ies) merchant(s) habitat(s) foundation(s) battalion(s) responsibilit**y**(ies) strateg**y**(ies) bit(s) ethic(s) ray(s) regiment(s) premise(s) chance(s) essay(s) gift(s) jet(s) apartment(s) ranking(s) bomb(s) fl**y**(ies) manager(s) local(s) consumer(s) monster(s) being(s) roll(s) batter**y**(ies) choice(s) agreement(s) lung(s) payment(s) national(s) exhibit(s) investment(s) trace(s) bomber(s) crew(s) mind(s) mechanism(s) governor(s) artifact(s) fragment(s) council(s) editor(s) contact(s) sculpture(s) molecule(s) raider(s) black(s) fear(s) label(s) ruler(s) complaint(s) soul(s) dancer(s) ceremon**y**(ies) contestant(s) graduate(s) disorder(s) raid(s) factor**y**(ies) cup(s) president(s) loan(s) row(s) sense(s) initiative(s) opinion(s) career(s) error(s) wi**fe**(ves) summer(s) dish(es) odd(s) ancestor(s) path(s) miner(s) genre(s) hunter(s) lane(s) fruit(s) sector(s) beat(s) vein(s) size(s) donation(s) setting(s) mother(s) expectation(s) pole(s) bat(s) curve(s) exercise(s) response(s) palm(s) shell(s) cave(s) tip(s) designer(s) tube(s) france(s) entit**y**(ies) speed(s) parish(es) request(s) lap(s) sailor(s) missile(s) objective(s) classic(s) frame(s) risk(s) recommendation(s) commander(s) maker(s) variable(s) advance(s) mammal(s) neighbor(s) missionar**y**(ies) remnant(s) patriot(s) profit(s) promise(s) fossil(s) squadron(s) bell(s) organism(s) lover(s) deput**y**(ies) carrier(s) prize(s) encounter(s) watt(s) saving(s) up(s) dynamic(s) toe(s) emission(s) back(s) conservative(s) statue(s) website(s) guideline(s) command(s) struggle(s) cowboy(s) reside(s) dispute(s) print(s) derive(s) coordinate(s) general(s) corporation(s) format(s) mode(s) bullet(s) blade(s) highland(s) highlight(s) exception(s) provider(s) intention(s) toy(s) galler**y**(ies) measurement(s) uniform(s) hectare(s) illustration(s) celebration(s) expense(s) qualit**y**(ies) theater(s) address(es) nurse(s) equation(s) villager(s) instance(s) pocket(s) sentence(s) similarit**y**(ies) scheme(s) dance(s) masse(s) protester(s) guitar(s) rumor(s) depth(s) planet(s) nail(s) vegetable(s) bab**y**(ies) defender(s) commissioner(s) suit(s) phase(s) third(s) translation(s) father(s) beach(es) spectator(s) celebrit**y**(ies) pick(s) metal(s) religion(s) surrounding(s) ashe(s) nerve(s) rose(s) distance(s) gain(s) lock(s) bottle(s) tomato(es) parameter(s) penalt**y**(ies) valley(s) expression(s) fist(s) holiday(s) talent(s) phone(s) winter(s) flag(s) damage(s) reporter(s) laborator**y**(ies) painter(s) survey(s) stamp(s) teaching(s) humanit**y**(ies) kingdom(s) revenue(s) finish(es) sander(s) rat(s) shirt(s) gentlem**an**(en) down(s) scot(s) tension(s) spell(s) regard(s) modification(s) sea(s) complication(s) conversation(s) quantit**y**(ies) differ(s) discipline(s) inscription(s) prince(s) employer(s) sword(s) save(s) convention(s) securit**y**(ies) predator(s) tonne(s) attribute(s) executive(s) communist(s) riot(s) landscape(s) native(s) feather(s) chemical(s) advocate(s) telecommunication(s) screen(s) limitation(s) helicopter(s) wage(s) red(s) storm(s) white(s) criminal(s) acid(s) button(s) diamond(s) commercial(s) bull(s) panther(s) professor(s) personalit**y**(ies) councillor(s) shield(s) transaction(s) teammate(s) ride(s) classroom(s) angle(s) theatre(s) trait(s) rocket(s) snail(s) costume(s) attraction(s) possibilit**y**(ies) advantage(s) preparation(s) cop(s) chronicle(s) trader(s) formation(s) commission(s) ram(s) ingredient(s) peer(s) quarterfinal(s) remark(s) domain(s) cluster(s) airport(s) addition(s) inmate(s) liberal(s) mineral(s) chip(s) escape(s) soil(s) peasant(s) stay(s) meal(s) sponsor(s) companion(s) venture(s) mistake(s) guide(s) violation(s) accusation(s) pool(s) sketch(es) crowd(s) frequenc**y**(ies) township(s) demonstration(s) marriage(s) carm**an**(en) outcome(s) pioneer(s) cycle(s) shift(s) teenager(s) suspect(s) watch(es) destroyer(s) trend(s) cousin(s) ghost(s) treatment(s) trick(s) sort(s) description(s) atom(s) strip(s) survive(s) holder(s) combine(s) accident(s) pipe(s) defense(s) dodger(s) affect(s) excavation(s) receptor(s) bay(s) arch(es) prayer(s) campus(es) grandchild(ren) possession(s) alien(s) ritual(s) shore(s) faction(s) dolphin(s) wrestler(s) moth(s) glove(s) rapid(s) destination(s) poll(s) concentration(s) patron(s) listener(s) success(es) observer(s) export(s) viking(s) partnership(s) devil(s) off(s) compete(s) snake(s) midland(s) whisper(s) keyboard(s) firearm(s) breast(s) tile(s) residence(s) debt(s) tone(s) barrier(s) standing(s) speech(es) meadow(s) disabilit**y**(ies) arrow(s) weekend(s) fisher**y**(ies) attitude(s) scream(s) minorit**y**(ies) teen(s) trouble(s) torre(s) installation(s) circuit(s) reinforcement(s) cliff(s) delay(s) stick(s) opera(s) marking(s) vision(s) fellow(s) exam(s) mutation(s) fortification(s) patent(s) cartoon(s) beam(s) pillar(s) tissue(s) tune(s) counterpart(s) debate(s) substance(s) tenant(s) localit**y**(ies) advertisement(s) spark(s) stranger(s) joke(s) impact(s) feed(s) interval(s) transfer(s) wanderer(s) brave(s) bug(s) flood(s) implication(s) introduce(s) seal(s) span(s) bee(s) finalist(s) elder(s) patch(es) brigade(s) passage(s) appointment(s) investigator(s) trade(s) plantation(s) touch(es) decoration(s) isle(s) craft(s) flat(s) collector(s) sensor(s) defeat(s) sand(s) stretch(es) robot(s) privilege(s) explore(s) parson(s) permit(s) practitioner(s) doubt(s) brake(s) script(s) satellite(s) folk(s) brow(s) sack(s) expedition(s) tape(s) plot(s) spider(s) bean(s) colt(s) correspond(s) extension(s) habit(s) brick(s) assignment(s) pit(s) top(s) tributar**y**(ies) reviewer(s) tire(s) examination(s) recruit(s) obligation(s) stroke(s) strikeout(s) constraint(s) beetle(s) crystal(s) falcon(s) air(s) notice(s)

Singular form with associated plural modifications
wrist(s) wagon(s) node(s) renovation(s) patrol(s) check(s) mirror(s) genetic(s) infection(s) wine(s) offering(s) promotion(s)
cr**y**(ies) capital(s) electron(s) shade(s) recipient(s) fingertip(s) cd(s) spear(s) inform(s) duck(s) background(s) pressure(s)
representation(s) complex(es) fort(s) poster(s) shareholder(s) witch(es) repl**y**(ies) reception(s) kni**fe**(ves) bowl(s) sunday(s)
package(s) explosive(s) colonist(s) charit**y**(ies) square(s) steal(s) dock(s) ad(s) commentator(s) cigarette(s) rail(s) neuron(s)
qualification(s) definition(s) commune(s) documentar**y**(ies) ideal(s) current(s) major(s) monaster**y**(ies) contest(s) migrant(s)
combination(s) potato(es) failure(s) pig(s) behavior(s) cadet(s) blow(s) innovation(s) histor**y**(ies) doll(s) predecessor(s)
obstacle(s) cable(s) buck(s) traveler(s) filmmaker(s) foreigner(s) dress(es) castle(s) scholarship(s) packer(s) amendment(s)
administrator(s) interpretation(s) wire(s) identit**y**(ies) captain(s) conclusion(s) enzyme(s) specification(s) ion(s) met(s)
breath(s) disc(s) petal(s) license(s) railroad(s) prison(s) treat**y**(ies) sanction(s) roof(s) discover**y**(ies) specialist(s) clip(s)
exchange(s) gas(es) server(s) defendant(s) ballot(s) clan(s) royal(s) candle(s) desire(s) terrorist(s) killing(s) yield(s) stock(s)
barrel(s) strand(s) push(es) warship(s) contemporar**y**(ies) hurt(s) phrase(s) wildcat(s) rope(s) mate(s) suggestion(s) finance(s)
ridge(s) breed(s) defence(s) originate(s) zombie(s) fixture(s) fang(s) heir(s) attorney(s) resort(s) wetland(s) butterfl**y**(ies)
scar(s) reading(s) rain(s) manner(s) opening(s) glance(s) confront(s) analys**is**(es) retailer(s) alteration(s) nest(s) subscriber(s)
linguistic(s) brain(s) dot(s) knot(s) landing(s) pack(s) adaptation(s) fisherm**an**(en) knuckle(s) denote(s) dam(s) express(es)
congregation(s) signature(s) hurricane(s) alternative(s) islander(s) propose(s) ant(s) deit**y**(ies) margin(s) ministr**y**(ies) over(s)
tomb(s) pond(s) contributor(s) earning(s) elbow(s) close(s) builder(s) dwelling(s) stable(s) pier(s) contractor(s) monitor(s)
arrest(s) melod**y**(ies) interception(s) warning(s) clash(es) bush(es) senior(s) mask(s) weight(s) creator(s) establishment(s)
jurisdiction(s) villain(s) denomination(s) vowel(s) container(s) pump(s) infant(s) calculation(s) intellectual(s) pathway(s)
relic(s) evening(s) flyer(s) motion(s) grape(s) elevation(s) exit(s) import(s) shelter(s) fiber(s) deed(s) prospect(s) build(s)
avenger(s) clue(s) switch(es) coat(s) meaning(s) businessm**an**(en) gap(s) sight(s) trouser(s) canal(s) stare(s) nationalist(s)
duke(s) utilit**y**(ies) curtain(s) dealer(s) selection(s) hostilit**y**(ies) objection(s) generator(s) dart(s) earthquake(s) care(s)
crack(s) surgeon(s) worr**y**(ies) participate(s) net(s) accessor**y**(ies) loop(s) junior(s) derivative(s) 80(s) beginning(s) worm(s)
slide(s) turtle(s) pill(s) scenario(s) fold(s) quote(s) elephant(s) donor(s) vendor(s) applicant(s) nominee(s) militant(s) buyer(s)
accomplishment(s) blanket(s) cent(s) norm(s) commitment(s) fit(s) sleeve(s) entrepreneur(s) noise(s) assistant(s) suffer(s)
photographer(s) cannon(s) affiliate(s) broadcaster(s) protocol(s) thie**f**(ves) myster**y**(ies) mouth(s) purchase(s) continent(s)
decrease(s) seminar(s) philosopher(s) smell(s) whale(s) search(es) press(es) valve(s) firework(s) rest(s) consideration(s) trap(s)
danger(s) perspective(s) citation(s) assembl**y**(ies) lyon(s) shel**f**(ves) organizer(s) portray(s) carriage(s) qualifier(s) entrance(s)
killer(s) respect(s) comparison(s) hurdle(s) beast(s) drain(s) successor(s) jewel(s) invite(s) dune(s) bite(s) listing(s) virus(es)
sin(s) analyst(s) jacket(s) wonder(s) fortune(s) ocean(s) invader(s) motive(s) convince(s) reservation(s) pilgrim(s) deliver(s)
assumption(s) terminal(s) elector(s) pin(s) hat(s) attacker(s) thread(s) shout(s) tablet(s) birth(s) certificate(s) wizard(s)
archaeologist(s) monkey(s) charger(s) cow(s) parallel(s) upgrade(s) flash(es) grain(s) mariner(s) taste(s) statute(s) morning(s)
kit(s) prosecutor(s) cottage(s) joint(s) penguin(s) hearing(s) explanation(s) textile(s) presentation(s) sk**y**(ies) fin(s)
libert**y**(ies) preserve(s) disappear(s) defect(s) sp**y**(ies) resolution(s) goat(s) offence(s) gang(s) icon(s) strain(s) vine(s)
profile(s) processor(s) mural(s) myth(s) centimeter(s) belt(s) shrub(s) textbook(s) will(s) cookie(s) 90(s) diver(s) priorit**y**(ies)
mount(s) woodland(s) hymn(s) strait(s) corridor(s) preference(s) database(s) criticism(s) actress(es) dinosaur(s) econom**y**(ies)
bike(s) 70(s) facult**y**(ies) bead(s) ambassador(s) sediment(s) tent(s) marsh(es) filter(s) reject(s) advisor(s) landowner(s)
justice(s) disaster(s) special(s) instinct(s) orchestra(s) register(s) swimmer(s) receiver(s) mayor(s) remember(s) emperor(s)
motorcycle(s) automobile(s) reward(s) detective(s) intervention(s) suspicion(s) sacrifice(s) morale(s) future(s) amenit**y**(ies)
boiler(s) crusader(s) treasure(s) pop(s) beer(s) bubble(s) punch(es) owl(s) seller(s) explorer(s) partisan(s) minor(s)
acquisition(s) depiction(s) needle(s) ambition(s) vector(s) average(s) bracket(s) cylinder(s) announce(s) rush(es) pitch(es)
explosion(s) proportion(s) prefer(s) youth(s) hostage(s) stoke(s) lawsuit(s) recipe(s) subsidiar**y**(ies) treat(s) offender(s)
launch(es) bombing(s) chord(s) fault(s) split(s) reeve(s) educator(s) reflection(s) generate(s) trench(es) prototype(s)
ecosystem(s) climb(s) lash(es) bath(s) coast(s) nutrient(s) bandit(s) diar**y**(ies) mound(s) husband(s) convert(s) proponent(s)
enthusiast(s) boxer(s) hang(s) musical(s) engagement(s) miracle(s) collaborator(s) simpson(s) crash(es) repeat(s) ferr**y**(ies)
headline(s) pot(s) thumb(s) sim(s) old(s) instructor(s) concession(s) brewer(s) spaniard(s) corpse(s) medicine(s) skirt(s)
noun(s) bedroom(s) eyelid(s) capture(s) orbit(s) admission(s) energ**y**(ies) rod(s) cyclist(s) pet(s) journey(s) gesture(s)
comrade(s) salar**y**(ies) apple(s) conviction(s) realm(s) interior(s) slot(s) cemeter**y**(ies) fatalit**y**(ies) apostle(s) reptile(s) reef(s)
supplier(s) prostitute(s) bid(s) ax**is**(es) sex(es) socialist(s) slip(s) insurgent(s) oil(s) elevator(s) puzzle(s) twent**y**(ies)
songwriter(s) crossing(s) insight(s) stall(s) deck(s) medication(s) tendenc**y**(ies) robe(s) vineyard(s) wedding(s) academ**y**(ies)
storyline(s) wait(s) breeder(s) abuse(s) echo(es) initial(s) assassin(s) policem**an**(en) surround(s) exploit(s) front(s) glacier(s)

Singular form with associated plural modifications
gig(s) schedule(s) traveller(s) equal(s) classmate(s) turret(s) fork(s) sphere(s) mosque(s) indicator(s) avenue(s) assessment(s)
bench(es) gathering(s) capacit**y**(ies) subdivision(s) headlight(s) impression(s) prediction(s) total(s) spike(s) weekday(s)
consonant(s) bind(s) sum(s) pad(s) altitude(s) cabin(s) technician(s) marche(s) gear(s) rein(s) belonging(s) hamlet(s) hodge(s)
buccaneer(s) reservoir(s) alliance(s) rabbit(s) creation(s) distribution(s) shortage(s) burst(s) collect(s) swing(s) mongol(s)
trust(s) tornado(es) ensemble(s) incentive(s) lay(s) franchise(s) bicycle(s) economist(s) aesthetic(s) tumor(s) dose(s)
periodical(s) strength(s) narrative(s) suite(s) interface(s) correction(s) waterway(s) audition(s) fluid(s) burial(s) fine(s) gut(s)
gem(s) toilet(s) comed**y**(ies) aquatic(s) dungeon(s) believer(s) optic(s) bump(s) canon(s) forum(s) stereotype(s) protagonist(s)
rall**y**(ies) coefficient(s) basin(s) ace(s) 60(s) biograph**y**(ies) transmission(s) troph**y**(ies) out(s) ratio(s) skin(s) disagreement(s)
bruise(s) galax**y**(ies) assault(s) lesion(s) firefighter(s) tract(s) flaw(s) grass(es) favorite(s) shrine(s) chick(s) beverage(s)
el**f**(ves) execution(s) construction(s) swamp(s) hormone(s) commentar**y**(ies) baron(s) dynast**y**(ies) hint(s) trunk(s) trailer(s)
projection(s) controller(s) cabinet(s) bengal(s) antiquit**y**(ies) clock(s) oversee(s) convoy(s) commodit**y**(ies) memorial(s)
residue(s) pitcher(s) eat(s) invention(s) guarantee(s) boast(s) bolt(s) sermon(s) credential(s) notion(s) syllable(s) fool(s)
fish(es) supervisor(s) burrow(s) pose(s) configuration(s) distributor(s) diplomat(s) compilation(s) hazard(s) orphan(s) grin(s)
spread(s) transition(s) palace(s) grove(s) pain(s) progress(es) surprise(s) astro(s) carving(s) recount(s) laborer(s) voyage(s)
fence(s) antibod**y**(ies) ending(s) grenade(s) invasion(s) contrast(s) magistrate(s) reactor(s) padre(s) regent(s) virtue(s) spice(s)
volcano(es) captive(s) email(s) orchard(s) devotee(s) robotic(s) bout(s) wasp(s) compare(s) trainer(s) freedom(s) pause(s)
transform(s) cue(s) transformation(s) charter(s) independent(s) consultant(s) lobe(s) warehouse(s) gland(s) occupation(s)
deliver**y**(ies) labourer(s) garment(s) chapel(s) van(s) walker(s) aspiration(s) cartridge(s) passport(s) creditor(s) stress(es)
tongue(s) cheer(s) radical(s) profession(s) crow(s) skull(s) sentiment(s) weakness(es) nucle**us**(i) weed(s) berr**y**(ies) plaintiff(s)
snap(s) balloon(s) assign(s) bow(s) frigate(s) superior(s) conductor(s) plea(s) realit**y**(ies) helmet(s) matri**x**(ces) favor(s)
outsider(s) maneuver(s) cumming(s) husk**y**(ies) friar(s) skater(s) climate(s) appropriation(s) werewol**f**(ves) soundtrack(s)
fresco(es) bearing(s) adviser(s) merit(s) implement(s) newscast(s) parasite(s) congress(es) hybrid(s) spore(s) suck(s)
atrocit**y**(ies) pursuit(s) shipment(s) secretar**y**(ies) shrug(s) inside(s) pub(s) revelation(s) radio(s) expenditure(s) pension(s)
original(s) realise(s) appliance(s) twist(s) fantas**y**(ies) republic(s) scripture(s) franc(s) cruise(s) furnishing(s) boss(es)
convict(s) constituent(s) paint(s) newcomer(s) petition(s) confession(s) yell(s) illness(es) clause(s) hal**f**(ves) surname(s)
runway(s) screening(s) tyre(s) perception(s) sandwich(es) banner(s) empire(s) confederate(s) juvenile(s) diocese(s) boulder(s)
antholog**y**(ies) siren(s) disturbance(s) comedian(s) illustrate(s) ornament(s) adjustment(s) airfield(s) legion(s) cake(s)
printer(s) san(s) blog(s) mat(s) fountain(s) revision(s) pamphlet(s) designation(s) sixt**y**(ies) thinker(s) badge(s) gibbon(s)
routine(s) primar**y**(ies) accolade(s) terrace(s) reduction(s) hornet(s) pearl(s) rehearsal(s) controvers**y**(ies) replacement(s)
sequel(s) royalt**y**(ies) cancer(s) griffith(s) buy(s) advise(s) cast(s) moor(s) mutant(s) inhibitor(s) archer(s) pasture(s)
indication(s) footballer(s) pepper(s) airplane(s) idol(s) waterfall(s) tariff(s) offense(s) indoor(s) psychologist(s) antibiotic(s)
astronaut(s) fleet(s) sensation(s) stadium(s) working(s) rollin(s) pulse(s) regular(s) fox(es) nationalit**y**(ies) decline(s) excuse(s)
astronomer(s) stride(s) amateur(s) adherent(s) four(s) celtic(s) dealing(s) loser(s) integer(s) outlaw(s) crown(s) rescue(s)
fr**y**(ies) anchor(s) cavalier(s) shower(s) visual(s) neck(s) contender(s) loyalist(s) pharmaceutical(s) promoter(s) gospel(s)
gator(s) seeker(s) craftsm**an**(en) piston(s) fabric(s) regime(s) mall(s) parade(s) presenter(s) insult(s) widow(s) prophet(s)
urge(s) grassland(s) classification(s) escort(s) striker(s) distinction(s) mushroom(s) pedestrian(s) announcement(s)
simulation(s) disk(s) fair**y**(ies) analytic(s) scroll(s) dialogue(s) dome(s) choir(s) seizure(s) pronoun(s) mandate(s) skeleton(s)
commando(s) keeper(s) laser(s) timber(s) vibration(s) leaflet(s) regret(s) cosmetic(s) packet(s) truth(s) civilization(s)
quarr**y**(ies) adolescent(s) detainee(s) budget(s) prop(s) highlander(s) chromosome(s) counter(s) elect(s) ribbon(s) banker(s)
cris**is**(es) regulator(s) horror(s) fingernail(s) summon(s) thug(s) transmitter(s) northward(s) operative(s) revolution(s)
breaker(s) gallon(s) clone(s) arrival(s) bend(s) staple(s) artisan(s) moon(s) collapse(s) brush(es) trojan(s) supplement(s)
pod(s) intersection(s) demographic(s) racer(s) rapper(s) caste(s) accent(s) demonstrator(s) guerrilla(s) shooter(s) showcase(s)
rodent(s) regulate(s) high(s) croat(s) submission(s) supermarket(s) marlin(s) legislature(s) militia(s) praise(s) spin(s) crane(s)
bottom(s) formula(s) combatant(s) basic(s) afternoon(s) groove(s) detector(s) accommodation(s) billion(s) plastic(s) basket(s)
tentacle(s) perpetrator(s) junction(s) chuckle(s) blessing(s) float(s) casino(s) learner(s) avoid(s) elite(s) planner(s) trooper(s)
lodge(s) onion(s) vocalist(s) aggie(s) quaker(s) extra(s) charm(s) referee(s) drawer(s) hedge(s) taxpayer(s) grime(s)
telescope(s) occurrence(s) sink(s) administration(s) fare(s) bulb(s) banknote(s) protection(s) exile(s) sculptor(s) clerk(s)
polymer(s) horizon(s) guitarist(s) flavor(s) hire(s) comet(s) taxi(s) op(s) barge(s) finisher(s) eruption(s) cart(s) plaque(s)
cascade(s) crater(s) mansion(s) excerpt(s) friday(s) barn(s) investigate(s) dane(s) merger(s) vaccine(s) conversion(s)
introduction(s) puppet(s) hine(s) serial(s) rhyme(s) confine(s) react(s) accuse(s) transformer(s) batter(s) sabre(s) theorist(s)
saxon(s) summit(s) extract(s) bundle(s) waiver(s) research(es) rent(s) principal(s) favourite(s) gunner(s) coyote(s) trigger(s)

Singular form with associated plural modifications
cord(s) shard(s) translator(s) footprint(s) moral(s) collision(s) bribe(s) terminate(s) emergenc**y**(ies) inquir**y**(ies) stat(s)
triangle(s) amphibian(s) curse(s) stint(s) sevent**y**(ies) massacre(s) inspector(s) portal(s) budd**y**(ies) bonus(es) compartment(s)
bodyguard(s) endeavor(s) sandal(s) eastward(s) jersey(s) fluctuation(s) blazer(s) engraving(s) theologian(s) gunm**an**(en)
solid(s) starter(s) receipt(s) slice(s) stocking(s) cook(s) herd(s) towel(s) italic(s) abnormalit**y**(ies) primate(s) fift**y**(ies)
fracture(s) pen(s) decree(s) flashback(s) lease(s) programmer(s) gasp(s) expansion(s) commit(s) alloy(s) apolog**y**(ies) stitch(es)
wavelength(s) drill(s) vent(s) scan(s) attendant(s) passion(s) storey(s) ape(s) vault(s) grocer**y**(ies) robber(s) thirt**y**(ies)
detachment(s) mortal(s) chore(s) harmon**y**(ies) beaver(s) affiliation(s) grip(s) byzantine(s) climber(s) illusion(s) fortress(es)
torch(es) psalm(s) desert(s) grader(s) flap(s) dragoon(s) billing(s) fitting(s) grower(s) shooting(s) scandal(s) mast(s)
subordinate(s) friendship(s) propeller(s) antique(s) tier(s) abortion(s) landlord(s) pigeon(s) carpet(s) tractor(s) feminist(s)
miniature(s) maverick(s) pace(s) vertebrate(s) dweller(s) blast(s) cuff(s) mortar(s) booth(s) chef(s) airm**an**(en) sweet(s) tap(s)
commuter(s) income(s) mollusk(s) pea(s) output(s) shifter(s) amplifier(s) canton(s) conspirator(s) biscuit(s) sighting(s)
chase(s) chunk(s) scissor(s) federation(s) inspection(s) brewer**y**(ies) depart(s) hinge(s) shipyard(s) extreme(s) frown(s)
shilling(s) ind**ex**(ices) dagger(s) hypothes**is**(es) snack(s) clement(s) monday(s) cossack(s) swede(s) siding(s) barbarian(s)
acquaintance(s) texture(s) eight**y**(ies) shock(s) pleasure(s) manifestation(s) punishment(s) spouse(s) invitation(s) raptor(s)
conquest(s) rack(s) hunt(s) motivation(s) triple(s) still(s) giggle(s) embass**y**(ies) bearer(s) peanut(s) ration(s) bowler(s)
explode(s) eel(s) youngster(s) playwright(s) administer(s) nickname(s) electrode(s) grouping(s) reformer(s) complete(s)
earring(s) ban(s) gather(s) droplet(s) toll(s) buddhist(s) yoursel**f**(ves) sidewalk(s) therap**y**(ies) semantic(s) sortie(s) pixel(s)
numeral(s) squirrel(s) slam(s) emigrant(s) organiser(s) moan(s) girlfriend(s) socket(s) symphon**y**(ies) manual(s) relay(s)
overlook(s) fringe(s) cage(s) yacht(s) stunt(s) surger**y**(ies) twig(s) dinner(s) recover(s) toxin(s) disadvantage(s) construct(s)
synthesizer(s) stain(s) constant(s) cal**f**(ves) bathroom(s) warrant(s) goddess(es) tit(s) comb(s) parod**y**(ies) ditch(es) thrust(s)
juror(s) noodle(s) blend(s) deficienc**y**(ies) housemate(s) specialt**y**(ies) tender(s) preview(s) entrant(s) crate(s) excursion(s)
activate(s) henchm**an**(en) messenger(s) font(s) nipple(s) immortal(s) precaution(s) nil(s) hub(s) probe(s) interrupt(s) mood(s)
murderer(s) lesbian(s) roller(s) rag(s) discharge(s) clipper(s) animation(s) ordinance(s) spelling(s) stalk(s) frontier(s)
signator**y**(ies) dictionar**y**(ies) assure(s) canoe(s) facet(s) strut(s) cushion(s) vandal(s) fingerprint(s) partition(s) restraint(s)
retreat(s) deficit(s) chant(s) weaver(s) launcher(s) duet(s) coil(s) pebble(s) befriend(s) pavilion(s) mentor(s) rebellion(s)
thorn(s) pesticide(s) pyramid(s) parliament(s) niche(s) camel(s) byte(s) ideolog**y**(ies) superhero(es) linguist(s) departure(s)
scorer(s) cooperative(s) dive(s) colonel(s) accountant(s) clutch(es) token(s) awaken(s) soloist(s) antagonist(s) condom(s)
tanker(s) challenger(s) lad(s) orchid(s) glider(s) cutter(s) fat(s) broker(s) anarchist(s) subscription(s) dependenc**y**(ies)
minion(s) shepherd(s) scratch(es) nose(s) forecast(s) cleric(s) mustang(s) pipeline(s) grievance(s) facade(s) pastor(s) debut(s)
don(s) remed**y**(ies) mon(s) slack(s) empt**y**(ies) telephone(s) photon(s) foe(s) 20(s) radar(s) densit**y**(ies) substitute(s)
adjective(s) tee(s) bunker(s) hound(s) funeral(s) thunderstorm(s) interchange(s) badger(s) headache(s) mosaic(s) maid(s)
simon(s) o**’clock**(’ clock) brief(s) trainee(s) bandage(s) homosexual(s) cafe(s) meyer(s) portfolio(s) alarm(s) smartphone(s)
lid(s) furnace(s) orange(s) captor(s) nightclub(s) neutron(s) blind(s) pajama(s) inhibit(s) equivalent(s) vacanc**y**(ies)
evaluation(s) premiere(s) garrison(s) trumpet(s) pancake(s) rune(s) licence(s) bet(s) blossom(s) delegation(s) forearm(s)
hoo**f**(ves) pan(s) chest(s) transcript(s) bower(s) sect(s) deployment(s) planter(s) pest(s) ruling(s) rivalr**y**(ies) physicist(s)
synagogue(s) liner(s) dye(s) skyscraper(s) click(s) millimeter(s) pointer(s) bicep(s) shutter(s) symbolize(s) flare(s) pollutant(s)
subtitle(s) steamer(s) tug(s) shoal(s) dispatch(es) greeting(s) projectile(s) gamer(s) baptist(s) mountaineer(s) ware(s) desk(s)
gown(s) roadway(s) idiot(s) paramedic(s) pursue(s) lieutenant(s) cater(s) talon(s) olive(s) installment(s) rupee(s) postcard(s)
setback(s) commandment(s) scare(s) quer**y**(ies) recorder(s) piano(s) whaler(s) guild(s) growl(s) riff(s) exploration(s) rerun(s)
shiver(s) achieve(s) jack(s) anomal**y**(ies) bang(s) housewi**fe**(ves) dan(s) recital(s) advertiser(s) philosoph**y**(ies) precursor(s)
flock(s) fur(s) princess(es) royalist(s) cathedral(s) cheekbone(s) pore(s) fate(s) romance(s) beneficiar**y**(ies) incarnation(s)
migration(s) arcade(s) altar(s) ensue(s) uprising(s) grizzl**y**(ies) auction(s) marshal(s) outcrop(s) cube(s) entertainer(s)
gangster(s) peach(es) crewm**an**(en) smith(s) may(s) rooster(s) implant(s) due(s) tribute(s) quotation(s) enhancement(s)
jur**y**(ies) embrace(s) ballet(s) constitution(s) gladiator(s) suffix(es) potential(s) percentage(s) drag(s) anecdote(s)
hallucination(s) gray(s) kidney(s) drummer(s) apologize(s) instruct(s) latitude(s) paratrooper(s) rocker(s) dormitor**y**(ies)
connector(s) manor(s) carpenter(s) traverse(s) iron(s) subgroup(s) slum(s) homeowner(s) steroid(s) flute(s) vulture(s)
treatise(s) outpost(s) workm**an**(en) slipper(s) medalist(s) quarterback(s) sympath**y**(ies) pope(s) pigment(s) liabilit**y**(ies)
surfer(s) coincide(s) declaration(s) acoustic(s) dissident(s) aide(s) parliamentarian(s) pupp**y**(ies) meat(s) aristocrat(s) pro(s)
overlap(s) scent(s) flake(s) traitor(s) fumble(s) amend(s) concentrate(s) math(s) vase(s) cereal(s) gust(s) crocodile(s)
bulletin(s) signing(s) freak(s) expose(s) prophec**y**(ies) pie(s) winning(s) subset(s) fable(s) fifth(s) pedal(s) revolt(s) bride(s)
editorial(s) cracker(s) mason(s) mosquito(es) correspondent(s) flip(s) admiral(s) probabilit**y**(ies) canyon(s) scorpion(s)

Singular form with associated plural modifications
throne(s) free(s) vassal(s) closure(s) chill(s) fort**y**(ies) currenc**y**(ies) maroon(s) balcon**y**(ies) friendl**y**(ies) consultation(s)
sideline(s) proposition(s) eater(s) cavern(s) edmond(s) gen(s) mixture(s) vamp(s) litre(s) feat(s) seam(s) counselor(s)
mormon(s) superstar(s) bucket(s) rug(s) tutor(s) gunboat(s) intruder(s) bachelor(s) armament(s) umpire(s) alert(s) pat(s)
ladder(s) hobb**y**(ies) berth(s) incursion(s) concerto(s) burger(s) chieftain(s) sewer(s) breach(es) skier(s) compliment(s) mail(s)
motorist(s) shutout(s) exemption(s) geologist(s) fraction(s) hull(s) pacer(s) bellow(s) ounce(s) reparation(s) constable(s)
leopard(s) apprentice(s) insert(s) catalogue(s) proton(s) waste(s) fibre(s) masterpiece(s) architecture(s) anthropologist(s)
sticker(s) hydrocarbon(s) reflex(es) plug(s) shipwreck(s) camper(s) balance(s) spectacle(s) biologist(s) billiard(s) fashion(s)
scrap(s) briton(s) gull(s) magnet(s) connor(s) tre(s) smuggler(s) resume(s) axiom(s) plead(s) rape(s) critique(s) bobcat(s)
filament(s) eliminate(s) census(es) coral(s) initiate(s) shopper(s) television(s) milestone(s) adversar**y**(ies) bracelet(s)
transistor(s) planting(s) longhorn(s) trademark(s) necessit**y**(ies) impulse(s) resolve(s) depression(s) rental(s) rotation(s)
dwar**f**(ves) directive(s) viewpoint(s) medium(s) affection(s) vitamin(s) privateer(s) hispanic(s) strive(s) fix(es) 30(s) comfort(s)
western(s) cult(s) inventor(s) traged**y**(ies) backer(s) golfer(s) dividend(s) quest(s) solvent(s) harbor(s) negative(s) gyps**y**(ies)
welcome(s) doorway(s) fairground(s) edit(s) toss(es) whistle(s) glimpse(s) envelope(s) ambulance(s) grunt(s) grandson(s)
aviator(s) claimant(s) stronghold(s) proprietor(s) buckeye(s) shingle(s) crusade(s) sentinel(s) contradiction(s) bandmate(s)
pon**y**(ies) gender(s) feast(s) methodolog**y**(ies) breech(es) whip(s) mortgage(s) staff(s) wednesday(s) stumble(s) robber**y**(ies)
exposure(s) enclosure(s) lancer(s) kidnap(s) mutter(s) freedm**an**(en) therapist(s) chaplain(s) brothel(s) censor(s) reappear(s)
hank(s) confrontation(s) deacon(s) velocit**y**(ies) procession(s) buttress(es) nav**y**(ies) bark(s) airstrike(s) sonnet(s) pickup(s)
tapestr**y**(ies) lighthouse(s) gold(s) upland(s) wither(s) fundamental(s) calorie(s) shut(s) browser(s) freeze(s) expatriate(s)
mare(s) sparrow(s) tribunal(s) rotate(s) metaphor(s) lender(s) watcher(s) pellet(s) railing(s) walkway(s) howitzer(s)
pediatric(s) strawberr**y**(ies) interpreter(s) main(s) steward(s) podcast(s) anger(s) wipe(s) commoner(s) willow(s) ancient(s)
advancement(s) essential(s) tray(s) refiner**y**(ies) riding(s) discourse(s) specific(s) breathe(s) tenet(s) rampart(s) catalyst(s)
thermodynamic(s) goon(s) layout(s) on(s) oath(s) iris(es) laptop(s) stallion(s) fugitive(s) smoker(s) assurance(s) lick(s) lee(s)
flop(s) aunt(s) slug(s) roommate(s) tab(s) coalition(s) fascist(s) bust(s) connotation(s) bull**y**(ies) warlord(s) beggar(s)
checkpoint(s) hood(s) hitter(s) adventurer(s) gradient(s) recollection(s) chemist(s) consume(s) alternate(s) mug(s) chorus(es)
gardener(s) clump(s) juice(s) prime(s) canvas(es) oyster(s) wash(es) belle(s) islet(s) evil(s) wander(s) mystic(s) appendage(s)
seminole(s) temptation(s) alley(s) allowance(s) squeeze(s) ailment(s) testimon**y**(ies) admirer(s) cutting(s) feud(s) injection(s)
rainforest(s) ass(es) deviation(s) fraternit**y**(ies) persist(s) inspiration(s) gauge(s) great(s) staircase(s) marseille(s) spill(s)
attachment(s) miller(s) gorge(s) hen(s) assertion(s) scrub(s) mangrove(s) reel(s) coating(s) announcer(s) throat(s) restroom(s)
violate(s) melt(s) cellar(s) surveyor(s) convey(s) allusion(s) zealander(s) vacation(s) thunderbird(s) dio(s) chairm**an**(en)
methodist(s) musket(s) lagoon(s) batch(es) sitcom(s) inconsistenc**y**(ies) franciscan(s) ranch(es) dominion(s) constellation(s)
ski(s) manifest(s) lecturer(s) goth(s) animator(s) marble(s) drift(s) volt(s) principalit**y**(ies) sacrament(s) gadget(s) reaper(s)
coaster(s) dip(s) punt(s) coffin(s) souvenir(s) tremor(s) eclipse(s) reprint(s) portrayal(s) thes**is**(es) billboard(s) pencil(s)
sergeant(s) attention(s) solicitor(s) spoil(s) malay(s) forensic(s) palisade(s) viper(s) sorrow(s) quit(s) extremist(s) tenth(s)
formulation(s) playground(s) hoop(s) clown(s) pelican(s) streetcar(s) brace(s) skate(s) gay(s) locker(s) eyeball(s) galley(s)
rip(s) analyze(s) rogue(s) trio(s) assassination(s) overhear(s) placement(s) seedling(s) discrepanc**y**(ies) roster(s) incumbent(s)
potter(s) low(s) endorsement(s) parrot(s) lil**y**(ies) beginner(s) sooner(s) anthem(s) predate(s) gag(s) favour(s) grid(s) bye(s)
proverb(s) condominium(s) salon(s) muse(s) suicide(s) cleaner(s) psychiatrist(s) forbid(s) cocktail(s) quartet(s) additive(s)
railcar(s) collar(s) certification(s) otter(s) jockey(s) motorway(s) hacker(s) archbishop(s) mold(s) rancher(s) precinct(s)
lifeboat(s) phantom(s) happening(s) magpie(s) pageant(s) iteration(s) pour(s) courtier(s) dessert(s) harrier(s) fill**y**(ies) duct(s)
terrier(s) brazilian(s) turnover(s) pledge(s) mist(s) addict(s) ethnicit**y**(ies) coal(s) crest(s) aggregate(s) kidnapper(s)
reconstruction(s) etching(s) prejudice(s) protector(s) chimpanzee(s) handler(s) venetian(s) prank(s) bumper(s) pal(s)
emblem(s) suitcase(s) sprite(s) hoosier(s) ember(s) tor**y**(ies) evoke(s) epic(s) saracen(s) grace(s) cymbal(s) platoon(s)
gunshot(s) composite(s) allele(s) loyalt**y**(ies) carbohydrate(s) samuel(s) corvette(s) swell(s) kiln(s) spire(s) reproduction(s)
scanner(s) blogger(s) contraction(s) poison(s) gopher(s) cheese(s) zero(s) foll**y**(ies) baker(s) pregnanc**y**(ies) speculation(s)
con(s) curator(s) behave(s) tick(s) prairie(s) goalkeeper(s) eight(s) winer**y**(ies) index(es) capacitor(s) innocent(s) reunite(s)
bra(s) beating(s) theorem(s) tweet(s) sniper(s) ore(s) voltage(s) sponge(s) sleeper(s) folder(s) garage(s) projector(s) helper(s)
membership(s) councilor(s) sage(s) oscillation(s) safeguard(s) sheriff(s) fireplace(s) disguise(s) novelist(s) swear(s)
sensibilit**y**(ies) complexit**y**(ies) pianist(s) diagnostic(s) porch(es) biker(s) precede(s) lunch(es) truss(es) epistle(s) correlation(s)
abstract(s) delusion(s) discount(s) ayer(s) revival(s) pear(s) fourth(s) shred(s) warhead(s) interruption(s) mimic(s) excess(es)
feeder(s) illustrator(s) rookie(s) orbital(s) intestine(s) cadre(s) cheerleader(s) cure(s) congressm**an**(en) industrialist(s) taunt(s)

Singular form with associated plural modifications
pumpkin(s) crook(s) permutation(s) kink(s) distraction(s) riddle(s) saying(s) raft(s) accomplice(s) atheist(s) director**y**(ies)
couch(es) reprisal(s) inductee(s) liar(s) novice(s) fertilizer(s) freemason(s) countrym**an**(en) hatch(es) beacon(s) sprint(s)
extremit**y**(ies) liter(s) suitor(s) bias(es) foal(s) tilt(s) olympian(s) calculate(s) amazon(s) hon(s) lodging(s) harm(s)
separatist(s) shotgun(s) sausage(s) commodore(s) crevice(s) st(s) pickle(s) vulnerabilit**y**(ies) bitch(es) stud(s) password(s)
oar(s) rendition(s) valuable(s) healer(s) template(s) chocolate(s) entertainment(s) ache(s) rim(s) gymnast(s) loom(s)
paragraph(s) barricade(s) prescription(s) disparit**y**(ies) replay(s) horde(s) grand(s) wildfire(s) progressive(s) correlate(s)
chariot(s) bureaucrat(s) footnote(s) aerodynamic(s) sultan(s) glare(s) footpath(s) reverse(s) scribe(s) mage(s) frustration(s)
heron(s) abbreviation(s) rotor(s) restoration(s) spiral(s) shove(s) complement(s) gorilla(s) errand(s) auxiliar**y**(ies) victor(s)
revolver(s) staffer(s) semiconductor(s) compatriot(s) upset(s) mussel(s) abbot(s) rarit**y**(ies) watercolor(s) bristle(s)
premium(s) nucleotide(s) ligament(s) courtyard(s) spoon(s) kicker(s) greet(s) auditor(s) rooftop(s) dyke(s) arie(s)
disappearance(s) midwi**fe**(ves) templar(s) sanctuar**y**(ies) intervene(s) hell(s) midshipm**an**(en) instrumental(s) alter(s) fuse(s)
soap(s) spitfire(s) handgun(s) endeavour(s) nick(s) beaut**y**(ies) cohort(s) cherr**y**(ies) forgive(s) penn**y**(ies) assailant(s) cease(s)
informant(s) borrow(s) donut(s) pawn(s) suspension(s) resist(s) lament(s) conception(s) bypass(es) drawback(s) babe(s)
carcass(es) mep(s) intrigue(s) diner(s) gee(s) lifetime(s) nuance(s) deliberation(s) clam(s) flick(s) gatherer(s) pharmac**y**(ies)
oscar(s) splinter(s) tavern(s) dressing(s) spinner(s) sinner(s) patriarch(s) pixie(s) attendance(s) silver(s) hollow(s)
caricature(s) perk(s) naturalist(s) warbler(s) linebacker(s) saloon(s) toad(s) heroine(s) posting(s) vest(s) reagent(s) stool(s)
porter(s) pastr**y**(ies) enclave(s) rescuer(s) niece(s) medic(s) monologue(s) crutch(es) tendon(s) moss(es) tom(s) advert(s) fig(s)
therapeutic(s) dial(s) superpower(s) genital(s) savage(s) diameter(s) capitalist(s) forget(s) magnate(s) outing(s) octave(s)
exponent(s) first(s) virgin(s) lever(s) elm(s) jurist(s) stance(s) speculate(s) pagan(s) blueprint(s) wildflower(s) biographer(s)
skit(s) caregiver(s) hog(s) freeway(s) rower(s) reminder(s) bystander(s) cuisine(s) uncertaint**y**(ies) butcher(s) bass(es)
enforcer(s) misunderstanding(s) mom(s) recess(es) sweat(s) booster(s) inspire(s) plume(s) carol(s) chandelier(s)
environmentalist(s) majorit**y**(ies) chop(s) nurser**y**(ies) suture(s) orientation(s) gum(s) caller(s) boom(s) converge(s) halt(s)
barrow(s) cedar(s) spit(s) gateway(s) multiple(s) residenc**y**(ies) diagnos**is**(es) prohibition(s) puff(s) steppe(s) copyright(s)
transcription(s) compromise(s) hesitate(s) life(s) registration(s) duel(s) battlefield(s) chancellor(s) goldsmith(s) pundit(s)
examiner(s) rap(s) clearance(s) pharmacist(s) deduction(s) endowment(s) particular(s) manoeuvre(s) boutique(s) bronze(s)
infrastructure(s) innovator(s) transplant(s) bot(s) six(es) best(s) democrac**y**(ies) distortion(s) west(s) borrower(s) binding(s)
spare(s) leaning(s) offensive(s) pheromone(s) increment(s) byrd(s) donkey(s) wallab**y**(ies) mascot(s) triplet(s) summar**y**(ies)
deflection(s) opposite(s) commence(s) benefactor(s) disagree(s) bible(s) bowel(s) payload(s) thruster(s) bogie(s) delight(s)
necklace(s) bother(s) guess(es) hun(s) airship(s) affinit**y**(ies) sauce(s) harvest(s) heater(s) allerg**y**(ies) canine(s) stove(s)
bookstore(s) medallion(s) dentist(s) disposal(s) outage(s) woe(s) grasshopper(s) persecution(s) snippet(s) mole(s) precedent(s)
agenda(s) kernel(s) whar**f**(ves) earth(s) injustice(s) dodd(s) silhouette(s) disruption(s) girder(s) salad(s) completion(s)
workplace(s) seagull(s) confide(s) cloth(s) bastion(s) linkage(s) prom(s) waiter(s) bounce(s) botanist(s) bookseller(s)
huguenot(s) timer(s) spasm(s) mumble(s) checker(s) homer(s) haunt(s) druid(s) alias(es) amp(s) termite(s) dormer(s)
alcohol(s) sel**f**(ves) fell(s) edict(s) verandah(s) fisher(s) exclaim(s) fender(s) approve(s) contingent(s) gargoyle(s) greenhouse(s)
leftover(s) gambler(s) foster(s) computation(s) peg(s) clay(s) sayer(s) execute(s) withdrawal(s) bohemian(s) shriek(s)
overture(s) cramp(s) shank(s) tiptoe(s) weave(s) oboe(s) dorm(s) englishm**an**(en) sociologist(s) annotation(s) hierarch**y**(ies)
emerald(s) showing(s) pylon(s) roar(s) alkaloid(s) intermediate(s) skeptic(s) thrill(s) dr**y**(ies) educate(s) troupe(s) legac**y**(ies)
ulcer(s) appointee(s) blotch(es) choreographer(s) baker**y**(ies) evangelist(s) tithe(s) hale(s) mennonite(s) retainer(s) lace(s)
alignment(s) allocation(s) converter(s) conifer(s) turkey(s) prospector(s) diode(s) rambler(s) fleck(s) simulator(s) tense(s)
purse(s) heal(s) marathon(s) cache(s) cartel(s) orphanage(s) observator**y**(ies) emissar**y**(ies) ascend(s) contradict(s)
mistress(es) birthday(s) millionaire(s) escalator(s) eyewitness(es) brochure(s) impairment(s) lemur(s) butter(s) serpent(s)
were(s) cam(s) ravine(s) mover(s) alphabet(s) heartbeat(s) dissolve(s) ovar**y**(ies) surge(s) bungalow(s) concubine(s)
livelihood(s) workstation(s) booklet(s) purchaser(s) mint(s) vice(s) navigator(s) soup(s) swim(s) sentr**y**(ies) perceive(s)
wrong(s) sucker(s) mumm**y**(ies) insider(s) quilt(s) airliner(s) analogue(s) houseguest(s) runaway(s) soybean(s) repositor**y**(ies)
buffer(s) puddle(s) whore(s) pressing(s) sting(s) duch**y**(ies) cyst(s) tabloid(s) trajector**y**(ies) tracker(s) speck(s) insecurit**y**(ies)
griffin(s) dame(s) resin(s) potion(s) cargo(es) catcher(s) lemon(s) rainbow(s) refreshment(s) snatch(es) pilgrimage(s) scrape(s)
loudspeaker(s) inclusion(s) lev**y**(ies) forester(s) steamboat(s) newborn(s) podium(s) deadline(s) bison(s) napkin(s)
intermediar**y**(ies) coconut(s) tight(s) booking(s) parachute(s) interfere(s) fundraiser(s) convent(s) dependent(s) minefield(s)
blister(s) byrne(s) diaper(s) spray(s) admire(s) evidence(s) homicide(s) martian(s) digger(s) cockroach(es) affirm(s) par(s)
abolitionist(s) hive(s) antiquar**y**(ies) blaster(s) inventor**y**(ies) chronicler(s) covenant(s) clipping(s) trapper(s) axon(s)

Singular form with associated plural modifications
dais**y**(ies) quarrel(s) grey(s) interrogation(s) midfielder(s) lobbyist(s) powder(s) confer(s) sportsm**an**(en) meditation(s)
enroll(s) imprint(s) woodpecker(s) homeland(s) wild(s) newsletter(s) moderate(s) good**y**(ies) tease(s) lander(s) pouch(es)
adjust(s) farmhouse(s) clamp(s) ambush(es) frenchm**an**(en) casting(s) forge(s) overall(s) transporter(s) purge(s) consort(s)
maritime(s) substitution(s) reunion(s) mandible(s) debtor(s) swap(s) advisor**y**(ies) blocker(s) lobster(s) clarinet(s) headliner(s)
embankment(s) conqueror(s) ledge(s) dare(s) outburst(s) router(s) hose(s) calculator(s) siege(s) intrusion(s) scar**f**(ves) suv(s)
inherit(s) canop**y**(ies) isolate(s) trombone(s) pinnacle(s) buckle(s) storefront(s) hike(s) integral(s) howl(s) pray(s) cape(s)
screenwriter(s) heli**x**(ces) hippie(s) dale(s) flashlight(s) cand**y**(ies) reputation(s) failing(s) stork(s) scavenger(s) ink(s)
mould(s) david(s) quark(s) thicket(s) going(s) harper(s) sweater(s) torture(s) gemstone(s) dash(es) conglomerate(s)
abraham(s) long(s) gaul(s) furrow(s) cornice(s) subsystem(s) recover**y**(ies) identifier(s) symmetr**y**(ies) breakdown(s) trope(s)
resignation(s) lorr**y**(ies) competenc**y**(ies) addiction(s) fur**y**(ies) header(s) widower(s) unicorn(s) kidnapping(s) headstone(s)
undertaking(s) steel(s) notation(s) supreme(s) financier(s) manipulation(s) ever(s) warm(s) disclosure(s) modalit**y**(ies)
umbrella(s) magnitude(s) approximation(s) stag(s) firefl**y**(ies) secretion(s) regional(s) aboriginal(s) bookshel**f**(ves) caution(s)
mock(s) enquir**y**(ies) talkie(s) clap(s) emplacement(s) campaigner(s) evangelical(s) preside(s) postulate(s) knockout(s)
luxur**y**(ies) outcast(s) drinker(s) thunderbolt(s) imitation(s) dail**y**(ies) aker(s) sororit**y**(ies) slovak(s) mattress(es) jackal(s)
peripheral(s) judgement(s) alchemist(s) campground(s) pogrom(s) silk(s) pun(s) supercar(s) shipbuilder(s) brownie(s) bun(s)
toddler(s) hiss(es) saucer(s) superintendent(s) fed(s) pairing(s) shaker(s) sorcerer(s) suffering(s) snort(s) understanding(s)
crawl(s) paddle(s) homestead(s) moore(s) paperback(s) remake(s) geometr**y**(ies) swarm(s) perfume(s) conspirac**y**(ies) assay(s)
trawler(s) reproduce(s) interlude(s) odor(s) heading(s) training(s) plague(s) atmosphere(s) blacksmith(s) geographer(s)
chaser(s) typhoon(s) slayer(s) lineup(s) adapt(s) typeface(s) indictment(s) amulet(s) germ(s) monarch**y**(ies) lance(s) boil(s)
covert(s) taylor(s) vial(s) petitioner(s) townhouse(s) wedge(s) coupon(s) panic(s) exporter(s) genealog**y**(ies) salamander(s)
raine(s) tutorial(s) lintel(s) jain(s) tailor(s) livre(s) aquarium(s) bidder(s) piercing(s) chat(s) adventist(s) collage(s)
efficienc**y**(ies) tenement(s) tectonic(s) cybernetic(s) autograph(s) minutem**an**(en) foul(s) dude(s) alligator(s) tortoise(s)
filing(s) raccoon(s) columnist(s) muppet(s) sub(s) preliminar**y**(ies) caress(es) overcome(s) boulevard(s) platelet(s) scooter(s)
levee(s) fake(s) private(s) thrower(s) spa(s) racehorse(s) crag(s) passer(s) reinforce(s) exterior(s) overseer(s) coup(s) pardon(s)
barrister(s) absence(s) dub(s) lookout(s) steer(s) condemn(s) notch(es) hurr**y**(ies) freighter(s) peril(s) dreamer(s)
coordinator(s) fen(s) voucher(s) buster(s) canister(s) aquifer(s) rub**y**(ies) formalit**y**(ies) gravestone(s) negotiator(s) banquet(s)
antidepressant(s) dismissal(s) centurion(s) evacuation(s) cru**x**(ces) forger**y**(ies) rectangle(s) raisin(s) devote(s) trapping(s)
shear(s) theft(s) prefect(s) inlet(s) positive(s) foray(s) permission(s) exposition(s) vesper(s) verdict(s) picnic(s) snow(s)
prelate(s) elective(s) ural(s) crush(es) flirt(s) flourish(es) supervise(s) cough(s) equate(s) spoke(s) harness(es) superstition(s)
gnome(s) idiom(s) uncover(s) mobster(s) timeline(s) tub(s) collectible(s) generalization(s) faerie(s) archetype(s) writ(s)
schooner(s) liaison(s) cannibal(s) nature(s) prism(s) obstruction(s) follicle(s) billionaire(s) covering(s) mature(s) intent(s)
dumm**y**(ies) sophomore(s) surrender(s) caption(s) sofa(s) supersonic(s) 10(s) rake(s) instrumentalist(s) intake(s) bulkhead(s)
beet(s) ascent(s) compressor(s) harlequin(s) colonial(s) apparition(s) misconception(s) recognition(s) cheetah(s) lattice(s)
accelerator(s) hydraulic(s) blonde(s) importer(s) shovel(s) thoroughfare(s) lumberjack(s) splash(es) tryout(s) panelist(s)
access(es) mag(s) woolworth(s) inclination(s) misfortune(s) fil(s) pronunciation(s) fissure(s) strengthen(s) hallmark(s)
assemblage(s) hanger(s) cipher(s) gull**y**(ies) diesel(s) divorce(s) mediator(s) dwell(s) trolley(s) liberator(s) accelerate(s)
repetition(s) buoy(s) palacio(s) hover(s) sprout(s) adapter(s) notification(s) sprinter(s) approval(s) glide(s) improvisation(s)
loophole(s) sailboat(s) scoop(s) secure(s) landfill(s) oscillator(s) contraceptive(s) retire(s) inference(s) conquistador(s)
passageway(s) more(s) elk(s) hummingbird(s) backpack(s) dick(s) transept(s) thrasher(s) timetable(s) warlock(s) couplet(s)
fanatic(s) aqueduct(s) sloop(s) archeologist(s) diverge(s) neanderthal(s) refrigerator(s) ford(s) obituar**y**(ies) philanthropist(s)
seduce(s) dragonfl**y**(ies) anxiet**y**(ies) cream(s) lobb**y**(ies) ballistic(s) briefing(s) gym(s) burner(s) playlist(s) intercept(s)
hamburger(s) maple(s) queue(s) registr**y**(ies) minstrel(s) meteor(s) coastline(s) finch(es) terror(s) taboo(s) crease(s) slate(s)
racket(s) baptism(s) tyrant(s) antecedent(s) executor(s) trim(s) appreciate(s) leftist(s) waitress(es) spotlight(s) dad(s)
dissertation(s) echelon(s) basement(s) basque(s) bunn**y**(ies) casing(s) cask(s) wrapper(s) nettle(s) piper(s) ironclad(s)
culprit(s) carnival(s) snapshot(s) bombardment(s) intensit**y**(ies) sore(s) vigilante(s) pelt(s) oracle(s) snowflake(s) refinement(s)
muffin(s) ram**an**(en) platter(s) consult(s) cookbook(s) veil(s) coward(s) humanist(s) cistern(s) ghetto(s) recite(s)
disposition(s) exhale(s) ornithologist(s) cucumber(s) pretender(s) ode(s) transcend(s) shin(s) paleontologist(s) skid(s)
omission(s) capillar**y**(ies) proclamation(s) irregular(s) sparkle(s) printing(s) hanging(s) questionnaire(s) yellow(s) yarn(s)
gcse(s) mishap(s) width(s) atoll(s) locust(s) coincidence(s) slash(es) stairway(s) vapor(s) forerunner(s) shamrock(s)
neurotransmitter(s) deduce(s) revert(s) granddaughter(s) pallet(s) boomer(s) walnut(s) baron**y**(ies) awake(s) diversion(s)

Singular form with associated plural modifications
insecticide(s) lotter**y**(ies) offset(s) hermit(s) contemplate(s) scam(s) mixer(s) crossover(s) hack(s) quay(s) correspondence(s)
upheaval(s) posture(s) shrink(s) bell**y**(ies) thoroughbred(s) primitive(s) cloister(s) saxophone(s) curiosit**y**(ies) traditionalist(s)
snarl(s) linger(s) conform(s) veer(s) pizza(s) filling(s) canar**y**(ies) watercourse(s) poke(s) modem(s) tumour(s) retort(s)
blackmail(s) inquire(s) conservationist(s) opioid(s) phoenician(s) penetrate(s) chute(s) nonprofit(s) shopkeeper(s) perch(es)
smack(s) tangle(s) bard(s) syndicate(s) lubricant(s) silence(s) stipulation(s) avalanche(s) courthouse(s) foundr**y**(ies) dor(s)
storyteller(s) rosette(s) ill(s) imbalance(s) calm(s) shack(s) hampton(s) abstraction(s) elaborate(s) litter(s) plum(s) tuck(s)
wreath(s) taping(s) annual(s) viaduct(s) jacobite(s) exert(s) bab(s) conjecture(s) façade(s) searchlight(s) barb(s)
transponder(s) herder(s) bestseller(s) engraver(s) baton(s) literature(s) drifter(s) cod**ex**(ices) courier(s) resistor(s) lapel(s)
groom(s) enforce(s) dramatist(s) minaret(s) flicker(s) shudder(s) bree(s) wraith(s) parapet(s) creole(s) assemble(s)
specialization(s) wheeler(s) refuge(s) radiator(s) aeroplane(s) memo(s) bassoon(s) plunge(s) transgression(s) equit**y**(ies)
epithet(s) prelude(s) grandfather(s) rowd**y**(ies) rom(s) cheque(s) carton(s) injunction(s) ease(s) blackout(s) perturbation(s)
clearing(s) glade(s) vive(s) ale(s) dreadnought(s) rendering(s) shallow(s) squeal(s) elimination(s) lapse(s) imp(s) signer(s)
caliph(s) leech(es) garland(s) allegiance(s) pastel(s) hitm**an**(en) recruiter(s) parkland(s) tester(s) interceptor(s) stingray(s)
scold(s) referral(s) three(s) utterance(s) bivalve(s) cobblestone(s) cartographer(s) coupling(s) parlor(s) tribune(s) fetus(es)
vicar(s) grill(s) coffee(s) gutter(s) quad(s) tulip(s) backup(s) cancellation(s) moulding(s) weather(s) never(s) cour(s) clench(es)
mytholog**y**(ies) nave(s) slur(s) dictator(s) fragrance(s) orchestration(s) roadblock(s) clove(s) trucker(s) stripper(s) violet(s)
asylum(s) tenor(s) deformit**y**(ies) progression(s) identification(s) primer(s) observance(s) borrowing(s) meme(s) revue(s)
wrangler(s) overrun(s) asshole(s) nerd(s) opener(s) grocer(s) potter**y**(ies) mac(s) veterinarian(s) revive(s) anniversar**y**(ies)
cormorant(s) snipe(s) loa**f**(ves) sash(es) civic(s) phonetic(s) salute(s) vaccination(s) owing(s)

A.1.2. Past

Base form with past inflection modifications
be(was) ha**ve**(d) **be**(were) sa**y**(id) d**o**(id) ma**ke**(de) use(d) bec**ome**(ame) call(ed) t**ake**(ook) **have**('d) f**ind**(ound) release(d)
play(ed) beg**in**(an) le**ave**(ft) h**old**(eld) w**in**(on) locate(d) c**ome**(ame) **go**(went) buil**d**(t) receive(d) serve(d) move(d) start(ed)
ask(ed) die(d) look(ed) g**et**(ot) publish(ed) le**ad**(d) include(d) turn(ed) open(ed) s**ee**(aw) want(ed) los**e**(t) work(ed) g**ive**(ave)
establish(ed) fe**el**(lt) return(ed) kn**ow**(ew) continue(d) t**ell**(old) produce(d) announce(d) join(ed) see(n) create(d) wr**ite**(ote)
describe(d) appear(ed) consider(ed) follow(ed) reach(ed) elect(ed) record(ed) form(ed) sign(ed) finish(ed) develop(ed)
appoint(ed) br**ing**(ought) sen**d**(t) remain(ed) decide(d) s**ell**(old) add(ed) kill(ed) r**un**(an) design(ed) me**et**(t) hear(d)
perform(ed) complete(d) involve(d) tr**y**(ied) direct(ed) report(ed) need(ed) feature(d) change(d) st**and**(ood) replace(d)
provide(d) award(ed) spen**d**(t) live(d) allow(ed) pass(ed) seem(ed) introduce(d) end(ed) raise(d) ke**ep**(pt) enter(ed) attend(ed)
score(d) force(d) defeat(ed) stud**y**(ied) walk(ed) f**all**(ell) represent(ed) note(d) require(d) cover(ed) stop(ped) help(ed)
offer(ed) cause(d) launch(ed) happen(ed) manage(d) r**ise**(ose) carr**y**(ied) increase(d) believe(d) arrive(d) select(ed) fail(ed)
retire(d) operate(d) mean(t) smile(d) support(ed) claim(ed) gr**ow**(ew) ca**tch**(ught) earn(ed) drop(ped) present(ed) refer(red)
agree(d) compose(d) discover(ed) sh**ake**(ook) expect(ed) order(ed) acquire(d) propose(d) graduate(d) promote(d) plan(ned)
remove(d) show(ed) construct(ed) issue(d) watch(ed) intend(ed) reveal(ed) transfer(red) fill(ed) pa**y**(id) recognize(d)
destroy(ed) accept(ed) divide(d) settle(d) participate(d) mention(ed) compare(d) combine(d) t**each**(aught) bur**y**(ied)
suggest(ed) purchase(d) organize(d) conduct(ed) assign(ed) occur(red) capture(d) declare(d) sp**eak**(oke) gain(ed) adopt(ed)
suffer(ed) realize(d) dr**aw**(ew) b**uy**(ought) push(ed) refuse(d) step(ped) pick(ed) estimate(d) stare(d) connect(ed) designate(d)
focus(ed) repl**y**(ied) nominate(d) result(ed) expand(ed) identif**y**(ied) reduce(d) grant(ed) face(d) inspire(d) share(d) laugh(ed)
l**ie**(ay) confirm(ed) la**y**(id) mark(ed) define(d) love(d) attack(ed) succeed(ed) occup**y**(ied) obtain(ed) suppose(d) achieve(d)
arrest(ed) whisper(ed) point(ed) attempt(ed) glance(d) visit(ed) learn(ed) stay(ed) approve(d) invite(d) explain(ed)
abandon(ed) lean(ed) control(led) train(ed) employ(ed) grab(bed) consist(ed) contain(ed) oppose(d) s**eek**(ought) lift(ed)
influence(d) prove(d) convert(ed) derive(d) cho**ose**(se) hire(d) schedule(d) educate(d) contribute(d) admit(ted) cross(ed)
engage(d) register(ed) maintain(ed) miss(ed) observe(d) collect(ed) express(ed) paint(ed) argue(d) remember(ed) regard(ed)
thr**ow**(ew) str**ike**(uck) survive(d) answer(ed) resign(ed) press(ed) merge(d) vote(d) attach(ed) conclude(d) improve(d) fix(ed)

Base form with past inflection modifications
tie(d) rule(d) roll(ed) like(d) surround(ed) wait(ed) charge(d) accuse(d) distribute(d) deliver(ed) praise(d) protect(ed) separate(d) arrange(d) damage(d) accompan**y**(ied) hope(d) threaten(ed) reject(ed) dr**ive**(ove) kiss(ed) approach(ed) dress(ed) mount(ed) wonder(ed) w**ear**(ore) link(ed) enjoy(ed) repeat(ed) b**ind**(ound) restore(d) translate(d) burn(ed) underst**and**(ood) respond(ed) adapt(ed) shrug(ged) wrap(ped) exist(ed) sound(ed) touch(ed) fl**y**(ew) disappear(ed) land(ed) criticize(d) date(d) treat(ed) shift(ed) indicate(d) emerge(d) save(d) act(ed) lock(ed) travel(ed) edit(ed) demand(ed) file(d) attract(ed) decline(d) promise(d) celebrate(d) den**y**(ied) expose(d) gather(ed) modif**y**(ied) cite(d) jump(ed) sponsor(ed) command(ed) execute(d) print(ed) eliminate(d) aim(ed) preserve(d) pause(d) clear(ed) g**et**(otten) portray(ed) talk(ed) dominate(d) fund(ed) recommend(ed) post(ed) recover(ed) comment(ed) suspend(ed) slip(ped) escape(d) donate(d) animate(d) erect(ed) fle**e**(d) view(ed) discuss(ed) cease(d) contest(ed) attribute(d) snap(ped) wish(ed) demolish(ed) display(ed) tour(ed) generate(d) prefer(red) dismiss(ed) withdr**aw**(ew) organise(d) shout(ed) characterize(d) assist(ed) rebuil**d**(t) relocate(d) knock(ed) ignore(d) isolate(d) request(ed) fit(ted) deploy(ed) struggle(d) sail(ed) kick(ed) suspect(ed) initiate(d) trade(d) implement(ed) switch(ed) recall(ed) pitch(ed) rush(ed) cr**y**(ied) decorate(d) climb(ed) demonstrate(d) recognise(d) manufacture(d) lower(ed) travel(led) depart(ed) submit(ted) hate(d) restrict(ed) ban(ned) twist(ed) cancel(led) insist(ed) review(ed) resume(d) fold(ed) inherit(ed) reflect(ed) exhibit(ed) figure(d) test(ed) convict(ed) spot(ted) permit(ted) check(ed) suppl**y**(ied) swe**ep**(pt) warn(ed) orient(ed) collaborate(d) stretch(ed) depict(ed) deem(ed) commence(d) challenge(d) delay(ed) dissolve(d) collapse(d) remind(ed) scream(ed) compile(d) administer(ed) wave(d) scatter(ed) seize(d) trap(ped) disband(ed) relieve(d) revise(d) swallow(ed) update(d) recruit(ed) murmur(ed) document(ed) imprison(ed) perceive(d) prevent(ed) sustain(ed) abolish(ed) question(ed) encounter(ed) transform(ed) block(ed) load(ed) honor(ed) renew(ed) listen(ed) alter(ed) race(d) guide(d) proceed(ed) acknowledge(d) discontinue(d) plant(ed) toss(ed) enable(d) invent(ed) reform(ed) dub(bed) handle(d) accomplish(ed) crash(ed) practice(d) prompt(ed) divorce(d) pronounce(d) examine(d) drag(ged) slam(med) target(ed) absorb(ed) assemble(d) pursue(d) authorize(d) inhabit(ed) fear(ed) shove(d) squeeze(d) quote(d) count(ed) desire(d) brush(ed) reserve(d) possess(ed) imagine(d) advise(d) fade(d) transport(ed) impose(d) wipe(d) deal(t) govern(ed) activate(d) avoid(ed) interpret(ed) w**ake**(oke) chuckle(d) endanger(ed) pack(ed) hesitate(d) assure(d) enhance(d) rest(ed) witness(ed) predict(ed) criticise(d) proclaim(ed) respect(ed) widen(ed) expel(led) curl(ed) interview(ed) specif**y**(ied) render(ed) condemn(ed) trace(d) search(ed) decrease(d) **eat**(ate) flash(ed) tighten(ed) prove(n) revive(d) rescue(d) calculate(d) ar**ise**(ose) investigate(d) renovate(d) dr**y**(ied) breathe(d) explore(d) stain(ed) resolve(d) park(ed) clench(ed) arch(ed) surrender(ed) allocate(d) groan(ed) suck(ed) strip(ped) invade(d) trust(ed) prohibit(ed) embark(ed) import(ed) remark(ed) consume(d) s**ink**(unk) slow(ed) induce(d) contract(ed) emigrate(d) flip(ped) reinforce(d) infect(ed) ruin(ed) match(ed) seal(ed) complain(ed) spell(ed) enclose(d) flood(ed) rip(ped) invest(ed) value(d) forg**et**(ot) welcome(d) detect(ed) bat(ted) explode(d) tuck(ed) adjust(ed) comprise(d) glare(d) project(ed) reunite(d) dispute(d) highlight(ed) exchange(d) aid(ed) precede(d) tend(ed) blame(d) lie(d) endorse(d) reopen(ed) crack(ed) persuade(d) bless(ed) realise(d) crush(ed) anticipate(d) appeal(ed) lease(d) burn(t) enact(ed) protest(ed) regain(ed) double(d) drift(ed) reverse(d) enlarge(d) terminate(d) chart(ed) weaken(ed) transmit(ted) kidnap(ped) smell(ed) greet(ed) retreat(ed) piss(ed) vanish(ed) exclude(d) drain(ed) pave(d) tap(ped) borrow(ed) constitute(d) utilize(d) loan(ed) encode(d) garner(ed) shorten(ed) sue(d) displace(d) plead(ed) confine(d) guarantee(d) attain(ed) dance(d) rent(ed) summon(ed) tip(ped) deserve(d) peer(ed) pop(ped) label(ed) detach(ed) access(ed) migrate(d) cultivate(d) emphasize(d) ease(d) halt(ed) distract(ed) confront(ed) shatter(ed) tug(ged) align(ed) haunt(ed) survey(ed) appreciate(d) evacuate(d) assert(ed) straighten(ed) cancel(ed) annex(ed) amend(ed) fortif**y**(ied) disturb(ed) owe(d) negotiate(d) slap(ped) admire(d) weigh(ed) bow(ed) finance(d) guess(ed) trigger(ed) insert(ed) pin(ned) judge(d) assess(ed) erupt(ed) repair(ed) track(ed) bother(ed) expire(d) scan(ned) bre**ed**(d) clean(ed) solve(d) stroke(d) pose(d) strengthen(ed) rejoin(ed) yank(ed) drown(ed) impl**y**(ied) resist(ed) stress(ed) detain(ed) suppress(ed) spark(ed) lick(ed) tease(d) exile(d) reorganize(d) process(ed) discharge(d) testif**y**(ied) violate(d) sanction(ed) account(ed) exceed(ed) dispatch(ed) anchor(ed) volunteer(ed) reconstruct(ed) chase(d) excavate(d) accumulate(d) fr**y**(ied) haul(ed) presume(d) embrace(d) correct(ed) retrieve(d) moan(ed) whip(ped) speculate(d) reward(ed) accelerate(d) choke(d) postpone(d) devise(d) conceal(ed) confess(ed) doubt(ed) vacate(d) beg(ged) oblige(d) guard(ed) export(ed) pace(d) hook(ed) soak(ed) disperse(d) debate(d) flow(ed) prolong(ed) punish(ed) confer(red) intercept(ed) cool(ed) wander(ed) stir(red) bounce(d) flick(ed) shiver(ed) taste(d) dump(ed) len**d**(t) honour(ed) culminate(d) ensure(d) react(ed) neglect(ed) float(ed) leak(ed) waste(d) compel(led) illuminate(d) cop**y**(ied) extract(ed) evaluate(d) scramble(d) rape(d) root(ed) spill(ed) hunt(ed) wink(ed) intensif**y**(ied) discard(ed) hail(ed) enforce(d) dip(ped) flank(ed) analyze(d) pledge(d) reissue(d) mandate(d) pray(ed) roar(ed) concede(d) clutch(ed) darken(ed) assassinate(d) dream(ed) advertise(d) exit(ed) reprint(ed) grunt(ed) resemble(d) giggle(d) smash(ed) pound(ed) redesign(ed) confiscate(d) forge(d) propel(led) flatten(ed) slump(ed) denounce(d) snatch(ed) thank(ed) duck(ed)

Base form with past inflection modifications
tune(d) substitute(d) deport(ed) overturn(ed) puzzle(d) matter(ed) dismantle(d) pretend(ed) consult(ed) circulate(d)
harden(ed) yield(ed) fulfil(led) elongate(d) compress(ed) bestow(ed) omit(ted) tangle(d) coincide(d) screw(ed) besiege(d)
abuse(d) flourish(ed) label(led) exercise(d) assault(ed) mask(ed) drape(d) sever(ed) reinstate(d) inflict(ed) endure(d)
spawn(ed) delete(d) dare(d) scratch(ed) attest(ed) heal(ed) envision(ed) depend(ed) dart(ed) span(ned) offend(ed) depose(d)
spare(d) stiffen(ed) research(ed) flare(d) exploit(ed) deprive(d) hang(ed) tremble(d) snarl(ed) convene(d) divert(ed) raid(ed)
refurbish(ed) commemorate(d) persist(ed) grasp(ed) picture(d) ensue(d) scowl(ed) fuel(ed) apologize(d) lecture(d) differ(ed)
grimace(d) dislike(d) facilitate(d) inhale(d) pierce(d) cede(d) exhale(d) boil(ed) skip(ped) stack(ed) tow(ed) wrinkle(d)
wreck(ed) provoke(d) deteriorate(d) craft(ed) augment(ed) denote(d) stalk(ed) tempt(ed) categorize(d) conserve(d) poison(ed)
slice(d) liberate(d) prevail(ed) soften(ed) notif**y**(ied) scale(d) implicate(d) plunge(d) disrupt(ed) supplement(ed) improvise(d)
popularize(d) practise(d) batter(ed) spe**ed**(d) mediate(d) fuse(d) poke(d) adorn(ed) function(ed) bark(ed) blush(ed) arouse(d)
grumble(d) prop(ped) verif**y**(ied) experiment(ed) sip(ped) stall(ed) ascend(ed) clarif**y**(ied) intrigue(d) sample(d) intervene(d)
await(ed) leap(ed) inject(ed) repeal(ed) glow(ed) twitch(ed) squint(ed) disclose(d) bolt(ed) blurt(ed) wrestle(d) amass(ed)
inspect(ed) petition(ed) linger(ed) dash(ed) whirl(ed) glaze(d) articulate(d) venture(d) showcase(d) enslave(d) summarize(d)
grit(ted) strap(ped) phase(d) shuffle(d) reign(ed) hamper(ed) rotate(d) cough(ed) subdue(d) convey(ed) fumble(d) flinch(ed)
bruise(d) blast(ed) spoil(ed) chew(ed) revoke(d) surge(d) sacrifice(d) reproduce(d) clash(ed) clap(ped) audition(ed)
harvest(ed) bump(ed) hover(ed) finalize(d) discourage(d) entertain(ed) incur(red) recount(ed) excel(led) muster(ed) lunge(d)
smoke(d) befriend(ed) nudge(d) tumble(d) recycle(d) cure(d) eject(ed) reclaim(ed) deepen(ed) peek(ed) awaken(ed) clamp(ed)
recapture(d) resurrect(ed) correlate(d) penetrate(d) synthesize(d) pinch(ed) alert(ed) afford(ed) mess(ed) rattle(d) blind(ed)
clip(ped) emit(ted) stamp(ed) shave(d) download(ed) encompass(ed) dissipate(d) rall**y**(ied) worship(ped) swirl(ed) signal(ed)
stroll(ed) dictate(d) smack(ed) trim(med) simulate(d) publicize(d) filter(ed) inquire(d) diversif**y**(ied) reappear(ed) evict(ed)
scrape(d) perish(ed) elaborate(d) calm(ed) revere(d) huddle(d) heave(d) stimulate(d) rebel(led) steer(ed) restart(ed) pluck(ed)
trip(ped) repulse(d) dial(ed) strangle(d) compute(d) contemplate(d) caress(ed) cheat(ed) slaughter(ed) poise(d) drill(ed)
communicate(d) peel(ed) defect(ed) muse(d) streak(ed) swell(ed) obscure(d) mock(ed) degrade(d) shriek(ed) recite(d)
synchronize(d) hunch(ed) mobilize(d) radiate(d) relinquish(ed) correspond(ed) breach(ed) splash(ed) spearhead(ed) stock(ed)
rig(ged) manifest(ed) erase(d) cramp(ed) sprawl(ed) disuse(d) retort(ed) glue(d) worsen(ed) escalate(d) loosen(ed) rumble(d)
buzz(ed) tolerate(d) loot(ed) commute(d) lurch(ed) empower(ed) oust(ed) stomp(ed) encircle(d) remodel(ed) fuck(ed)
stem(med) modernize(d) fasten(ed) harass(ed) rake(d) compliment(ed) trick(ed) stabilize(d) boost(ed) traverse(d) log(ged)
tackle(d) cringe(d) claw(ed) plot(ted) accommodate(d) w**ake**(oken) slash(ed) hurl(ed) reconcile(d) resent(ed) forgive(n)
recreate(d) chop(ped) lure(d) humiliate(d) bang(ed) harm(ed) compensate(d) shower(ed) retract(ed) drip(ped) sire(d)
streamline(d) outlaw(ed) complement(ed) spray(ed) hinder(ed) contend(ed) tag(ged) thwart(ed) subside(d) hush(ed)
revisit(ed) gape(d) patrol(led) interfere(d) whistle(d) behave(d) discipline(d) starve(d) detonate(d) flop(ped) censor(ed)
materialize(d) freak(ed) spike(d) bathe(d) assimilate(d) extinguish(ed) swap(ped) shield(ed) customize(d) amplif**y**(ied)
undermine(d) hoist(ed) bellow(ed) falter(ed) repel(led) weld(ed) exert(ed) defer(red) squeal(ed) fool(ed) nail(ed) we**ep**(pt)
bombard(ed) mold(ed) graze(d) despise(d) whimper(ed) lash(ed) taint(ed) relay(ed) topple(d) restructure(d) profess(ed)
emphasise(d) smuggle(d) infer(red) salvage(d) intimidate(d) smear(ed) pardon(ed) ram(med) lengthen(ed) swipe(d)
rewrit**e**(ten) roast(ed) revamp(ed) fare(d) optimize(d) applaud(ed) stifle(d) plunder(ed) dismount(ed) hack(ed) foster(ed)
discount(ed) apprehend(ed) sag(ged) broaden(ed) sharpen(ed) tick(ed) analyse(d) lev**y**(ied) chime(d) buckle(d) glide(d)
massacre(d) sparkle(d) wage(d) torment(ed) hum(med) exempt(ed) rain(ed) mingle(d) fake(d) howl(ed) adore(d) ripple(d)
revolve(d) avert(ed) enlighten(ed) tweet(ed) caution(ed) contradict(ed) wiggle(d) officiate(d) pant(ed) jog(ged) disapprove(d)
scrub(bed) wrench(ed) replicate(d) deflect(ed) chant(ed) reckon(ed) forfeit(ed) moor(ed) vibrate(d) puff(ed) charm(ed)
abort(ed) insure(d) scold(ed) dread(ed) skim(med) brew(ed) dodge(d) esteem(ed) grill(ed) creak(ed) comfort(ed) revolt(ed)
deduce(d) propagate(d) solidif**y**(ied) cherish(ed) adhere(d) thump(ed) relent(ed) bull**y**(ied) lament(ed) hinge(d) wane(d)
roam(ed) rehabilitate(d) whine(d) precipitate(d) herald(ed) seep(ed) def**y**(ied) dive(d) taunt(ed) reestablish(ed) evoke(d)
sp**y**(ied) tickle(d) evaporate(d) shr**ink**(ank) entail(ed) cruise(d) rearrange(d) disintegrate(d) soothe(d) rescind(ed) implant(ed)
chronicle(d) devour(ed) diverge(d) crop(ped) wedge(d) wield(ed) derail(ed) postulate(d) inhibit(ed) quicken(ed) trickle(d)
quiver(ed) hatch(ed) allude(d) rehearse(d) zip(ped) imitate(d) crumble(d) rack(ed) signal(led) mould(ed) tax(ed) waver(ed)
congratulate(d) deceive(d) duplicate(d) purif**y**(ied) apologise(d) worship(ed) skid(ded) refute(d) ridicule(d) hone(d)
snicker(ed) latch(ed) secrete(d) glitter(ed) disarm(ed) lapse(d) boycott(ed) colonize(d) dangle(d) crackle(d) subscribe(d)
snag(ged) confide(d) flatter(ed) sketch(ed) mash(ed) shimmer(ed) legalize(d) gag(ged) recede(d) groom(ed) veer(ed) limp(ed)
prod(ded) fuel(led) churn(ed) disregard(ed) replay(ed) stead**y**(ied) splinter(ed) squat(ted) lessen(ed) bolster(ed) refit(ted)
scout(ed) twirl(ed) moderate(d) chip(ped) interact(ed) tingle(d) stutter(ed) salute(d) warrant(ed) faint(ed) stitch(ed)

Base form with past inflection modifications
hasten(ed) leave(d) repa**y**(id) evade(d) croak(ed) redirect(ed) straddle(d) disembark(ed) discredit(ed) comb(ed)
disseminate(d) rouse(d) signif**y**(ied) pivot(ed) overhaul(ed) shun(ned) rust(ed) memorize(d) wail(ed) nip(ped) reprimand(ed)
grate(d) consent(ed) refrain(ed) regroup(ed) redefine(d) redeem(ed) equate(d) untie(d) splatter(ed) decay(ed) quer**y**(ied)
catalogue(d) buck(ed) consummate(d) patch(ed) elicit(ed) mimic(ked) solicit(ed) swivel(ed) refresh(ed) swoop(ed)
annihilate(d) overlap(ped) knit(ted) coerce(d) plow(ed) stow(ed) bandage(d) recoil(ed) abdicate(d) minimize(d) marvel(ed)
tutor(ed) stray(ed) rap(ped) elude(d) sneak(ed) emulate(d) parade(d) pr**y**(ied) channel(ed) misle**ad**(d) conjure(d) reel(ed)
drum(med) broker(ed) gush(ed) stake(d) wobble(d) aspire(d) disgrace(d) peg(ged) dispense(d) bleach(ed) nurture(d)
indulge(d) clone(d) mourn(ed) stoop(ed) pout(ed) bail(ed) clatter(ed) massage(d) stash(ed) mist**ake**(ook) conjecture(d)
ditch(ed) blindfold(ed) bulge(d) obstruct(ed) reload(ed) burden(ed) crunch(ed) camouflage(d) hike(d) remand(ed) impede(d)
toast(ed) twinkle(d) bristle(d) intersect(ed) lighten(ed) flap(ped) triumph(ed) resonate(d) perk(ed) sprout(ed) refill(ed)
fanc**y**(ied) decompose(d) perpetuate(d) divest(ed) predate(d) reconsider(ed) dredge(d) eradicate(d) rinse(d) paste(d) fray(ed)
incite(d) intrude(d) conjugate(d) splice(d) flirt(ed) hustle(d) rot(ted) spook(ed) instill(ed) impart(ed) ascertain(ed) fiddle(d)
abstain(ed) env**y**(ied) visualize(d) swat(ted) sober(ed) blacklist(ed) cuddle(d) degenerate(d) relish(ed) sabotage(d) wade(d)
saddle(d) ferr**y**(ied) rebuke(d) secede(d) beep(ed) weave(d) ooze(d) itch(ed) revitalize(d) critique(d) herd(ed) cascade(d)
pounce(d) whitewash(ed) blackmail(ed) squash(ed) dine(d) mitigate(d) deter(red) cleave(d) disjoint(ed) outweigh(ed) vent(ed)
gallop(ed) spice(d) delve(d) cushion(ed) deviate(d) dampen(ed) rival(ed) negate(d) lag(ged) profit(ed) replenish(ed)
benefit(ted) discern(ed) neutralize(d) overflow(ed) disobey(ed) rustle(d) catapult(ed) tack(ed) diffuse(d) dissent(ed) vomit(ed)
whack(ed) email(ed) entice(d) revel(ed) tamper(ed) burrow(ed) exterminate(d) harness(ed) barricade(d) spiral(ed)
prioritize(d) regenerate(d) conform(ed) stabilise(d) decode(d) pelt(ed) sizzle(d) stoke(d) scorn(ed) punt(ed) cleanse(d)
dent(ed) vie(d) censure(d) reciprocate(d) misuse(d) videotape(d) rein(ed) paddle(d) buoy(ed) smudge(d) distrust(ed)
quell(ed) dazzle(d) stereotype(d) endeavour(ed) merit(ed) twine(d) lead(ed) excise(d) butcher(ed) dissuade(d) rekindle(d)
decipher(ed) barge(d) reboot(ed) congregate(d) alleviate(d) parachute(d) lumber(ed) keel(ed)

A.1.3. Present Participle

Base form with present participle inflection modifications
be(ing) includ**e**(ing) follow(ing) go(ing) hav**e**(ing) mak**e**(ing) us**e**(ing) work(ing) liv**e**(ing) play(ing) tak**e**(ing) look(ing)
lead(ing) run(ning) try(ing) com**e**(ing) begin(ning) do(ing) win(ning) leav**e**(ing) get(ting) say(ing) mov**e**(ing) start(ing)
stand(ing) becom**e**(ing) giv**e**(ing) hold(ing) serv**e**(ing) feel(ing) remain(ing) think(ing) fight(ing) grow(ing) talk(ing) wait(ing)
act(ing) sit(ting) los**e**(ing) operat**e**(ing) featur**e**(ing) return(ing) fly(ing) watch(ing) surround(ing) see(ing) turn(ing) walk(ing)
speak(ing) scor**e**(ing) provid**e**(ing) allow(ing) reach(ing) call(ing) result(ing) set(ting) perform(ing) exist(ing) creat**e**(ing)
pass(ing) end(ing) kill(ing) wear(ing) tell(ing) support(ing) driv**e**(ing) miss(ing) receiv**e**(ing) sell(ing) increas**e**(ing) keep(ing)
know(ing) contain(ing) caus**e**(ing) regard(ing) finish(ing) show(ing) develop(ing) carry(ing) represent(ing) produc**e**(ing)
join(ing) fall(ing) ris**e**(ing) bring(ing) cross(ing) break(ing) l**ie**(ying) help(ing) consist(ing) ask(ing) cover(ing) dat**e**(ing)
fac**e**(ing) star**e**(ing) chang**e**(ing) put(ting) pull(ing) form(ing) study(ing) enter(ing) continu**e**(ing) add(ing) clos**e**(ing)
depend(ing) rid**e**(ing) visit(ing) defeat(ing) burn(ing) involv**e**(ing) appear(ing) hop**e**(ing) danc**e**(ing) replac**e**(ing) surviv**e**(ing)
beat(ing) seek(ing) compet**e**(ing) graduat**e**(ing) breath**e**(ing) rais**e**(ing) shak**e**(ing) smil**e**(ing) send(ing) hang(ing) eat(ing)
want(ing) sleep(ing) let(ting) consider(ing) defend(ing) listen(ing) claim(ing) d**ie**(ying) cut(ting) tour(ing) rul**e**(ing)
manag**e**(ing) attend(ing) laugh(ing) earn(ing) hid**e**(ing) arriv**e**(ing) push(ing) connect(ing) happen(ing) forc**e**(ing)
establish(ing) attempt(ing) stay(ing) complet**e**(ing) travel(ing) concern(ing) search(ing) belong(ing) pay(ing) report(ing)
deal(ing) promot**e**(ing) throw(ing) describ**e**(ing) refer(ring) contribut**e**(ing) participat**e**(ing) correspond(ing) hit(ting)
wonder(ing) bear(ing) lean(ing) indicat**e**(ing) breed(ing) attack(ing) fail(ing) focus(ing) sav**e**(ing) point(ing) reduc**e**(ing)
bind(ing) command(ing) drop(ping) cry(ing) releas**e**(ing) stop(ping) resid**e**(ing) extend(ing) gain(ing) cit**e**(ing) govern(ing)
maintain(ing) rush(ing) emerg**e**(ing) handl**e**(ing) prepar**e**(ing) struggl**e**(ing) retir**e**(ing) not**e**(ing) strik**e**(ing) climb(ing)
feed(ing) relat**e**(ing) suggest(ing) approach(ing) travel(ling) press(ing) requir**e**(ing) expand(ing) pick(ing) float(ing)

Base form with present participle inflection modifications
scream(ing) believe(ing) kiss(ing) flow(ing) improve(ing) threaten(ing) reveal(ing) protect(ing) expect(ing) jump(ing)
argue(ing) conduct(ing) rest(ing) collect(ing) slide(ing) cast(ing) comprise(ing) remove(ing) skate(ing) love(ing) buy(ing)
occur(ring) direct(ing) spread(ing) mark(ing) lift(ing) measure(ing) demand(ing) oppose(ing) enjoy(ing) control(ling)
realize(ing) destroy(ing) settle(ing) promise(ing) step(ping) explain(ing) match(ing) depict(ing) introduce(ing) ignore(ing)
advance(ing) view(ing) shift(ing) sound(ing) spin(ning) check(ing) avoid(ing) achieve(ing) accompany(ing) prevent(ing)
discuss(ing) organize(ing) present(ing) lay(ing) assume(ing) accept(ing) precede(ing) bleed(ing) reflect(ing) choose(ing)
decide(ing) sink(ing) obtain(ing) owe(ing) count(ing) shine(ing) mix(ing) combine(ing) link(ing) apply(ing) investigate(ing)
block(ing) wish(ing) grab(bing) generate(ing) capture(ing) curl(ing) refuse(ing) pursue(ing) engage(ing) steal(ing) pound(ing)
stretch(ing) question(ing) secure(ing) enable(ing) dig(ging) launch(ing) cool(ing) knock(ing) hurl(ing) identify(ing) dress(ing)
switch(ing) remember(ing) retain(ing) tremble(ing) glance(ing) need(ing) affect(ing) determine(ing) function(ing)
succeed(ing) treat(ing) practice(ing) blow(ing) define(ing) address(ing) deliver(ing) purchase(ing) ensure(ing) shout(ing)
exclude(ing) ensue(ing) celebrate(ing) chase(ing) incorporate(ing) care(ing) compare(ing) order(ing) declare(ing) recover(ing)
decline(ing) examine(ing) acquire(ing) span(ning) announce(ing) ste(ing) charge(ing) experience(ing) demonstrate(ing)
marry(ing) escape(ing) occupy(ing) slip(ping) express(ing) specialize(ing) rotate(ing) pretend(ing) post(ing) wave(ing)
transfer(ring) discover(ing) fold(ing) recruit(ing) target(ing) assist(ing) freeze(ing) drag(ging) aim(ing) originate(ing) flee(ing)
solve(ing) rebuild(ing) observe(ing) review(ing) hurt(ing) scout(ing) prove(ing) limit(ing) hire(ing) praise(ing) undergo(ing)
wake(ing) respond(ing) render(ing) melt(ing) separate(ing) judge(ing) wander(ing) grant(ing) crash(ing) recognize(ing)
implement(ing) issue(ing) comment(ing) ring(ing) signal(ing) beg(ging) flash(ing) pour(ing) fear(ing) attract(ing) lower(ing)
employ(ing) eliminate(ing) preserve(ing) tie(ying) pump(ing) twist(ing) dream(ing) conclude(ing) yell(ing) graze(ing)
sense(ing) construct(ing) brush(ing) mate(ing) log(ging) reign(ing) depart(ing) guide(ing) adopt(ing) weigh(ing)
resemble(ing) stick(ing) descend(ing) pack(ing) divide(ing) invite(ing) consume(ing) convert(ing) expose(ing) pray(ing)
culminate(ing) agree(ing) restore(ing) strengthen(ing) whisper(ing) compose(ing) repeat(ing) crush(ing) decrease(ing)
merge(ing) await(ing) appeal(ing) fade(ing) mount(ing) bend(ing) portray(ing) pierce(ing) differ(ing) suck(ing) drown(ing)
accuse(ing) disappear(ing) chart(ing) utilize(ing) intend(ing) lock(ing) commit(ting) impose(ing) fix(ing) deny(ing) dare(ing)
drift(ing) clutch(ing) tap(ping) wipe(ing) supply(ing) request(ing) boil(ing) drip(ping) exceed(ing) preside(ing) ache(ing)
restructure(ing) transport(ing) arise(ing) linger(ing) retreat(ing) satisfy(ing) worry(ing) invade(ing) concentrate(ing)
project(ing) select(ing) crack(ing) proceed(ing) scatter(ing) rely(ing) dry(ing) preach(ing) like(ing) encode(ing) guard(ing)
grind(ing) slam(ming) plead(ing) remind(ing) cheat(ing) squeeze(ing) negotiate(ing) admit(ting) possess(ing) carve(ing)
mention(ing) prevail(ing) encompass(ing) seem(ing) oversee(ing) advocate(ing) trace(ing) collaborate(ing) confirm(ing)
crawl(ing) inspire(ing) violate(ing) welcome(ing) glare(ing) complain(ing) distribute(ing) pant(ing) widen(ing) blink(ing)
criticize(ing) compile(ing) bet(ting) invest(ing) arrange(ing) research(ing) coordinate(ing) kneel(ing) roar(ing) imagine(ing)
shut(ting) chew(ing) shove(ing) commemorate(ing) highlight(ing) abandon(ing) pertain(ing) swell(ing) overlap(ping)
echo(ing) distinguish(ing) thrive(ing) slow(ing) swirl(ing) imply(ing) award(ing) peer(ing) starve(ing) guess(ing) organise(ing)
ban(ning) bother(ing) forget(ting) analyze(ing) draft(ing) evolve(ing) lick(ing) enhance(ing) rip(ping) stir(ring) protest(ing)
adjust(ing) waste(ing) blaze(ing) honor(ing) prefer(ring) cling(ing) pose(ing) withdraw(ing) creep(ing) signal(ling) drain(ing)
balance(ing) regulate(ing) yield(ing) calculate(ing) inform(ing) absorb(ing) reject(ing) propose(ing) assess(ing) plot(ting)
murder(ing) taste(ing) spill(ing) fulfil(ling) interact(ing) glide(ing) weaken(ing) steam(ing) relax(ing) recall(ing) mock(ing)
aspire(ing) specialise(ing) circulate(ing) exhibit(ing) communicate(ing) tighten(ing) swallow(ing) emphasize(ing)
acknowledge(ing) integrate(ing) exit(ing) dangle(ing) sustain(ing) ascend(ing) debate(ing) injure(ing) supervise(ing)
translate(ing) sweat(ing) bury(ing) insist(ing) cruise(ing) terminate(ing) seal(ing) stalk(ing) raid(ing) rename(ing)
explode(ing) resign(ing) smash(ing) blind(ing) trust(ing) stumble(ing) sob(bing) resist(ing) speed(ing) free(ing) smell(ing)
execute(ing) flirt(ing) escort(ing) scratch(ing) pop(ping) advise(ing) weld(ing) pause(ing) rescue(ing) dump(ing)
evaluate(ing) collapse(ing) brake(ing) grasp(ing) drum(ming) burst(ing) aid(ing) permit(ting) chant(ing) sneak(ing)
radiate(ing) flush(ing) embrace(ing) hum(ming) penetrate(ing) interfere(ing) glitter(ing) adapt(ing) spy(ing) filter(ing)
cough(ing) click(ing) surpass(ing) educate(ing) flank(ing) relocate(ing) strain(ing) witness(ing) seize(ing) register(ing)
commence(ing) blame(ing) influence(ing) forage(ing) facilitate(ing) shatter(ing) poke(ing) hack(ing) prohibit(ing) chat(ting)
upgrade(ing) giggle(ing) slap(ping) scale(ing) rot(ting) leak(ing) sway(ing) assert(ing) soar(ing) gape(ing) repair(ing)
appoint(ing) tumble(ing) unite(ing) detect(ing) tend(ing) spawn(ing) restrict(ing) strip(ping) interpret(ing) intersect(ing)
intervene(ing) pin(ning) sprawl(ing) showcase(ing) exchange(ing) soak(ing) bang(ing) swear(ing) realise(ing) roam(ing)

Base form with present participle inflection modifications
deteriorat**e**(ing) loom(ing) react(ing) lodg**e**(ing) enforc**e**(ing) illustrat**e**(ing) growl(ing) pav**e**(ing) shimmer(ing) go(in’)
migrat**e**(ing) howl(ing) eas**e**(ing) moan(ing) calm(ing) whip(ping) figur**e**(ing) cheer(ing) intimidat**e**(ing) admir**e**(ing)
stem(ming) hat**e**(ing) anticipat**e**(ing) condemn(ing) quot**e**(ing) knit(ting) interview(ing) scrap**e**(ing) sip(ping) tingl**e**(ing)
spit(ting) tow(ing) loot(ing) reopen(ing) regain(ing) arrest(ing) distract(ing) whistl**e**(ing) rain(ing) mentor(ing) trigger(ing)
cop**e**(ing) shield(ing) flourish(ing) reform(ing) administer(ing) crumbl**e**(ing) delay(ing) shuffl**e**(ing) tip(ping) spot(ting)
practis**e**(ing) round(ing) exercis**e**(ing) caress(ing) vibrat**e**(ing) heav**e**(ing) submit(ting) elect(ing) respect(ing) flutter(ing)
resolv**e**(ing) vanish(ing) induc**e**(ing) encounter(ing) rig(ging) access(ing) scrambl**e**(ing) blast(ing) suppress(ing) hop(ping)
restrain(ing) predict(ing) blush(ing) accelerat**e**(ing) shrink(ing) risk(ing) revers**e**(ing) clench(ing) bow(ing) lectur**e**(ing)
assembl**e**(ing) tuck(ing) conced**e**(ing) import(ing) thump(ing) dazzl**e**(ing) thank(ing) decorat**e**(ing) opt(ing) dart(ing)
ruin(ing) export(ing) donat**e**(ing) foster(ing) deploy(ing) rattl**e**(ing) assign(ing) prob**e**(ing) illuminat**e**(ing) straighten(ing)
rent(ing) decay(ing) bypass(ing) churn(ing) rumbl**e**(ing) volunteer(ing) inspect(ing) sacrific**e**(ing) brood(ing) updat**e**(ing)
smooth(ing) assault(ing) twitch(ing) proclaim(ing) reinforc**e**(ing) spray(ing) flick(ing) diminish(ing) bulg**e**(ing) crav**e**(ing)
shudder(ing) clap(ping) contest(ing) offend(ing) contact(ing) activat**e**(ing) betray(ing) surrender(ing) constitut**e**(ing)
chuckl**e**(ing) recommend(ing) irritat**e**(ing) resum**e**(ing) quit(ting) sniff(ing) reward(ing) bud(ding) forg**e**(ing) punish(ing)
prosecut**e**(ing) yank(ing) deepen(ing) reviv**e**(ing) signify(ing) unfold(ing) griev**e**(ing) reel(ing) dub(bing) embark(ing)
plung**e**(ing) thrash(ing) peel(ing) striv**e**(ing) dissent(ing) recognis**e**(ing) halt(ing) summon(ing) forbid(ding) taunt(ing)
reunit**e**(ing) accumulat**e**(ing) flap(ping) peek(ing) tick(ing) extract(ing) combat(ing) accommodat**e**(ing) dissolv**e**(ing)
rustl**e**(ing) insert(ing) invent(ing) straddl**e**(ing) remodel(ing) disrupt(ing) deriv**e**(ing) scar**e**(ing) transition(ing) bloom(ing)
abus**e**(ing) inhabit(ing) hatch(ing) fry(ing) behav**e**(ing) tortur**e**(ing) strid**e**(ing) attach(ing) shrug(ging) estimat**e**(ing)
crouch(ing) retriev**e**(ing) escalat**e**(ing) specify(ing) dodg**e**(ing) reliev**e**(ing) recit**e**(ing) screw(ing) remark(ing) dash(ing)
enclos**e**(ing) splash(ing) wan**e**(ing) stock(ing) apologiz**e**(ing) harass(ing) whirl(ing) worsen(ing) wreck(ing) correct(ing)
stain(ing) consolidat**e**(ing) wad**e**(ing) crackl**e**(ing) clon**e**(ing) compromis**e**(ing) imitat**e**(ing) persuad**e**(ing) inflict(ing) do(in’)
dismantl**e**(ing) inhibit(ing) anchor(ing) shorten(ing) empower(ing) navigat**e**(ing) rejoin(ing) deserv**e**(ing) groom(ing)
conceal(ing) winc**e**(ing) snarl(ing) denounc**e**(ing) undermin**e**(ing) wrench(ing) rippl**e**(ing) travers**e**(ing) slug(ging) seep(ing)
impress(ing) grappl**e**(ing) minimiz**e**(ing) convey(ing) stress(ing) soften(ing) stomp(ing) authoriz**e**(ing) wail(ing) cultivat**e**(ing)
snor**e**(ing) symboliz**e**(ing) chatter(ing) benefit(ing) itch(ing) boost(ing) mistak**e**(ing) twinkl**e**(ing) surg**e**(ing) confess(ing)
nag(ging) fumbl**e**(ing) degrad**e**(ing) overtak**e**(ing) criticis**e**(ing) craft(ing) stall(ing) taper(ing) commut**e**(ing) spar(ring)
slash(ing) overflow(ing) collid**e**(ing) snowboard(ing) boast(ing) nominat**e**(ing) harm(ing) affirm(ing) stifl**e**(ing) reckon(ing)
stroll(ing) bruis**e**(ing) harden(ing) recount(ing) enlist(ing) classify(ing) duck(ing) flar**e**(ing) savor(ing) grat**e**(ing) incit**e**(ing)
fak**e**(ing) defy(ing) designat**e**(ing) rap**e**(ing) stack(ing) skim(ming) liberat**e**(ing) shriek(ing) associat**e**(ing) rov**e**(ing)
stabiliz**e**(ing) decod**e**(ing) suspect(ing) crunch(ing) blur(ring) ponder(ing) meander(ing) nurtur**e**(ing) mumbl**e**(ing) ration(ing)
uphold(ing) revert(ing) endors**e**(ing) flex(ing) download(ing) divert(ing) twirl(ing) ceas**e**(ing) overpower(ing) displac**e**(ing)
amass(ing) tickl**e**(ing) reproduc**e**(ing) breach(ing) cascad**e**(ing) isolat**e**(ing) inject(ing) inherit(ing) erect(ing) renew(ing)
fuel(ing) devour(ing) pry(ing) rehears**e**(ing) slay(ing) allud**e**(ing) tamper(ing) mask(ing) swap(ping) officiat**e**(ing) stamp(ing)
guarantee(ing) squat(ting) enact(ing) smack(ing) adher**e**(ing) clinch(ing) spring(ing) snatch(ing) discharg**e**(ing) fractur**e**(ing)
abid**e**(ing) invok**e**(ing) dissipat**e**(ing) scrub(bing) mingl**e**(ing) shear(ing) formulat**e**(ing) spiral(ing) sag(ging) swarm(ing)
testify(ing) pictur**e**(ing) utilis**e**(ing) roast(ing) shelter(ing) gallop(ing) alert(ing) sketch(ing) babbl**e**(ing) squirm(ing) rak**e**(ing)
ventur**e**(ing) reply(ing) omit(ting) massag**e**(ing) reced**e**(ing) evad**e**(ing) sever(ing) spar**e**(ing) wink(ing) clasp(ing) glaz**e**(ing)
conform(ing) neglect(ing) panic(king) rous**e**(ing) clip(ping) stink(ing) pinch(ing) rewrit**e**(ing) nudg**e**(ing) renovat**e**(ing)
trot(ting) tax(ing) su**e**(ing) suspend(ing) desir**e**(ing) reclaim(ing) doubt(ing) audition(ing) clash(ing) cancel(ling)
propagat**e**(ing) amend(ing) fus**e**(ing) infect(ing) justify(ing) besieg**e**(ing) expel(ling) analys**e**(ing) puff(ing) attribut**e**(ing)
steady(ing) arous**e**(ing) certify(ing) simulat**e**(ing) intensify(ing) revis**e**(ing) succumb(ing) trickl**e**(ing) rebound(ing) rack(ing)
shunt(ing) fling(ing) equip(ping) lov**e**(in’) phras**e**(ing) honour(ing) v**ie**(ying) divorc**e**(ing) shower(ing) worship(ping)
disabl**e**(ing) differentiat**e**(ing) propel(ling) elevat**e**(ing) root(ing) cock(ing) grumbl**e**(ing) stutter(ing) piss(ing) plunder(ing)
contend(ing) trim(ming) deceiv**e**(ing) wag**e**(ing) pledg**e**(ing) fiddl**e**(ing) ram(ming) streak(ing) conserv**e**(ing) puzzl**e**(ing)
lag(ging) motivat**e**(ing) wiggl**e**(ing) marshal(ling) dispens**e**(ing) intercept(ing) dispos**e**(ing) inquir**e**(ing) wither(ing) bust(ing)
squeal(ing) lumber(ing) appreciat**e**(ing) clarify(ing) drool(ing) replicat**e**(ing) disregard(ing) paddl**e**(ing) scoop(ing) eras**e**(ing)
endanger(ing) channel(ing) strangl**e**(ing) furnish(ing) evacuat**e**(ing) peck(ing) erupt(ing) comply(ing) confer(ring) damp(ing)
maximiz**e**(ing) matur**e**(ing) slaughter(ing) obey(ing) resort(ing) speculat**e**(ing) depriv**e**(ing) batter(ing) prop(ping)

Base form with present participle inflection modifications
reconstruct(ing) disclos**e**(ing) convers**e**(ing) obstruct(ing) disarm(ing) aveng**e**(ing) discard(ing) emphasis**e**(ing) allocat**e**(ing) grit(ting) deem(ing) sum(ming) phas**e**(ing) crop(ping) dispatch(ing) coax(ing) flatten(ing) duel(ing) exhal**e**(ing) infiltrat**e**(ing) rav**e**(ing) dread(ing) indulg**e**(ing) scold(ing) revolt(ing) contradict(ing) expir**e**(ing) simplify(ing) gush(ing) diverg**e**(ing) falter(ing) toppl**e**(ing) disseminat**e**(ing) bleach(ing) blossom(ing) utter(ing) damn(ing) sharpen(ing) mitigat**e**(ing) verify(ing) reciprocat**e**(ing) faint(ing) clatter(ing) lighten(ing) swip**e**(ing) slump(ing) summariz**e**(ing) reliv**e**(ing) dispers**e**(ing) torment(ing) sow(ing) fetch(ing) perceiv**e**(ing) upload(ing) mediat**e**(ing) complicat**e**(ing) blackmail(ing) recuperat**e**(ing) snort(ing) relay(ing) disintegrat**e**(ing) bash(ing) snoop(ing) reconcil**e**(ing) provision(ing) oust(ing) strap(ping) bait(ing) rebel(ling) intrud**e**(ing) incur(ring) plow(ing) hinder(ing) brib**e**(ing) madden(ing) complement(ing) tread(ing) memoriz**e**(ing) conjur**e**(ing) relish(ing) twin**e**(ing) hoist(ing) catalog(ing) decompos**e**(ing) discriminat**e**(ing) exert(ing) cancel(ing) daydream(ing) petition(ing) tangl**e**(ing) prescrib**e**(ing) smother(ing) gag(ging) buckl**e**(ing) diagnos**e**(ing) emulat**e**(ing) drap**e**(ing) unravel(ing) trick(ing) impart(ing) confin**e**(ing) ador**e**(ing) interrogat**e**(ing) reload(ing) bug(ging) redeem(ing) alienat**e**(ing) optimiz**e**(ing) supplement(ing) conven**e**(ing) cring**e**(ing) pivot(ing) object(ing) cuddl**e**(ing) purg**e**(ing) leverag**e**(ing) mandat**e**(ing) skid(ding) graft(ing) gaug**e**(ing) distort(ing) squeak(ing) erod**e**(ing) hoard(ing) fasten(ing) congratulat**e**(ing) splinter(ing) zip(ping) replay(ing) disobey(ing) perpetuat**e**(ing) kindl**e**(ing) annex(ing) quak**e**(ing) gossip(ing) reorganiz**e**(ing) ply(ing) extinguish(ing) immigrat**e**(ing) worship(ing) corrupt(ing) quicken(ing) lung**e**(ing) plough(ing) afford(ing) elaborat**e**(ing) manifest(ing) stray(ing) wrinkl**e**(ing) rethink(ing) compost(ing) streamlin**e**(ing) prickl**e**(ing) ced**e**(ing) marvel(ing) sober(ing) consent(ing) masquerad**e**(ing) plug(ging) peep(ing) overhear(ing) detonat**e**(ing) mop(ping) fend(ing) imprison(ing) stunt(ing) spur(ring) revisit(ing) vent(ing)

A.1.4. Third Person

Base form with third person inflection modifications
be(is) ha**ve**(s) do(es) work(s) force(s) point(s) include(s) stud**y**(ies) say(s) show(s) run(s) play(s) make(s) take(s) contain(s) change(s) remain(s) match(es) plan(s) come(s) form(s) use(s) appear(s) consist(s) provide(s) lie(s) leave(s) go(es) race(s) look(s) follow(s) give(s) tell(s) offer(s) seem(s) serve(s) continue(s) become(s) report(s) begin(s) claim(s) figure(s) vote(s) attack(s) get(s) train(s) cover(s) step(s) want(s) set(s) call(s) act(s) sound(s) attempt(s) turn(s) end(s) start(s) find(s) occur(s) refer(s) pass(es) ask(s) win(s) date(s) lead(s) object(s) cause(s) value(s) cost(s) tr**y**(ies) charge(s) return(s) operate(s) measure(s) represent(s) concern(s) face(s) suggest(s) matter(s) practice(s) account(s) require(s) round(s) move(s) park(s) experience(s) process(es) bear(s) feel(s) wave(s) share(s) exist(s) benefit(s) judge(s) write(s) support(s) reach(es) release(s) drum(s) link(s) decide(s) produce(s) stop(s) believe(s) tie(s) reveal(s) help(s) involve(s) increase(s) focus(es) protest(s) explain(s) think(s) demand(s) spot(s) talk(s) visit(s) influence(s) challenge(s) control(s) bring(s) present(s) comment(s) enter(s) answer(s) cut(s) travel(s) signal(s) walk(s) approach(es) sit(s) receive(s) assist(s) arrive(s) hope(s) pound(s) comprise(s) reform(s) ring(s) connect(s) love(s) aim(s) grant(s) display(s) extend(s) broadcast(s) argue(s) ruin(s) keep(s) appeal(s) die(s) manage(s) discover(s) rival(s) create(s) lecture(s) deal(s) maintain(s) depend(s) put(s) count(s) read(s) wish(es) strike(s) mention(s) drive(s) rise(s) estimate(s) deposit(s) smile(s) drink(s) fight(s) refuse(s) agree(s) rebound(s) kiss(es) add(s) fl**y**(ies) roll(s) consider(s) exhibit(s) trace(s) fear(s) send(s) realize(s) graduate(s) reflect(s) beat(s) bat(s) exercise(s) seek(s) tip(s) speed(s) request(s) advance(s) profit(s) promise(s) encounter(s) command(s) struggle(s) reside(s) derive(s) coordinate(s) highlight(s) address(es) develop(s) dance(s) suit(s) pick(s) teach(es) gain(s) lock(s) damage(s) survey(s) stamp(s) appl**y**(ies) finish(es) tend(s) spell(s) regard(s) differ(s) save(s) attribute(s) advocate(s) ride(s) impl**y**(ies) lose(s) chronicle(s) ram(s) peer(s) remark(s) escape(s) stay(s) accept(s) sponsor(s) venture(s) guide(s) shift(s) enable(s) suspect(s) watch(es) employ(s) wear(s) strip(s) survive(s) combine(s) affect(s) arch(es) incorporate(s) reduce(s) compete(s) whisper(s) scream(s) promote(s) delay(s) stick(s) admit(s) debate(s) spark(s) joke(s) feed(s) transfer(s) conclude(s) flood(s) introduce(s) seal(s) span(s) trade(s) touch(es) defeat(s) stretch(es) explore(s) permit(s) raise(s) correspond(s) stroke(s) notice(s) check(s) cr**y**(ies) inform(s) duck(s) repl**y**(ies) steal(s) commune(s) blow(s) dress(es) exchange(s) desire(s) yield(s)

Base form with third person inflection modifications
push(es) contribute(s) hurt(s) breed(s) originate(s) resort(s) demonstrate(s) choose(s) rain(s) glance(s) enjoy(s) confront(s) denote(s) express(es) prevent(s) spend(s) propose(s) close(s) intersect(s) arrest(s) clash(es) pump(s) insist(s) prove(s) exit(s) build(s) switch(es) stare(s) dart(s) care(s) crack(s) worr**y**(ies) participate(s) recall(s) slide(s) fold(s) respond(s) quote(s) fit(s) suffer(s) purchase(s) decrease(s) smell(s) search(es) press(es) constitute(s) rest(s) attract(s) portray(s) respect(s) hurdle(s) drain(s) invite(s) bite(s) wonder(s) convince(s) deliver(s) conduct(s) pin(s) shout(s) revolve(s) upgrade(s) flash(es) taste(s) separate(s) preserve(s) disappear(s) strain(s) protect(s) mount(s) arise(s) filter(s) reject(s) assert(s) register(s) remember(s) declare(s) reward(s) deserve(s) pop(s) punch(es) announce(s) rush(es) pitch(es) prefer(s) treat(s) launch(es) split(s) reeve(s) generate(s) climb(s) lash(es) fill(s) convert(s) hang(s) crash(es) repeat(s) ferr**y**(ies) capture(s) divide(s) bid(s) slip(s) puzzle(s) stall(s) wait(s) abuse(s) echo(es) surround(s) exploit(s) examine(s) bind(s) sum(s) burst(s) collect(s) swing(s) trust(s) lay(s) remove(s) interface(s) ensure(s) bump(s) utilize(s) rall**y**(ies) assault(s) celebrate(s) organize(s) oversee(s) eat(s) guarantee(s) boast(s) fool(s) burrow(s) pose(s) replace(s) spread(s) surprise(s) recount(s) manufacture(s) compare(s) pause(s) transform(s) stress(es) snap(s) assign(s) bow(s) maneuver(s) exceed(s) merit(s) suck(s) shrug(s) realise(s) twist(s) direct(s) cruise(s) paint(s) engage(s) yell(s) illustrate(s) buy(s) advise(s) cast(s) decline(s) stride(s) rescue(s) anchor(s) improve(s) parade(s) insult(s) recommend(s) expand(s) escort(s) prohibit(s) mandate(s) regret(s) prop(s) elect(s) summon(s) bend(s) expect(s) staple(s) collapse(s) brush(es) showcase(s) regulate(s) praise(s) spin(s) crane(s) chuckle(s) float(s) avoid(s) lodge(s) sink(s) fare(s) hire(s) barge(s) cascade(s) investigate(s) acknowledge(s) react(s) accuse(s) batter(s) extract(s) facilitate(s) waiver(s) research(es) rent(s) trigger(s) terminate(s) curse(s) arrange(s) ignore(s) inhabit(s) slice(s) encode(s) cook(s) fracture(s) oppose(s) lease(s) commit(s) drill(s) scan(s) flap(s) lower(s) pace(s) dominate(s) blast(s) tap(s) chase(s) depart(s) hunt(s) giggle(s) blame(s) explode(s) administer(s) complete(s) ban(s) gather(s) toll(s) slam(s) moan(s) adopt(s) earn(s) relay(s) overlook(s) recover(s) stain(s) settle(s) warrant(s) comb(s) blend(s) activate(s) probe(s) interrupt(s) discharge(s) stalk(s) assure(s) strut(s) retreat(s) chant(s) befriend(s) pretend(s) dive(s) clutch(es) awaken(s) scratch(es) substitute(s) bandage(s) induce(s) blind(s) inhibit(s) specialise(s) dye(s) click(s) defend(s) symbolize(s) tug(s) dispatch(es) organise(s) pursue(s) cater(s) scare(s) quer**y**(ies) growl(s) shiver(s) achieve(s) bang(s) ensue(s) recognise(s) predict(s) embrace(s) persuade(s) drag(s) obtain(s) apologize(s) instruct(s) latitude(s) traverse(s) coincide(s) overlap(s) flake(s) fumble(s) concentrate(s) freak(s) expose(s) appoint(s) revolt(s) flip(s) scull(s) free(s) chill(s) exclude(s) alert(s) pat(s) breach(es) compliment(s) bellow(s) attain(s) insert(s) catalogue(s) waste(s) plug(s) balance(s) enhance(s) resume(s) plead(s) rape(s) acquire(s) critique(s) eliminate(s) initiate(s) resolve(s) strive(s) fix(es) comfort(s) welcome(s) collaborate(s) edit(s) toss(es) whistle(s) grunt(s) slow(s) feast(s) whip(s) stumble(s) kidnap(s) mutter(s) reappear(s) bark(s) shut(s) freeze(s) expatriate(s) rotate(s) govern(s) contend(s) wipe(s) breathe(s) lick(s) flop(s) slug(s) bust(s) bull**y**(ies) consume(s) complain(s) wash(es) wander(s) squeeze(s) persist(s) characterize(s) spill(s) gorge(s) scrub(s) reel(s) violate(s) melt(s) convey(s) dio(s) manifest(s) drift(s) dip(s) punt(s) spoil(s) quit(s) brace(s) skate(s) rip(s) analyze(s) overhear(s) predate(s) gag(s) muse(s) forbid(s) mold(s) pour(s) pledge(s) evoke(s) telecast(s) swell(s) poison(s) behave(s) tick(s) reunite(s) emit(s) accommodate(s) safeguard(s) disguise(s) swear(s) precede(s) discount(s) abandon(s) mimic(s) cure(s) taunt(s) hatch(es) sprint(s) tilt(s) calculate(s) harm(s) separatist(s) ache(s) loom(s) replay(s) correlate(s) glare(s) reverse(s) spiral(s) shove(s) complement(s) upset(s) detect(s) bristle(s) greet(s) culminate(s) intervene(s) alter(s) fuse(s) endeavour(s) forgive(s) cease(s) borrow(s) resist(s) lament(s) bypass(es) intrigue(s) solve(s) clam(s) flick(s) restrict(s) evaluate(s) locate(s) perk(s) dial(s) forget(s) stimulate(s) speculate(s) inspire(s) chop(s) converge(s) halt(s) unite(s) spit(s) enlist(s) puff(s) compromise(s) hesitate(s) rap(s) manoeuvre(s) designate(s) spare(s) undertake(s) evolve(s) commence(s) disagree(s) bother(s) guess(es) retrieve(s) expire(s) restore(s) confide(s) betray(s) bounce(s) mumble(s) haunt(s) exclaim(s) approve(s) foster(s) execute(s) shriek(s) weave(s) roar(s) dr**y**(ies) educate(s) heal(s) invest(s) ascend(s) contradict(s) deem(s) infect(s) dissolve(s) surge(s) swim(s) perceive(s) snatch(es) scrape(s) lev**y**(ies) interfere(s) spray(s) admire(s) affirm(s) confer(s) enroll(s) imprint(s) moderate(s) tease(s) adjust(s) clamp(s) forge(s) purge(s) consort(s) swap(s) dare(s) inherit(s) isolate(s) buckle(s) hike(s) howl(s) pray(s) unlock(s) compose(s) torture(s) dash(es) furrow(s) warm(s) caution(s) mock(s) clap(s) preside(s) postulate(s) omit(s) prevail(s) snort(s) crawl(s) remake(s) reproduce(s) manipulate(s) adapt(s) boil(s) fulfill(s) wedge(s) panic(s) chat(s) caress(es) overcome(s) fake(s) reinforce(s) pardon(s) dub(s) steer(s) condemn(s) hurr**y**(ies) devote(s) seize(s) crush(es) flirt(s) flourish(es) supervise(s) cough(s) equate(s) pertain(s) uncover(s) minimize(s) mature(s) surrender(s) rake(s) shovel(s) splash(es) tryout(s) access(es) strengthen(s) dwell(s) accelerate(s) hover(s) glide(s) scoop(s) secure(s) retire(s) correct(s) diverge(s) seduce(s) intercept(s) queue(s) trim(s) appreciate(s) donate(s) pelt(s) consult(s) recite(s) exhale(s) transcend(s) skid(s) sparkle(s) slash(es) deduce(s) revert(s) offset(s) contemplate(s) hack(s) shrink(s) snarl(s) linger(s) conform(s) veer(s) rejoin(s) disrupt(s) poke(s) retort(s) blackmail(s) inquire(s) penetrate(s) perch(es) smack(s) calm(s) elaborate(s) tuck(s) exert(s) enforce(s) flicker(s) shudder(s) assemble(s) plunge(s) simulate(s) ease(s) squeal(s) lapse(s) clean(s) submit(s) scold(s) grill(s) clench(es) distract(s) slur(s) undermine(s) overrun(s) revive(s) snipe(s) salute(s)

A.2. Manual

This word list was generated manually by extracting arbitrary countries from the vocabulary and looking to see if the associated nationality was also present in the vocabulary. We stopped adding words when significant changes stopped being observed in our results.

A.2.1. Countries and Nationalities

Countr with nationality motifications
canad**a**(ian) australia(n) russia(n) ital**y**(ian) norw**ay**(egian) fr**ance**(ench) hungar**y**(ian)
brazil(ian) austria(n) belgi**um**(an) egypt(ian) iran(ian) croatia(n) palestin**e**(ian)
ukrain**e**(ian) romania(n) serbia(n) armenia(n) bulgaria(n) albania(n) nigeria(n)
malaysia(n) hawaii(an) colombia(n) ethiopia(n) bavaria(n) scandinavia(n)

Appendix B

Bit Differences - Unordered

B.1. Plurals

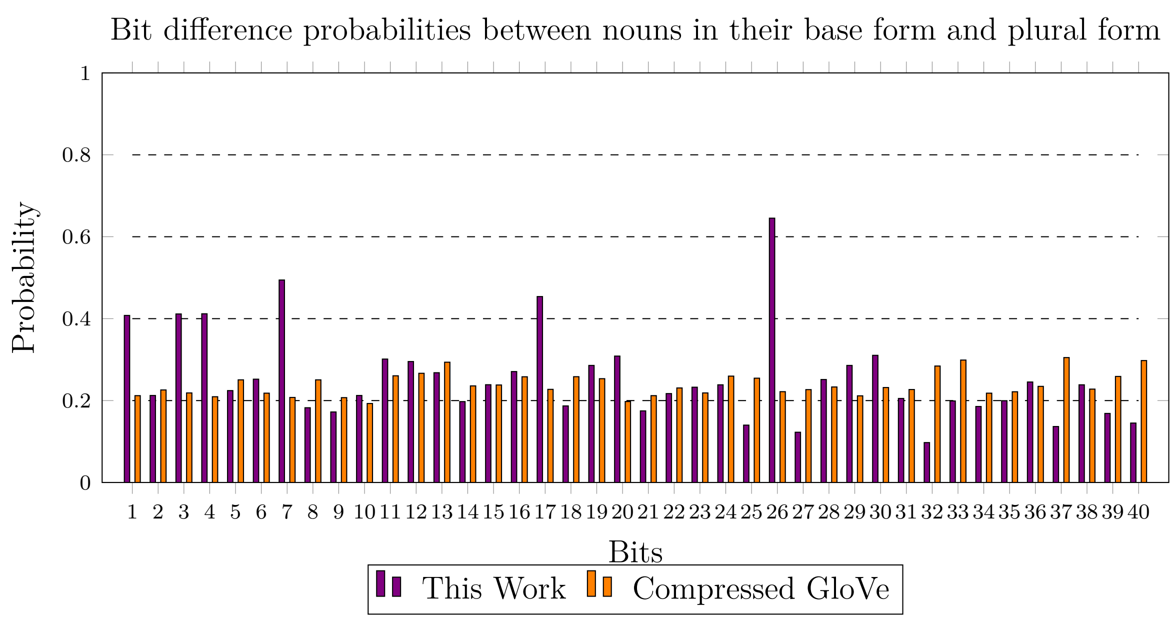

Fig. B.1. The unordered probabilities of a certain bit being different between the embedding of the base form of a noun and the embedding of its plural form. A semi-automatic list of 4948 nouns present in the vocabulary was used (see appendix A.1.1)

B.2. Past

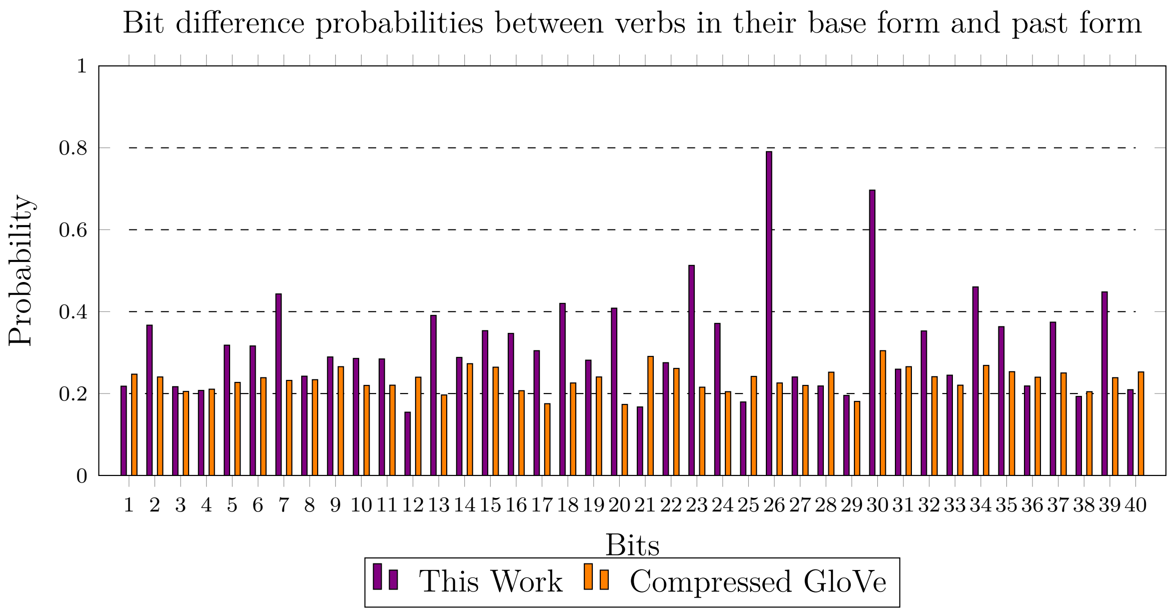

Fig. B.2. The unordered probabilities of a certain bit being different between the embedding of the base form of a verb and the embedding of its past form. A semi-automatic list of 1638 verbs present in the vocabulary was used (see appendix A.1.2)

B.3. Present Participle

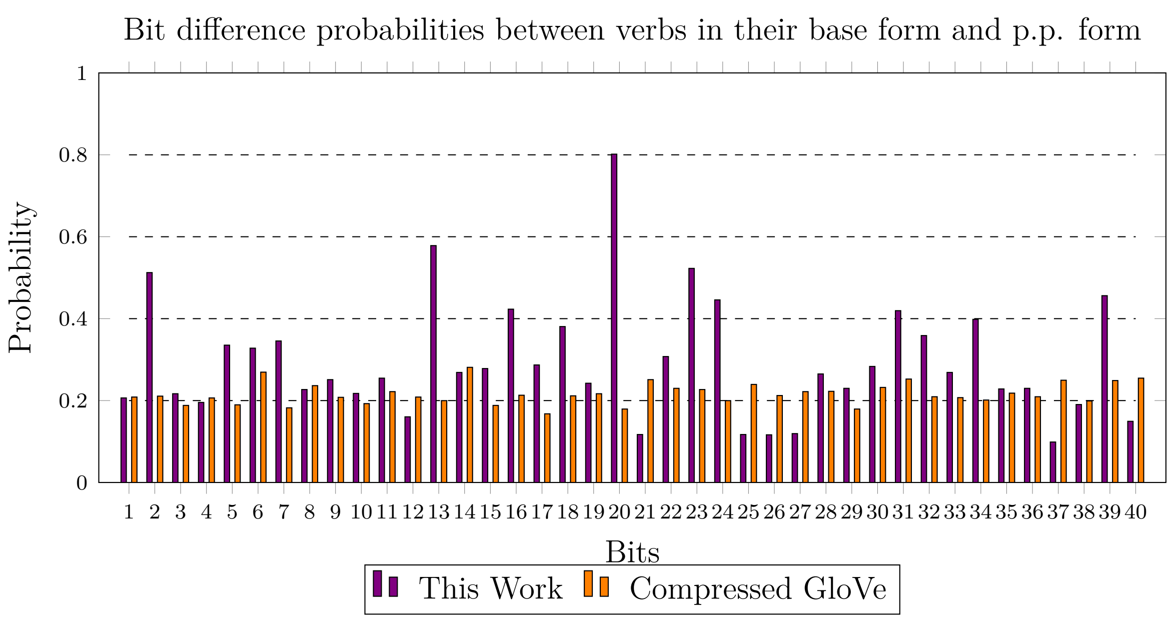

Fig. B.3. The unordered probabilities of a certain bit being different between the embedding of the base form of a verb and the embedding of its present participle form. A semi-automatic list of 1366 verbs present in the vocabulary was used (see appendix A.1.3)

B.4. Third Person

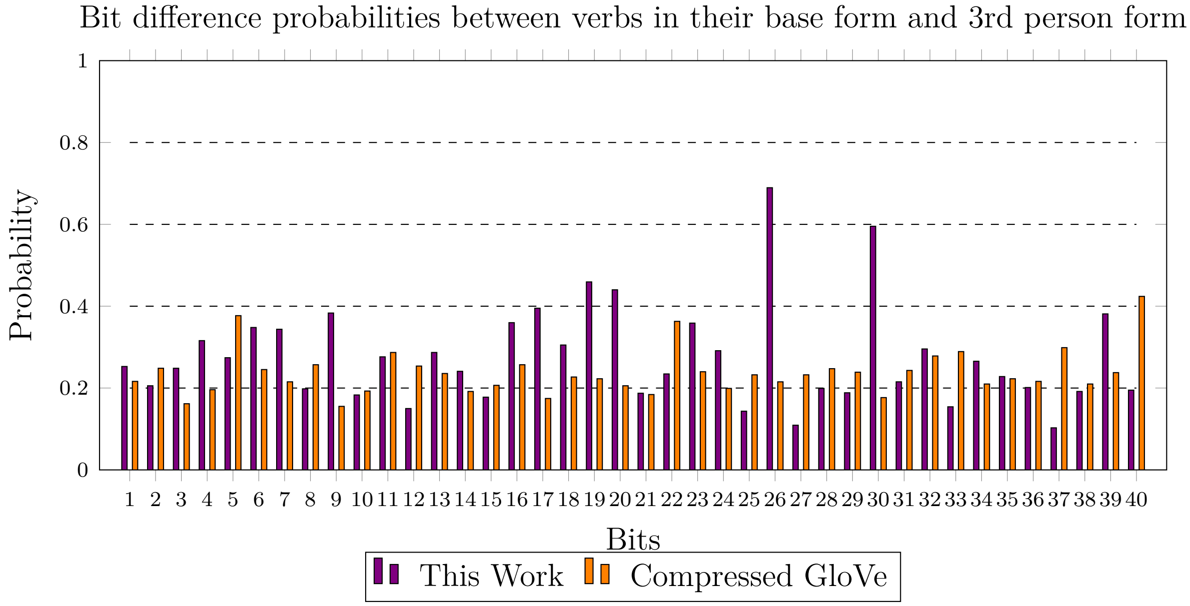

Fig. B.4. The unordered probabilities of a certain bit being different between the embedding of the base form of a verb and the embedding of its third person form. A semi-automatic list of 934 verbs present in the vocabulary was used (see appendix A.1.4)

B.5. Countries and Nationalities

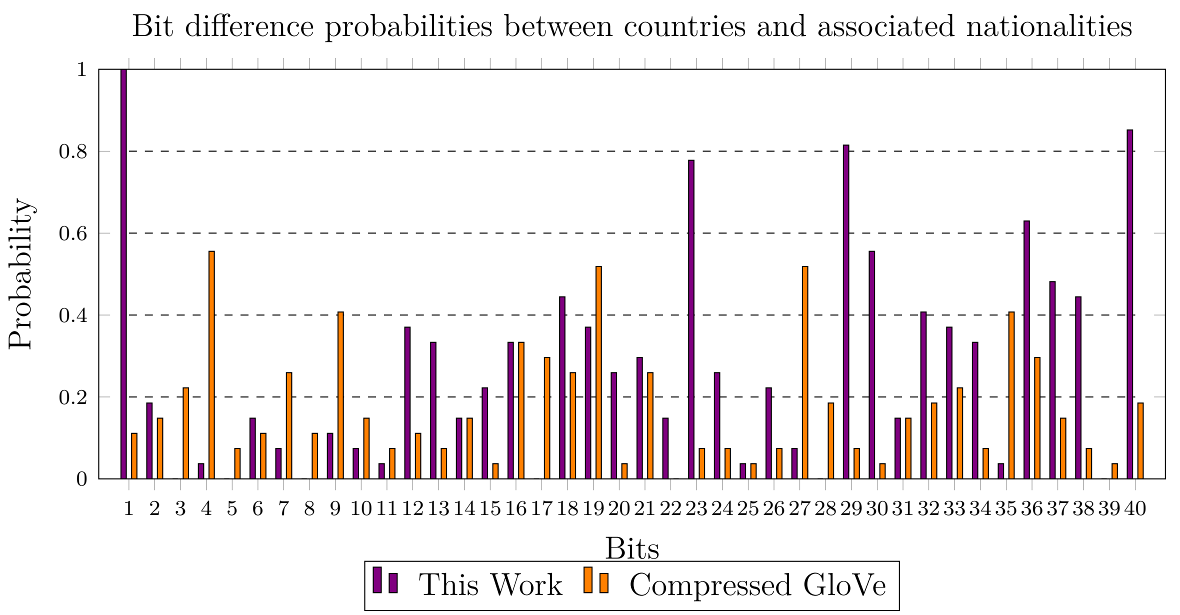

Fig. B.5. he unordered probabilities of a certain bit being different between the embedding of a country and the embedding of its associated nationality. A manual list of 27 countries present in the vocabulary was used (see appendix A.2.1)

Appendix C

Additional Vocabularies

C.1. POS tags

Complete list of POS tags considered					
$	AFX	JJ	NNPS	SP	VBZ
“	CC	JJR	PDT	SYM	WDT
”	CD	JJS	POS	TO	WP
,	DT	LS	PRP	UH	WP$
-LRB-	EX	MD	PRP$	VB	WRB
-RRB-	FW	NFP	RB	VBD	WRB
.	GW	NIL	RBR	VBG	XX
:	HYPH	NN	RBS	VBN	_SP
ADD	IN	NNP	RP	VBP	NNS

C.2. WordNet Lexical Categories

Complete list of WordNet lexical categories			
adj.all	noun.event	noun.quantity	verb.consumption
adj.pert	noun.feeling	noun.relation	verb.contact
adj.ppl	noun.food	noun.shape	verb.creation
adv.all	noun.group	noun.state	verb.emotion
cntlist	noun.location	noun.substance	verb.Framestext
noun.act	noun.motive	noun.time	verb.motion
noun.animal	noun.object	noun.Tops	verb.perception
noun.artifact	noun.person	verb.body	verb.possession
noun.attribute	noun.phenomenon	verb.change	verb.social
noun.body	noun.plant	verb.cognition	verb.stative
noun.cognition	noun.possession	verb.communication	verb.weather
noun.communication	noun.process	verb.competition	

Appendix D

WordNet Training Words

D.1. noun.animal

Words from the *noun.animal* lexical group matching our vocabulary
world down head game hand young royal style blue hair female bay primary queen jack test male wing martin baby horse host tom emperor grade foot fish jersey scale cricket rail dog cancer throat ray fox adult grey bear bass bird turkey fly web flag investment arizona dam wolf giant lincoln morgan kid billy cat soldier moth argentina relative coat dragon shell tail ass hide welsh assault pan admiral lip runner robin jay layer hampshire sole cattle copper monster carrier eagle spider worker lucy tiger bull beetle cardinal lion drum breast pen poll sierra virus seal snake argentine bat hood hart horn drake prey egg sheep monitor deer chicken devon mouse belly armor ram hybrid dock permit brush pastor cannon fisher durham fur butterfly kit vector rod bitch swift dug spine pet chen tang bacteria duck boxer gastropod elbow amazon circus snail colt hawk rat raja bee livestock coral frog crest beard raven valve buck monarch roller swan fin rabbit elephant shark cow flank newfoundland salmon monkey citation crane crow tooth mollusk swallow wool survivor dove shire cock falcon cerambycidae pad mare pike omaha gill beef bug venus drill antenna vein whale saddle trout beaver arabian lamb turtle knot insect blackburn owl goat goose weaver secretariat ant lizard pomeranian vanessa racer sawyer stud predator skate stunt dinosaur panther mutant chestnut worm kitty snout chick sire spat penguin claw feather pony finch thoroughbred hare zeus cobra tick jaguar canary stray leopard phoebe mara hobby mole cornish cod pigeon wren maltese hind bunny viceroy elk affirmed crab orphan fang dolphin muzzle puppy greyhound drone vent lama squirrel ling shrimp camel peacock calf horseback python stallion doe pest wasp feeder mustang otis sparrow dominique roe parasite hereford flicker bison pembroke nanny viper kitten kangaroo galloway siren mule penelope roach beak ridley herring oyster otter perch liza mosquito rana hack filly racehorse kite coyote crocodile talon lynx wolverine mammal catfish pointer scorpion geometridae parrot noctuidae heron hound grunt ara guernsey eel fledgling slug boar pathogen sable bullock comb argus paw cub homo ox badger gorilla mammoth pinto tuna chow primate tortricidae partridge ape reeve helix hen sponge springer embryo nightingale ala pollard rattle vane mane spawn cisco lark squid sleeper octopus rudd gecko hackney rhea trumpeter primates brit maverick forewing sonora gnu fetus larva vulture flea brant hornet passerine rook carp hog sucker thorax lobster moray pup cochin chihuahua hydra ayrshire antelope emu manx newt canine rodent dobson cougar aix gelechiidae marlin wildcat caterpillar bulldog cuckoo alligator char gull zebra condor tortoise vertebrate reptile salamander weasel grizzly stag sturgeon redhead dun pelican woodpecker whiting castor quill bot hedgehog medusa warbler nestor guan grasshopper nit cocoon brahman terrier caribou jackal locust shrew boa uta cyclops barb charger swine silvia kelp dragonfly parr triton uma chameleon coney seaweed grenadier mya cairn nymph finisher

Words from the *noun.animal* lexical group matching our vocabulary
peregrine leech mantis raccoon drosophila jellyfish withers waterfowl apis chinook mite ani cardigan teal clam gar schoolmaster mink beagle rhinoceros operculum kiwi passer longhorn firefly porcupine cheetah feline stifle marten falco quail kingfisher carabidae magpie gibbon tusk amphibian fawn arista skunk asp stork lhasa hummingbird columba broodmare trotter equine hoof watchdog selene starling milt rattlesnake tench osprey galleria carapace brill sow bongo arca whiff canis ibis gazelle agua grouse booby flamingo albatross ruff bunting hippocampus lycaenidae bantam buzzard tabby foal harrier stingray megachile blackbird thrush bronco carrion streptomyces slider weevil pheasant comma loon starfish ostrich snapper plankton steed lepidoptera vixen bluebird invertebrate nymphalidae lemur plasmodium giraffe chrysalis bumblebee tern dalmatian capuchin skink chimpanzee scavenger gelding smelt atticus walrus bighorn collie fluke scarab recombinant ferret mussel goldfish solitaire aquila sus tuft gopher crayfish mackerel pollack bobcat snipe cirrus nematode poodle diploid bovineshad roan darter mako margate roebuck mollie mongoose nutcracker coronet impala yolk dingo sloth halter capra gander anaconda whitefish midge hamster draco reticulum woodcock kestrel shiner mallard maja conch bos proteus stinger roadrunner setter basset jaeger cyanobacteria fossa petrel phytoplankton manta theropod hyena vermin cockroach gnat bacillus anguilla bivalve suckling wishbone flathead tachinidae plover harpy takin moa cob tarantula ewe creeper mullet termite haddock bristle kea mockingbird yearling loach gallus swordfish vole saki grub bulbul siphon herpes cormorant cockatoo tentacle thrasher teg manatee armadillo yak chlorophyll gosling pupa barracuda dory sweeper sunfish instar megachilidae merino griffon cotswold sphingidae retriever spitz shrike cichlid anas tetra walleye sula springbok sepia pseudomonas dodo placental adder seabird wallaby hock cygnus ermine salmonella largemouth koala pacer zooplankton mew songbird spaniel piranha chub escherichia chipmunk proboscis conspecific minnow humpback goby pug mus curculionidae hippopotamus haploid opossum llama lioness parakeet damselfly corvus staphylococcus irena centipede

D.2. noun.body

Words from the *noun.body* lexical group matching our vocabulary
a back part down head area right left side small hand system face body center heart middle person process hair blood mouth bridge wall arm structure b feature temple skin eye shoulder neck ball foot joint zone throat brain leg crown imperial bob wave quick chamber finger stomach gene nose tongue ear hip o plate ridge lap palm cheek passage vessel column index lip bone chin lock jaw knee organ flesh waist milk sole blade costa muscle root tube breast poll skull arch axis landmark tissue matrix radiation wrist tear trap fist belly thumb humor rod vegetation pot nerve brow lens pin elbow receptor membrane specimen bang pupil eyebrow ab lung beard liver valve os ankle bypass juice plaque thigh ink fold radius tooth iris fiber atlas plasma cock tract nucleus toe partition cone facial vein beaver heel kidney hemisphere corpus node abdomen gore apparatus marcel nail mantle shin groove chromosome pulp attic cortex limb nipple digit sac cavity sperm torso capsule artery afro antagonist hormone forearm keel hooks nasal calf penis rib serum abdominal lobe groin bulb palatine roach beak corona mohawk colon buff ponytail insulin flap pouch tensor socket stump ligament bladder scalp sheath duct gland whitehead stigma tunic appendix mummy queue marrow nape lambda bile ala mane atrium chop braid xx enamel cleavage saliva thorax spindle dorsum tendon canine pore chopper fascia vagina mons isthmus stubble adhesion mustache biceps secretion pelvis cartilage pectoral crotch intestine suture vestibule smear extremity hamstring retina molar sinus germ buttocks knuckle allele agonist uterus palate haircut cytoplasm collagen lymph graft semen filament navel bloodstream carina countenance genitalia ovary femur cunt xxx bosom fingertip capillary diaphragm shank tibia phalanx spleen arse hippocampus pons anus mucus chromatin haw pituitary aster epithelium dimple pate wattle suppressor hairline tarsus earlobe girdle fissure perspiration cheekbone lumen pancreas placenta fingernail fuzz clot beehive bursa plexus cusp peduncle vertebra clitoris thatch drool blister eyelid synapse pacemaker cornea ventricle axon reticulum epidermis sternum esophagus eyeball humerus snot aorta tiptoe amygdala ganglion foramen centrum septum underbelly lamina shunt gyrus rima hypothalamus cerebellum dentition venter goatee fibula larynx ribosome follicle exoskeleton cuticle rectum modifier trachea cadaver thalamus proboscis armpit vesicle sulcus orifice thumbnail

D.3. noun.time

Words from the *noun.time* lexical group matching our vocabulary
time over then may school while year now second season day life game end go march old september january june october july august april november december age century set february top death times history night point days half run period summer present today term past middle week leave moment festival turn beginning hours dead deep future morning date fall month youth chapter majority speed distance winter pass rate breath hour era spring prime window rule minute value watch sleep quarter evening birth bottom none generation christmas tonight bell sunday phase flow extension saturday decade anniversary afternoon cycle weekend tomorrow childhood reign occupation shift birthday frequency depression friday restoration absence holiday occasion pace flower grave eve minority tenure monday dawn calendar attendance midnight stretch trick renaissance presidency meal span reconstruction delay yesterday pulse duration wednesday interim thursday lease pause tuesday tide millennium baroque bout ab easter assumption sunset vacation snap threshold wartime overtime noon leisure inning shiva prohibition cease advent teens jubilee drought tempo tertiary jerk 1870s semester bloom deadline regency halloween 1860s lent twilight eternity centennial float 1900s thanksgiving 1850s incarnation antiquity maturity weekday acceleration cretaceous preseason termination wee 1840s halftime monsoon probation 1830s jurassic flux adulthood ascension twenties cradle sabbath honeymoon rag morrow immortality limitation infancy 9/11 octave pleistocene sixties continuum epoch caliphate miocene afterlife 1820s olympiad triassic seventies fifties thirties adolescence annum cambrian blackout eighties spacing peacetime permian eocene fortnight kip devonian nineties curfew lapse forties puberty turnaround interlude epiphany carboniferous incubation bicentennial 24/7 solstice lunchtime sabbatical gestation quaternary ramadan deathbed pinpoint boyhood respite bedtime ordovician enlistment moratorium prehistory annunciation hertz pliocene circumcision passover equinox holocene episcopate 1790s tiebreaker mesozoic silurian pentecost oligocene allegro paleozoic 1780s novitiate hospitalization checkout mississippian apogee precambrian 1770s airspeed 1760s tet eon cenozoic vespers trimester retardation transfiguration weeknight interregnum midterm paleocene clocking 1750s downtime sivan midweek embolism payday workday menopause

D.4. verb.body

Words from the *verb.body* lexical group matching our vocabulary
have make water play look help cup round act give feel call rest attack contract hold cut smile cover sun train break spring doctor hurt draw sleep entrance drug shirt labor bob wave condition nick tone pull dress drop trouble gain catch grow wear roll reduce laugh wet stream pale shock wake twin coat pack waste correct cry hat expect grin exercise blow leaf shake affect stare jacket shower nurse sigh treat beam soap comfort stretch harm bat nod shed breathe tear sweat breed joke torture faint suffer costume tense boot powder wash tan relax twist habit poison shoe hatch alter crane highlight soup heal frown gown lamb tire roar rouge barber freeze bundle shrug spit marcel exhaust slick blink relieve rack shiver wan stool spat gum chuckle clown razor smirk strap scrub manipulate pod vest perfume choke blush fracture groom litter wink giggle gag martyr remedy reproduce sob administer cough scarf tease nap kitten evacuate twitch bray hack bleed stimulate menace scowl massage purge bonnet comb cub brood lipstick overdose spawn puff vomit prank bandage stale snort infect grimace shave dung awaken vet tumble emit sniff perm clap soothe conceive pant wince hamstring shampoo inhale dope exhale yawn propagate optimize bathe presume sneer disable wail suction relapse inject fawn resurrect regenerate stub swagger refresh cripple pout injure undress foal corset abort animate dimple snuff sprout atrophy prim wanton squint crick cleanse recuperate salve farrow cramp secrete douche irritate quicken fart reek quack hock chloroform frock sneeze snore sedate snicker

D.5. verb.consumption

Words from the *verb.consumption* lexical group matching our vocabulary
use go want play power need put board give hit range provide meet mine port drive serve fast horse host answer pop tax content receive drug carry trip wine drop drink address stomach eat tank wolf feed lap grass waste smoke enjoy cry kick lunch milk breakfast sample shower mess prey accommodate diet spare afford corn feast strain bolt raven sip swallow quarry drain delight fare picnic suck patronage gorge sustain consume toast exploit harness peck digest mainline cater chew avail huff plank smack crunch puff champ forage addict gulp spree exert quell dine dunk indulge inhale dope devour fodder starve gutter partake inject crave mumble ply browse abstain recycle brunch nibble repose overuse