# Computational Lexical Semantics
# Final Assignment

Thuong-Hai Pham

January 27, 2017

## 1 Question 1

The Distributional Hypothesis dated back to the discussion of Harris [Harris, 1954] about meaning as a function of distribution when he observed that,

> [I]f we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution.

Having that observation and transposition rule, the similarity between the distributions of (A, B) and (A, C) leads to the similarity in meaning of those considered words (or morphemes).

Please note that the hypothesis stated above is also called Weak Distributional Hypothesis, in which it only assumes the correlation between semantic content and contextual distributions. A stronger version of the Distributional Hypothesis which take into account the assumption of causal role in the creation of semantic content is not covered in this report as underlying idea of Distributional Semantics Models (DSMs).

## 2 Question 2

Although lexical output of DSMs and lexical semantic content of WordNet both generate lists of semantically similar words, the difference between those two school are still evident. While DSMs results in a ranked list of similar words (by using a specific score function), manually built lexical resource such as such as WordNet does not provide the ranking system. However, these resources clearly declare the relation labels between words (e.g. synonym, antonym, hypernym...) which are missing in DSMs.

The main differences between those two approaches are listed in table 1 below

| | Manually built (WordNet...) | DSM output |
|---|---|---|
| Pros | Precise<br>Well-defined relations | Quantifiable semantic relatedness<br>Empirical, cheap to construct<br>Domain-independent<br>Language-independent<br>Flexible<br>Scalable, depending on the corpus size |
| Cons | Expensive to be manually built<br>Domain-dependent characterisations<br>No frequency information<br>Binary-relation (related {synonym, antonym...} or not)<br>Static resource | Approximative<br>Unexplainable relation label |

Table 1: Advantages and drawbacks of DSMs output and manually built lexical resources

# 3 Question 3

The monolingual text goes with DISSECT provides us lemmas for both words and contexts.

Using lemma instead of word (after tokenisation) will prevent our co-occurrence matrix getting sparse. Hence, it will improve the latter process which is matrix decomposition or dimension reduction. More importantly, our ultimate goal is to analyse the underlying meaning of words, not the morphological process or meaning aspects produced by the morphological process. Therefore, choosing lemma to be our experiment subject is a better choice in compare with morphological words.

# 4 Question 4

To do the lemmatisation and POS tagging for our bilingual corpus, please follow the instruction to set up the environment in README.md, then run the following code:

```
bash ./data_get.sh
bash ./data_preprocess.sh
```

These two bash file will help you to download all required data from `http://opus.lingfil.uu.se/download.php?f=OpenSubtitles2016\%2Fen-vi.txt.zip`. Then data_preprocess will do tokenisation, lemmatisation and POS tagging as follow:

| | English | Vietnamese |
|---|---|---|
| Tokenisation | TreebankWordTokenizer[1] in nltk library | vnTokenizer [Lê Hng Phng et al., 2008] for Apache Spark™ |
| POS tagging | Stanford POS Tagger [Toutanova et al., 2003] | vnTagger [Le-Hong et al., 2010] for Apache Spark™ |
| Lemmatisation | WordNetLemmatizer[2] in nltk library which finds lemma for each words in noun, verb, adjective categories (content words we aim to get). The POS information is taken from the previous phase. | Vietnamese words are not going under any morphological process, so lemmatisation is not needed |
| Contractions | Hand-defined contractions found in corpus: 's/v: is, 're/v: are, 'm/v: am, 've/v: have, ma'am/n: madam | no contraction |

Table 2: Bilingual corpus tokenisation, lemmatisation and POS tagging

For the POS tagging, at first, pos_tag in nltk was used. However, this default PerceptronTagger is purely based on probability, so it made a lot of mistakes, such as:

```
gamma/N :/. none/N of/ADP your/PRON *mailman/A* friend/N can/V
hear/V you/PRON ./.
every/DET time/N you/PRON say/V ``/. *mailman/A* ,/. ''/. jay/N
,/. i/PRON 'm/V just/ADV gon/V na/PRT hit/V you/PRON ./.
```

Pos_tag is mistaken in assigning *mailman* as adjective because it stands before nouns. While this is solved using Stanford tagger.

```
gamma/N :/: none/N of/IN your/PRP$ *mailman/N* friend/N can/MD
hear/V you/PRP ./.
every/DT time/N you/PRP say/V ``/`` *mailman/N* ,/, ''/'' jay/N
,/, i/PRP 'm/V just/RB gon/V na/TO hit/V you/PRP ./.
```

[1] http://www.nltk.org/api/nltk.tokenize.html#module-nltk.tokenize.treebank

[2] http://www.nltk.org/_modules/nltk/stem/wordnet.html#WordNetLemmatizer.lemmatize

# 5 Question 5

Without doubt, POS information helps us to distinguish words which are identical in spelling but have distinct meanings - the linguistic phenomenon of homonymy.

For examples, in DISSECT provided data, we have:

| Word | POS | Meaning[1] |
|---|---|---|
| present-j | adj. | (something presented as a gift) ”his tie was a present from his wife” |
| present-n | noun | (give an exhibition of to an interested audience) ”She shows her dogs frequently”; ”We will demo the new software in Washington” |
| present-v | verb | (temporal sense; intermediate between past and future; now existing or happening or in consideration) ”the present leader”; ”articles for present use”; ”the present topic”; ”the present system”; ”present observations” |

Table 3: Advantages and drawbacks of DSMs output and manually built lexical resources

It is clearly that those three words listed in 3 above all represented by “present”. However, these words read totally different meanings with their corresponding categories.

# 6 Question 6

Broadly speaking, features used in distributional semantics model can be categorised, at first, by how we construct the co-occurrence matrix. There are mainly two types: word-document and word-word matrix. The former is more common in data retrieval or web search. While the latter can be varied in term of context (the finite set of words which can be considered to have semantic relation with the current target word) in which we construct the word-word matrix:

[1] First sense for ”present” in each of its categories from Wordnet http://wordnetweb.princeton.edu/

- Window with fixed size of k consequent words, target word included. In this type of context, there are these parameters that needed to be taken care of:
  - Size k of the window.
  - The position of the target word in this sequence (left-only and right-only windows are specific cases).
  - Whether words within the window are weighted by their distance to target word or not.
- Linguistic boundaries such as sentences, paragraphs, words in the whole documents.
- Words with syntactic relation:
  - Adjective and its modified noun
  - Verb with its subject and/or object. In this case, a high-dimensional SVD can be applied to decompose the 3-D matrix.
  - Dependency relations extracted from dependency parsing.
- Aligned words in parallel corpora

In addition, there are some other feature type that we can take into consideration such as visual feature... For examples, words (red, oval...) that are used to described a picture can somehow related to visual features extracted from that picture.

# 7 Question 7

In short, the bigger the size of the window is, the looser (more general) our semantic relation be.

Let us make this clear by exaggerating the situation. For example, if we have the window size approximate our document size, the context of every single word are (almost) all of the words in that documents. Hence, it seems to be this context denoting the topic of that documents rather than meaning of that word itself.

# 8 Question 8

In my personal view, assuming that alignment is done with high accuracy, bilingual corpus is expected to propose higher percentage of synonyms than monolingual corpus. This is because of our "window size" in the bilingual corpus is exactly the word that it is aligned with. Hence, the context is more strictly limited.

# 9 Question 9

To extract a sparse co-occurrence matrix from the parallel aligned files. Please run:

```
python ./data_convert.py ./bilingual_data/en-vi.txt
        ./bilingual_data/en-vi.align
        ./bilingual_data/en-vi.rows
        ./bilingual_data/en-vi.cols
        ./bilingual_data/en-vi.sm
```

or

```
bash run_bilingual.sh
```

to run all the tasks from download data file to the DISSECT-formatted matrix extraction.

| | | |
|---|---|---|
| everyone/n | ơi/n | 5 |
| everyone/n | kể_cả/n | 1 |
| everyone/n | i/n | 44 |
| everyone/n | tất_cả/n | 370 |
| everyone/n | tâ/n | 9 |
| everyone/n | bất_cứ/n | 11 |
| everyone/n | khác/a | 17 |
| everyone/n | tập_trung/v | 9 |
| everyone/n | như/a | 3 |
| everyone/n | quý_khách/n | 2 |
| everyone/n | khác/n | 1 |
| everyone/n | bẫy/n | 1 |
| everyone/n | kích_cỡ/n | 1 |
| everyone/n | cu/n | 1 |
| everyone/n | thế_giới/n | 5 |
| everyone/n | dĩ_nhiên/a | 1 |
| everyone/n | đều/a | 10 |
| everyone/n | chú_ý/v | 3 |
| everyone/n | t/v | 7 |
| everyone/n | mo/n | 61 |

Figure 1: First 20 lines of the sparse co-occurrence matrix extracted from bilingual corpus

Before extraction, because our bilingual corpora are only sentence-aligned, not word-aligned. Hence, fast_align [Dyer et al., 2013] is used to do the word alignment over 100 iteration.

```
log_e likelihood: -7.94319e+07
log_2 likelihood: -1.14596e+08
cross entropy: 6.33762
perplexity: 80.8748
posterior p0: 0
posterior al-feat: 0
size counts: 3680
```

# 10 Question 10

For feature weighting scheme, we have these functions to be taken into account:

- Vector normalisation which normalises the co-occurrence count into probability distribution (over row or column)
- Term Frequency - Inverse Document Frequency (TF-IDF) in case of constructing word-document matrix

$$tfidf(f,d) = P(w|d) * log\left(\frac{|D|}{|d \in D : w \in d|}\right)$$

- Point-wise Mutual Information (PMI), in which t, f are target word, feature word respectively

$$PMI(t,f) = log\left(\frac{P(t,f)}{P(t)P(f)}\right)$$

- Exponential Point-wise Mutual Information (EPMI)

$$EPMI(t,f) = \frac{P(t,f)}{P(t)P(f)}$$

- Positive Point-wise Mutual Information

$$PPMI(t,f) = max(PMI(t,f), 0)$$

- PMI with contextual discounting [Pantel and Lin, 2002], let X be the co-occurrence matrix

$$cdPMI(t,f) = PMI(t,f) * \frac{X_{tf}}{X_{tf}+1} * \frac{min(\sum_{k=1}^{m} X_{kf}, \sum_{k=1}^{n} X_{tk})}{min(\sum_{k=1}^{m} X_{kf}, \sum_{k=1}^{n} X_{tk})+1}$$

- Positive Local Mutual Information (PLMI)

$$PLMI(t,f) = PPMI(t,f) * X_{t,f}$$

- t-test

$$\frac{p(t,f) - p(t)p(f)}{\sqrt{p(t)p(f)}}$$

- Positive Log Weighting (Plog)

$$Plog(t,f) = max(0, log(t,f))$$

By using raw frequencies without a weighting scheme, we are overrating meaningless words. In TF-IDF perspective, for example, "to be" (after lemmatisation) is a very common word and may occur in the context of many words whereas this does not count much into the target word meaning. In general, word that appears in less context carries more amount of meaning that those appear all over the corpus. TF-IDF helps to prevent these common words to be overrated during our calculation. These other weighting functions do the same thing.

# 11 Question 11

DISSECT comes with 4 weighting schemes: Positive Point-wise Mutual Information (PPMI), Positive Local Mutual Information (PLMI), Exponential Point-wise Mutual Information (EPMI) and Positive Log Weighting (Plog). This library also provides 4 similarity functions:

- Dot product

$$sim(\vec{u}, \vec{v}) = \vec{u}.\vec{v} = \sum_i u_i v_i$$

- Cosine similarity

$$sim(\vec{u}, \vec{v}) = \frac{\vec{u}.\vec{v}}{\sqrt{||\vec{u}||.||\vec{v}||}}$$

- Euclidean similarity

$$sim(\vec{u}, \vec{v}) = \frac{1}{||\vec{u} - \vec{v}|| + 1}$$

- Lin similarity

$$sim(\vec{u}, \vec{v}) = \frac{\sum_{i \in I}(u_i + v_i)}{\sum_i u_i + \sum_i v_i}, I = \{i | u_i > 0, v_i > 0\}$$

In this report, PPMI for weighting is chosen because it is mentioned to be robust by Baroni et al. About the similarity, cosine similarity will do the work, as obviously, direction is more important than distance in this vector space.

# 12 Question 12

| **Bilingual** | **Monolingual** |
|---|---|
| Top 5 highest frequency | |
| be/v | child-n |
| 1. do/v 0.644864852611 | 1. parent-n 0.917077574336 |
| 2. have/v 0.53856205691 | 2. adult-n 0.782701972496 |
| 3. seem/v 0.501341250193 | 3. age-v 0.766753507122 |
| 4. suppose/v 0.434208274879 | 4. mother-n 0.697207331389 |
| 5. appear/v 0.425135594284 | 5. care-v 0.672120216415 |
| do/v | company-n |
| 1. be/v 0.644864852611 | 1. firm-n 0.858734913616 |
| 2. job/n 0.482768427598 | 2. sale-n 0.703226771638 |
| 3. miranda/n 0.443403144377 | 3. share-n 0.692699749645 |
| 4. work/v 0.419430517203 | 4. finance-n 0.679246440312 |
| 5. happen/v 0.415200117797 | 5. asset-n 0.673388675912 |
| have/v | government-n |
| 1. be/v 0.53856205691 | 1. reform-n 0.858794556597 |
| 2. suppose/v 0.463524216234 | 2. opposition-n 0.812902399853 |
| 3. do/v 0.381634314346 | 3. official-n 0.790417186791 |
| 4. get/v 0.370700718953 | 4. citizen-n 0.783790421439 |
| 5. guess/v 0.370605030256 | 5. minister-n 0.775312049209 |

know/v

1. aware/a 0.703021663021
2. realize/v 0.671213997158
3. understand/v 0.63907783525
4. know/n 0.599892302029
5. knowledge/n 0.55807791604

get/v

1. pick/v 0.559988392181
2. take/v 0.539727747568
3. catch/v 0.528943028352
4. grab/v 0.470867631315
5. bring/v 0.465093376048

Top 5 lowest frequency

honor/v

1. honor/n 0.914345975167
2. respect/n 0.627452925122
3. proud/a 0.580979180612
4. pride/n 0.55512987187
5. pleasure/n 0.518680112168

cause/n

1. reason/n 0.609092817925
2. cause/v 0.565175858232
3. purpose/n 0.516723494415
4. excuse/n 0.326777208454
5. regret/v 0.306312326903

man-n

1. boy-n 0.808836412112
2. nothing-n 0.801337325917
3. suppose-v 0.797160387784
4. realize-v 0.79553290897
5. recall-v 0.792342694109

interest-n

1. debt-n 0.689492158541
2. asset-n 0.67486902297
3. finance-n 0.664138447874
4. investment-n 0.627399311431
5. fund-n 0.607052624799

flinch-v

1. face-n 0.634137640313
2. turn-v 0.594026517471
3. moment-n 0.589972389237
4. throw-v 0.589120941417
5. hand-n 0.586821526683

digress-v

1. sort-n 0.530770049145
2. suppose-v 0.506682152561
3. something-n 0.505936566787
4. seem-v 0.50282571827
5. wonder-v 0.500758937318

gut/n

1. courage/n 0.805742297034
2. brave/v 0.723577063421
3. dare/v 0.618285158884
4. capable/a 0.52971806174
5. ball/n 0.459042016931

friday/n

1. tomorrow/n 0.580693672328
2. yesterday/n 0.508509729613
3. sunday/n 0.500100539258
4. night/n 0.497346989695
5. tonight/n 0.491611345723

criminal/a

1. crime/n 0.887793115624
2. guilty/a 0.630346289689
3. robbery/n 0.605219555932
4. offense/n 0.568078418005
5. sin/n 0.55002560594

ricochet-v

1. shot-n 0.607339659037
2. ball-n 0.582938290069
3. hit-v 0.570853829087
4. strike-v 0.569048093428
5. throw-v 0.556888195528

slouch-v

1. sit-v 0.525704745576
2. back-n 0.500733637246
3. shoulder-n 0.499656537184
4. face-n 0.4974346316
5. notice-v 0.496175792746

gabble-v

1. talk-v 0.410311235872
2. wonder-v 0.401321902108
3. moment-n 0.398461557888
4. hear-v 0.392450811479
5. listen-v 0.387854584271

As stated above, running our DSM on bilingual corpora tends to produce higher percentage of synonyms than monolingual corpus. For example, "know/v" in bilingual corpora goes with "aware", "realise", "understand"... while "government-n" in monolingual corpus goes with "reform", "opposition", "official", "citizen"... It is obvious that monolingual corpus propose more topic-related words.

In comparison between high-frequency (HF) and low-frequency (LF) words. Related words for HF words tend to spread broader in meaning than LF words in both corpora. For instance, "be/v' or "do/v" in bilingual have a broader range of relatedness because in Vietnamese, these two words can be substituted by a wide range of words depending on context. The same range happens to "child-n" as well. On the other hand, "gut/n" in bilingual corpora and "ricochet-v" in monolingual corpus seem to have more closely related words.

# 13 Question 13

In brief, contextual variability stands for a problem or phenomenon that the meaning we assign for a particular word can be varied depending on the context where that word occurs. Let us take the word "gut/n" of our bilingual corpora in the list above as an example, its meaning varies from "intestine", "belly" to "courage" or "determination". However, because of the nature of film subtitles, most of its usage belong to the latter case, in which the related words are listed above.

The difference can be more subtle. For instance, "I will go to Washington next year" vs. "Washington declared their stand on the problem". In these sentences, the former "Washington" is the capital of the United States, while the latter is the U.S. government.

For polysemous word, let us choose the famous "bank-n".

1. bank-n 1.0
2. cash-n 0.637879447974
3. finance-n 0.553212523761
4. debt-n 0.552867349563
5. money-n 0.549037117541
6. account-n 0.542203646392
7. fund-n 0.536145442912
8. credit-n 0.529048101638
9. payment-n 0.510477460579
10. purchase-n 0.507018840735
11. asset-n 0.505840055635
12. capital-n 0.497043488885
13. investment-n 0.491719961594

14. river-n 0.486462985351
15. pay-v 0.479063629788
16. loan-n 0.476691468646
17. company-n 0.474462371703
18. invest-v 0.466238846801
19. firm-n 0.465827345318
20. exchange-n 0.460229621276

In the list of 19 related words above (except bank-n itself), most of those are about banking and finance (which is one of its sense). Fortunately, the $14^{th}$ word denotes its relation with river (in river bank). This happen because of the dominance of finance-related bank in the corpus, which crawled from newspapers and encyclopedia. The same situation occurs to our "gut/n" in bilingual corpora. Because of the nature of those corpora (film subtitles), "gut/n" denotes a kind of human virtue or mental instance instead of a part of human body.

In case of homonyms, let us meet our acquaintances "present-v" and "present-n" from question 5.

| **present-n** | **present-v** |
|---|---|
| 1. gift-n 0.518967672286 | 1. illustrate-v 0.693506315963 |
| 2. living-n 0.488776654395 | 2. presentation-n 0.691251540248 |
| 3. couple-n 0.484307429417 | 3. writing-n 0.597135409141 |
| 4. husband-n 0.44554892667 | 4. perspective-n 0.577479002089 |
| 5. household-n 0.414977623657 | 5. interpretation-n 0.573946360157 |
| 6. marry-v 0.40744417453 | 6. theme-n 0.567569668364 |
| 7. occasion-n 0.404849265528 | 7. describe-v 0.564004311533 |
| 8. wedding-n 0.355775154089 | 8. draw-v 0.563206571465 |
| 9. past-n 0.333187680494 | 9. discussion-n 0.552639114303 |
| 10. celebrate-v 0.331177180356 | 10. examine-v 0.54609084858 |

In this case, we can see clearly the difference between “present-n” as “gift” and “present-v” as “to illustrate”. This also reconfirms the advantages of using POS information we have discussed in question 5.

# 14 Question 14

According to the principle of compositionality, which is also called Frege’s principle, the meaning of a compound is determined by the meanings of its elements and the rules used to combine them. For example, the world “computer mouse” is constructed from “computer” and “mouse”, so its meaning should relates to these two words. Moreover, the way “computer” stands before “mouse” denotes that the main meaning is “mouse”, and “computer” specifies this kind of “mouse”, keeps it separated from others.

To be adapted to compositionality, there are several ways. One simple solution is to treat these compound word as a single units (which is somehow already solve in Vietnamese tokenisation). The other way is to perceive compositionality as a function. This approach is implemented by a various sets of functions: additive, full additive, weighted additive, multiplicative, lexical, full lexical, dilation...

# 15 Question 15

DISSECT provides us with full and weighted additive, multiplicative, dilation and lexical function. Hence, we choose lexical function to perform our tasks. This justification is based on a work of compositional DSM evaluation from Dinu et al. [Dinu et al., 2013]. In their explanation, lexical function works pretty well because it considers linguistic relations between words (verb-object, verb-subject or modification relations...).

To re-run the experiment, please use:

```
bash run_comp.sh
```

Our outputs from the command above are:

| | **Argument space** | **Observed phrase space** |
|---|---|---|

| | | |
|---|---|---|
| ball-n ricochet-v | 1. shot-n 0.60090820919<br>2. ball-n 0.555601914406<br>3. ricochet-v 0.549424934767<br>4. hit-v 0.512803151047<br>5. yard-n 0.486374884256<br>6. flick-v 0.483794412282<br>7. fire-v 0.463375541489<br>8. shoot-v 0.45805234635<br>9. drop-v 0.450721566791<br>10. strike-v 0.437304242189 | 1. ball-n_ricochet-v 1.0<br>2. shot-n_ricochet-v 0.943434619602<br>3. ball-n_rebound-v 0.469537442791<br>4. share-n_ricochet-v 0.417981793027<br>5. hand-n_kick-v 0.390882784654<br>6. optimism-n_ricochet-v 0.374202443085<br>7. head-n_shudder-v 0.323837743214<br>8. rifle-n_recoil-v 0.308846129507<br>9. screen-n_flick-v 0.307706335726<br>10. eye-n_flare-v 0.303568463715 |

| | | |
|---|---|---|
| vein-n pulse-v | 1. entrance-n 0.224424352834<br>2. tower-n 0.200311482731<br>3. paint-v 0.199348815366<br>4. hill-n 0.193616827569<br>5. road-n 0.193117821902<br>6. wall-n 0.188580909673<br>7. floor-n 0.186548341616<br>8. west-n 0.182714679164<br>9. stone-n 0.180126459453<br>10. north-n 0.1797992173 | 1. vein-n_pulse-v 1.0<br>2. voice-n_pulse-v 0.318874524154<br>3. export-n_thunder-v 0.199271833659<br>4. face-n_beam-v 0.198910407124<br>5. cigar-n_glow-v 0.169324448505<br>6. head-n_stoop-v 0.168784963436<br>7. child-n_roam-v 0.162331076115<br>8. cigar-n_burn-v 0.155946374982<br>9. fountain-n_erupt-v 0.139816664857<br>10. fire-n_beam-v 0.139524551862 |

It is clear that in the argument space, the neighbours are both related to one of the two words made up our compound word. While in the observed phrase space, the neighbours are those similar in both meanings of each constituent and meaning constructed by combining all of its constituents.

# 16 Question 16

We can not create a compositional model for our bilingual corpora as the way we did in question 15. The compositional models in DISSECT require observed phrase corpus to be extracted with linguistic relations (subject-verb in case of DISSECT data). Whereas, the relation of pairs in bilingual data is alignment between different language corpora, not within the corpus itself.

# 17 Question 17

In plain, prototype theory states that some members of a category are more central than others. Hence, it creates a graded categorisation for that category by "distance" to its central one, or prototype. By that theory, meaning of concept/object in lexical semantics can be expressed not by some well-defined features but its similarity in some of its features.

For example, if we defines the word "bird" for animal with "wing" and "ability to fly", "eagle" is clearly a hyponym of "bird", "chicken" is to be considered, while "ostrich" or "penguin" will be out due to their lack of needed feature(s). However, in prototype theory, "eagle", by all "bird" requirements, can be chosen as "prototype", and "chicken" or "ostrich" and "penguin" can be in the graded categorisation system (somehow farther than "pigeon") due to its similarity to the prototype. Consequently, the hyponymous relations in these cases are also constructed via this graded categorisation.

# 18 Question 18

Let compute the neighbours for "knife/n" in our bilingual corpora:

1. knife/n 1.0
2. blade/n 0.846837062401
3. sword/n 0.757256934822
4. throat/n 0.46765543662
5. cut/v 0.421440081426
6. cut/n 0.411966698543
7. stick/v 0.398805264072
8. wood/n 0.392601946913
9. stick/n 0.383138407667
10. tree/n 0.359358089655
11. tongue/n 0.349288518585

12. bite/v 0.339242334869
13. finger/n 0.312515569684
14. back/n 0.308099197181
15. wound/n 0.298108472021
16. side/n 0.271557471912
17. teeth/n 0.269856045383
18. skill/n 0.265075315182
19. arrow/n 0.255360859053
20. iron/n 0.252328951782

It is obvious that "blade", "sword" are closer to "knife" in the category of "personal hand-held ancient close-combat weapons". Those which are in outer level of this category can also be found as "stick" (not so harmful), teeth (not hand-held), arrow (not for close combat) with less score. Therefore, DSM seems to cater for typicality effect.

# References


[Dinu et al., 2013] Dinu, G. et al. (2013). General estimation and evaluation of compositional distributional semantic models.

[Dyer et al., 2013] Dyer, C., Chahuneau, V., and Smith, N. A. (2013). A simple, fast, and effective reparameterization of ibm model 2. Association for Computational Linguistics.

[Harris, 1954] Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3):146–162.

[Le-Hong et al., 2010] Le-Hong, P., Roussanaly, A., Nguyen, T. M. H., and Rossignol, M. (2010). An empirical study of maximum entropy approach for part-of-speech tagging of vietnamese texts. In *Traitement Automatique des Langues Naturelles-TALN 2010*, page 12.

[Lê Hng Phng et al., 2008] Lê Hng Phng, N. T. M., Huyn, A. R., and Vinh, H. T. (2008). A hybrid approach to word segmentation of vietnamese texts. In *Proceedings of the 2nd International Conference on Language and Automata Theory and Applications*.

[Pantel and Lin, 2002] Pantel, P. and Lin, D. (2002). Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 613–619. ACM.

[Toutanova et al., 2003] Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 173–180. Association for Computational Linguistics.