Literature Review

Edwin Zhou

Papers

Shutova et al. 2010 - Metaphor Identification Using Verb and Noun Clustering

https://www.semanticscholar.org/paper/Metaphor-Identification-Using-Verb-and-Noun-Shutova-Sun/6cf384ccea2943e059bd937f306fa601d72e3618

- Big idea
  - Detect metaphors in unrestricted text genres and with mostly unsupervised methods, by clustering verbs and nouns starting from a seeded set of metaphors, and then using those clusters to predict
  - These clusters model source-target mappings
- Abstract
  - Metaphor detection in **unrestricted** text (independent of genre of text). Scope of this paper is entirety of British National Corpus (BNC)
  - Starts with small, seeded, manually annotated set of metaphors, and gathers metaphors with similar syntactic structure
  - First to apply unsupervised methods for metaphor detection
  - Precision of 0.79
- Introduction
  - Metaphors are **mappings** between two conceptual domains: **source and target**. I.e., a metaphor is when one concept is viewed in terms of the properties of another (Lakoff and Johnson 1980)
  - Across many genres, a majority of metaphoricity are introduced by a verb
  - Metaphors represent a violation of **selectional restrictions** in a given context (Wilks 1978)
    - This tends to overclassify metaphors by detecting all kinds of non-literalness, for example metonymies unless using **hand-coded knowledge**
  - This paper differs from previous approaches in that it both applies to unrestricted text independent of domain or type of discourse, but is also unsupervised and does not rely on hand-crafted knowledge
  - **Target concepts** with same **source concept** (essentially the metaphor) should appear in similar lexico-syntactic environments (vocabulary and structure)
  - Since unsupervised, metaphors evaluated by human judges
- Conclusion

    - Although currently tested with verb-subject and verb-object metaphors, believe the techniques can be applied on a wider range of syntactic constructions
    - One possible limitation: seed-dependent
    - Future would could include more diverse seed set of common metaphor mappings
    - Captured many metaphors not directly related to those in seed set
-
- Why/Context
    - Knowledge deficit/conflict
        - Previous research for metaphor detection were either supervised or restricted to a specific domain
- What/Contribution
    - Clustering motivation
        - Possible to induce semantic word classes by clustering of contextual cues
        - Consensus is that lexical items with similar behaviour in a **large body of text** most likely have the same meaning, but **we say** that they can also be clustered together by **association with the same source domain**
- How/Approach
    - Data
        - Seed phrases
            - 62 phrases, taken from BNC sampled over a wide variety of genres
            - Single-word metaphors
            - Verb-subject and verb-object relations
        - Corpus
            - BNC parsed using RASP parser, which gave grammatical relations output
            - Corpus is searched for source/target domain vocabulary within a particular grammatical relation (verb-object or verb-subject)
        - Verb dataset is subset of VerbNet, extracted from raw corpus data, and including only verbs that appear more than 150 times
        - Noun dataset is 2000 most frequent nouns in BNC
    - Method
        - System first implicitly captures source-target associations underneath a few verb-subject/verb-object relations, then uses unsupervised verb clustering to expand this knowledge. These clusters can then be used to detect new metaphorical expressions
        - Feature extraction
            - Uses method proposed by Sun and Korhonen (2009)
            - Extracts syntactic and semantic features using shallow parser, normalizes, then applies clustering method suitable for high-dimensional feature space
            - Verb feature extraction

        - Verb features obtained by SCF acquisition system of Preiss et al. (2007)
            - System extracts SCFs by running rule-based classifier on GRs tagged by RASP parser
        - Obtain SPs by clustering “argument heads” in subject and object slots of the resulting SCFs
      - GRs used as features for noun clustering
    - Clustering algorithm: SPEC is used for both verbs and nouns
      - MNCut algorithm used for SPEC
      - SPEC takes **similarity matrix** as input
      - We construct similarity matrix using JSD as a measure, where similarities are viewed as weights in a graph
    - The noun clusters represent target concepts that we expect to be associated with the same source concept
    - The verb clusters contain coherent lists of source domain vocabulary
    - Selectional preference
      - Metaphor is violation of selectional restrictions
      - Not all verbs have strong capacity to constrain their arguments, therefore those ones are not as prone to metaphoricity. This criterion used to filter out those verbs with weak SP
      - SP classes formed based on SPEC clusters, with strength measure proposed by Resnik (1993)
- Discussion
  - Results/Conclusions
    - Evaluated against baseline of using WordNet **synsets**, also expanded on seed set to represent source/target domains, with help of human judges
    - Found that baseline synsets covers only 13% of the data identified using clustering, and does not go beyond concepts present in the seed set
      - Clustering allows us to discover concepts beyond just synonicity and thus tag novel
    - Human evaluation
      - Human agreement was 0.63 in terms of k
      - Precision of system was 0.79, while baseline was 0.44
  - Weaknesses
    - Only tested system on verb-subject and verb-object metaphors
    - Approach is seed-dependent
  - Takeaways
    - Lexical clustering in large body of text is useful for deducing **association with same source domain** of items in cluster. This could be useful for my research as VUAMC contains many large bodies of text
  - Other thoughts
    - Source/target domains are kind of like word parent topics

        - Would be interesting to do this first before doing using a supervised method
        - I feel like even though precision is okay, comparison with baseline is misleading because it has never been proposed as a valid system for metaphor detection outside of this paper, with a central idea being the same, in that some groups of words are constructed and each group represents “target words” mapped to the same “source domain”. Not only that, sysnet considers some metaphors to be synonyms, which would result in overtagging. The baseline seems like a poor strawman setup to beat. Would be nice to see it compared with other metaphor detection systems, supervised or unsupervised
        - Interestingly, there was no large metaphor-annotated corpus available at the time
    - Additional questions
        - How are SCFs parameterized by SPs when SC is syntactic and SP is semantic? (3.1.1)
        - The clustering algorithm is beyond my understanding currently. Will come back to review
        - How are noun clusters linked with verb clusters to detect metaphors? Does it consider all combinations before filtering them? This seems more like a metaphor generation method with a BNC dataset rather than identification/detection
            - Method could be used for novel metaphor generation?
    - Further research
        - Extend system to wider range of syntactic constructions, beyond verb-subject and verb-object metaphors
        - Creation of more diverse seet set representing whole variety of common metaphorical mappings
            - Could start with Master Metaphor List (Lakoff et al., 1991)

Turney et al. 2011 - Literal and Metaphorical Sense Identification through

https://www.semanticscholar.org/paper/Literal-and-Metaphorical-Sense-Identification-and-Turney-Neuman/cb0d80684acc65c6a6b7d291cb5a0e70b9317899

- Why
    - Past work metaphor detection as classical word sense disambiguation task
    - Metaphor as a method for transferring knowledge from concrete domain to more abstract domain (Lakoff and Johnson 1980)
    - Metaphor transfers associations from source to target domain (2003)
- What
    - Use of metaphors is correlated with the degree of abstractness of the word’s context

    - The abstractness rating algorithm is used to generate feature vectors from a word's context and training data is used to learn a logistic regression model that relates degrees of abstractness to the classes literal and metaphorical
- How
    - Dataset
        - One hundred adjective-noun phrases
        - TroFi (Trope Finder) Example Base of literal and nonliteral usage for fifty verbs, occurring in 3737 sentences from WSJ, with two types of models:
            - Separate model learned for each verb
            - Single model learned on all verbs
    - Classifier
        - Uses logistic regression implemented in Weka
    - Features
        - Abstractness ratings generated from word's context
        - A vector of five features for each sentence
            - Average abstractness ratings of all nouns, excluding proper nouns
            - Average abstractness ratings of all proper nouns
            - Average abstractness ratings of all verbs, excluding target verb
            - Average abstractness ratings of all adjectives
            - Average abstractness ratings of all adverbs
- Discussion
    - Results
        - 79% accuracy for adjective-noun phrases
    - Strengths
        - Readily generalizes to new words not present in training data
    - Weaknesses
        - Evaluated on adjective-noun phrases and TroFi example base only, unclear of effectiveness on running texts
    - Takeaways
        - Abstractness can be valuable component in system for metaphor detection

Dunn 2013 - What metaphor identification systems can tell us about metaphor-in-language

https://www.semanticscholar.org/paper/What-metaphor-identification-systems-can-tell-us-Dunn/8587fa3050e1f54d80fe93f16d077a5c01a058f0

- What
    - Evaluate four metaphor identification systems on the 200,000 word VU Amsterdam Metaphor Corpus
    - Compare results and rate of agreement between the systems between genres and sub-classes

- Why
    - Want to see what success or failure of metaphor systems can tell us about the essential linguistic properties of metaphor-in-language
- How
    - Dataset
        - VU Amsterdam Metaphor Corpus
    - Four metaphor identification systems
        - Using semantic similarity
        - Using word abstractness
        - Using source-target mappings
        - Using domain interactions
- Discussion
    - Results
        - Success of the identification systems varies significantly between genres and sub-classes of metaphor
            - Linguistic properties which can distinguish metaphors in one genre may not apply to other genres
            - Different genres are more likely to contain different types of metaphors
        - Different systems achieve similar success rates on each even though they show low-agreement among themselves
    - Weaknesses
        - Systems used in paper are forced to draw arbitrary line between metaphor/literal, even though metaphors are gradient
    - Takeaways
        - There are different types of metaphors in different genres with different linguistic properties which receive their metaphoric meanings from different sources
        - The ideal metaphor identification system will
            - First define a number of different types of metaphors
            - Model the linguistic properties which can distinguish these types from one another and from non-metaphors
        - Theories of metaphor-in-language are not mutually exclusive, but rather can apply to different types of metaphors
        - Will adopt a few of these feature systems in my project
        - Questions
            - Are metaphors really gradient as the paper claims?

Shutova 2013 - Metaphor Identification as Interpretation

https://www.semanticscholar.org/paper/Metaphor-Identification-as-Interpretation-Shutova/44707a4f10f1519fd52bd10eedf942ccc0374c78#paper-header

- What
  - Novel method that performs metaphor identification through interpretation
  - Identify metaphor interpretation as a task of finding a literal paraphrase for a metaphorically-used word
  - Symmetric reverse paraphrasing as a criterion for metaphor identification
    - Assumption is that literal paraphrases of literally-used words should yield original phrase when paraphrased in reverse, while non-literal would not have this property
- Why
  - Dependence on manually labelled data makes systems hard to scale and not applicable to real-world NLP
  - Automatic metaphor identification and interpretation in text have been traditionally considered as two separate tasks
    - However, humans are likely to perform these two tasks simultaneously
- How
  - Dataset
    - Single-word metaphors expressed by verb in verb-subject and verb-object relations (Shutova and Teufel 2010)
    - 14000-word subset of the BNC
  - Features
    - Determine likelihood of verb being metaphor based on selectional preference strength (Resnik 1993)
    - Selectional preference-based metaphor paraphrasing method (Shutova 2010) used to retrieve literal paraphrases of input verbs
    - Perform reverse paraphrasing of each identified phrase and check if same as original expression
    - If same, literal, otherwise metaphor
- Discussion
  - Results
    - Identifies metaphor with precision of 68% and recall of 66%
  - Strengths
    - Does not require manually-labeled data
    - First in NLP to combine the two tasks in one step
  - Weaknesses
    - Tested on verb-subject/verb-object metaphors only, and does not take context into account, which could improve system
  - Takeaways
    - Can be integrated with unsupervised paraphrasing and lexical substitution methods to replace WordNet filter, making system more robust and portable across domains/genres

Huang 2013 - Social Metaphor Detection via Topical Analysis

https://www.semanticscholar.org/paper/Social-Metaphor-Detection-via-Topical-Analysis-Huang/3c23687a7ec3818dd8cba5

- What
  - **Three-step framework** that is based on selection preference modeling and leverages **topical analysis techniques** to detect metaphors in social media data
- Why
  - Few works have focused on metaphor detection of social media data
  - Also, few works have focused on leveraging topical analysis techniques in metaphor detection
  - Both verbs and their arguments exhibit strong tendencies toward a few specific topics, and these topics can help identify selectional preference violations
- How
  - Dataset
    - Online support group
    - LDA model is applied to partition input corpus based on topics
    - System evaluated both on whole corpus and on each topic
  - Features
    - First extract tokens and cluster nouns
    - Then detect **selectional association outlier**
    - Then apply selectional preference strength filter to extract text snippets containing metaphors
- Discussion
  - Results
    - Topics do not have strong impacts on metaphor detection techniques
  - Weaknesses
    - Social media data is usually noisy
    - Discovering topical distribution for each term within open text is not a trivial problem

Tsvetkov et al. 2014 - Metaphor Detection with Cross-Lingual Model Transfer

https://www.semanticscholar.org/paper/Metaphor-Detection-with-Cross-Lingual-Model-Tsvetkov-Boytsov/0ddb6b3b7fd9fa86dec11eb1570886b70989e2b7

- Abstract
  - Using **model transfer** by **pivoting** through bilingual dictionary, the discussed model can identify metaphorical expressions in other languages

- Introduction
    - Problems:
        - Humans may disagree about whether something is a metaphor
        - Metaphors can be domain/context-dependent
    - Contributions
        - New detection system that uses conceptual semantic features like degree of abstractness and **semantic supersenses**
        - Created new metaphor-annotated corpora for Russian and English
        - Used **model transfer** to detect metaphors in other languages
- Conclusion
    - Model uses language-independent features and can identify metaphor across language without adapting the model
    - Study focuses on metaphors in the context of two kinds of syntactic relations: subject-verb-object (SVO) relations and adjective-noun (AN) relations, which account for majority of metaphorical phrases
- Read second time/Summary
    - What/Key Contribution
        - Identify metaphor across language with same model without adaptation
    - How/Approach
        - Data
            - SVO metaphor classifier
                - Dependency parsed and filtered TroFi dataset for SVO relations
                - TroFi (trope finder) is a dataset containing sentences containing literal/metaphorical use of English verbs
            - AN relations: constructed and made publicly available a training set from scratch
                - Phrases that depends on context (e.g. drowning students) were not included as metaphorsAdditional questions
            - Multilingual test sets
                - Collected and annotated equally balanced (metaphor vs literal) datasets for both SVO and AN relations
                - Sentences with more than one metaphor and non-SVO or non-AN metaphors were removed
                - “Low-agreement” (<.8) sentences were filtered out
        - Method
            - Study focuses on metaphors in the context of two kinds of syntactic relations: subject-verb-object (SVO) relations and adjective-noun (AN) relations, which account for majority of metaphorical phrases (84% in total)
            - Uses “coarse-grained conceptual features” rather than the “fine-grained lexical features” of individual phrases
            - These coarse semantic features can be language-invariant

- Model transfer is done by first translating syntactic constructions from other languages into English before applying English model
- Ignoring contexts, SVOs and ANs are taken as triplets/duplets and have feature vectors extracted
- For each syntactic construction (triplet/duplet), conceptual features are extracted for each word, and each pair of words, and then they're concatenated
- Three main feature categories: abstractness and imageability, supersenses, and unsupervised vector-space word representations
- Abstractness and imageability
    - E.g. abstract agent performing a concrete action is a strong signal of metaphorical usage
- Supersenses
    - Coarse semantic categories originating in WordNet (called lexicographer classes)
    - There are 26 supersense categories for nouns, and 15 for verbs
    - No high-level adjective categories in WordNet, so they used 13-class taxonomy constructed by Tsvetkov et al. (2014)
    - Supersense are preserved in translation (Schneider et al. 2013)
- Unsupervised vector-space word representations
    - Linear mapping easily captures similarity between vector space across language
    - Thus vector space models can be seen as vectors of latent semantic concepts, which preserve their "meaning" across languages
- Model
    - Random forest classifier is used
    - Good for **resource-scarce** scenario. Generalization error is limited as number of trees increases, so no hyperparameter tuning is required
    - Learn non-linear responses and often outperform logistic regression
    - This model gives probability that relation is metaphorical
- Feature extraction
    - Abstractness and imageability
        - MRC psycholinguistic database rates words by degree of abstractness and imageability

                - Propagate MRC ratings (seed set) using logistic regression classifier on vector space representations
                - Two separate models for abstractness and imageability posterior probabilities
                - These posterior probabilities are binarized based on thresholds
            - Supersenses
                - Supersenses of nouns and verbs: lexical item can belong to several synsets, each of which is associated with a supersense. Degree of membership represented by feature vectors of synset participation, whose values should add up to 1
                - Supersenses of adjectives: 13 top level classes from **GermaNet**
            - Vector space word representations
                - 64D vector-space word representation (**Faruqui and Dyer 2014)**
                - Synonymous words have similar vectors
        - Cross-lingual feature projection
            - Babylon bilingual dictionary used
            - For each non-English, first get all possible English translations, then average feature vectors (dimension==number of categories)
    - Results/Conclusions
        - Metrics
            - Accuracy, F1, and ROC curve
        - SVO: 10% cross-validation higher accuracy than Tsvetkov et al. (2013)
        - AN: 8% higher cross-validation accuracy than Turney et al. (2011)
        - Each of the 3 feature categories have good performance on their own
    - Further Research (i.e. next questions to be answered)
        - Expand scope by including **nominal** metaphoric relations
        - Incorporate **contextual feature**
        - Model transfer can be improved with more careful cross-lingual feature projection
- Discussion
    - Takeaways
        - Claim: metaphors are conceptual in nature rather than lexical
            - This sentence is unclear: what does “in nature” mean? How do they disambiguate “concept” and “lexeme”?
            - Ironically one feature they use in their model is “lexical abstractness”?

        - Further research mentions incorporating contextual features
        - Could use MRC psycholinguistic database for abstractness/concreteness in our implementation
        - Could use imageability as a supplementary feature
    - Other thoughts
        - Paper uses "coarse-grained conceptual features" rather than the "fine-grained lexical features" of individual phrases. This is opposite of approach taken by Jang 2015 which uses FrameNet over WordNet in favour of more fine-grained semantic categories to capture context
        - Phrases that depends on context (e.g. drowning students) were not included as metaphorsAdditional questions
        - Text sets have equal number of mets and literals, which would not be the case in everyday life. This might lead to higher accuracy than if it was tested on less balanced/more natural dataset with less ratio of metaphors

Klebanov 2014 - Difference Texts, Same Metaphors: Unigrams and Beyond

https://www.semanticscholar.org/paper/Different-Texts%2C-Same-Metaphors%3A-Unigrams-and-Klebanov-Leong/8bdef631341f6083cb47dc273d0966a12bab9dee

- What
    - Build metaphor detection system starting from unigram model
    - Improve recall of system
    - System should be able to classify content words in unrestricted and whole texts
- Why (Problem)
    - Current approaches in supervised learning detection systems tend to
        - Use sophisticated features based on theories of metaphor
        - Apply to certain lexical constructions, i.e. adj-noun or subject-verb-object groups
        - Restrict the type and/or contexts of metaphor. Typically snippets rather than whole texts
        - Often either rule-based or unsupervised
- How
    - Data
        - Fully annotated datasets
        - Use only content words
        - VU Amsterdam Corpus (Steen et al., 2010)
            - 117 text fragments from four genres sampled from BNC-Baby
            - Sample 23% as test set
            - All instances from same text placed in same fold of cross-validation during training to ensure generalization across texts

        - Annotated using MIP-VU procedure
        - Tagset organized hierarchically but this paper explores only top-level, i.e. function=mrw
    - Set of essays written by college graduates
        - 224 essays
        - 80 set aside for future experiments
        - 85 discuss “High-speed electronic communications media, such as electronic mail and television, tend to prevent meaningful and thoughtful communication” (Set A)
        - 79 discuss “In the age of television, reading books is not as important as it once was. People can learn as much by watching television as they can by reading books.” (Set B)
        - Multiple essays on same topic allows comparison of inter- as well as intra-topic performance
        - Annotation emphasizes connection between metaphor and arguments of the writer
        - 10-fold CV in each set as well as across the two sets of essays
    - Features
        - Unigrams
            - All content words, without lemmatization
        - Part-of-Speeach
            - Stanford POS tagger 3.3.0 and full Penn Treebank tagset for content words (tags starting with A, N, V, and J), where stopwords are have, be, do
        - Concreteness
            - Brysbaert et al. (2013) database of concreteness rating for 40 000 English words
            - Ratings are **binned**, bins used as binary features
        - Topic models
            - LDA
            - 100-topic model trained on lemmatized NYT corpus years 2003-2007
            - Gensim toolkit used for building models with default parameters
            - Score assigned to instance w on a topic t is **log((P(w|t)/P(w))**
                - P(w) estimated from Gigaword corpus
            - Based on hypothesis that certain topsic are likelier to be used as a source domain for metaphors than ours
    - Method
        - Logistic regression classifier
            - SKLL package based on scikit-learn
            - Optimizes for f1-score

            - First examine unigram model, then try all features
- Results
    - Discussion
        - Unigram model contributes most, followed by topic model
        - “One sense per discourse”, Gale et al. (1992): words keep their sense within the same text 98% of the time
            - I.e., If a word is used as a metaphor once in a text, it is very likely to be used as a metaphor again in the same text
        - Results contrast Dunn 2013 systems, e.g. News was hard for Dunn but easier for Klebanov
    - Positives
        - Baseline
            - Performs very well on essay sets, with precision above 70%
            - News fragments of VUAMC also high performance, suggesting sharing of metaphor across texts
            - Performance low for other VUAMC partitions but non-trivial
        - Beyond baseline
            - Improves recall significantly
                - Based on **McNemar’s test of significance of differences between correlated proportions**
    - Negatives
        - Does not take into account contexts of metaphors
        - Improvements upon baseline very small
        - Weak performance on non-news VUAMC texts
            - May be due to lack of common topics across texts
- Takeaways
    - Uses same dataset (VUAMC). I will base my experimentations from this paper, with similar dataset partition, so there can be a baseline for comparison
    - Can use McNemar’s test of significance
    - Used some similar features as Hyeju 2015
    - Shows usefulness of simple unigram model to complement different ways of detecting metaphoricity
- Thoughts/Questions
    - Surprised task of classifying all content words in running text hasn’t been addressed in literature prior to this
    - All instances from same text being in same fold means folds don’t have same size. Wonder how this impacts results
    - Implementing contexts in features might be able to solve problem of lack of common topics
    - LDA used to capture tendency of words to belong to a topic to used metaphorically, as opposed to difference in topics in Hyeju’s paper

Jang 2015 - Metaphor Detection in Discourse

https://www.semanticscholar.org/paper/Metaphor-Detection-in-Discourse-Jang-Moon/ed69fd3ef309118fcb239cf0afa0ffe34257229a

- What
    - Leverage global context to detect metaphors
    - Leverage syntactic information such as dependency structures to better describe local contextual information (as opposed to Klebanov 2014)
- Why
    - Most current work on detecting metaphors focus on local context (sentences)
        - Thus cannot detect atphorical cues outside of sentence
- How
    - Data
        - Breast cancer forum
        - Annotated by Amazon Mechanical Turk workers
        - Seven metaphor candidates: boat, candle, light, ride, road, spice, and train
            - Only use posts containing these seven words, and annotate posts as whether the use of the word was metaphorical or not metaphorical
        - Development set has 800 instances, cross-validation set contains 1870 instances
    - Features
        - Global
            - Semantic category
                - SEMAFOR/FrameNet
            - Topic distribution
                - LDA with GibbsC++
            - Lexical chain
                - ELKB toolkit using Roget's thesaurus
            - Context unigram
        - Local
            - Semantic category
                -
            - Semantic relatedness
                - AvgSR and DepSR
                - Cosine similarity of their topic distributions
            - Lexical abstractness/concreteness
                - Brysbaert's database of concreteness ratings
            - Grammatical dependencies
                - Stanford CoreNLP
    - Classifier

        - Logistic regression ten-fold CV
- Results
    - Discussion
        - In general, local features more successful at capturing information than global features
    - Positives
        - By using set of metaphor candidates, we can choose words that have clear distinctions between metaphor/non-metaphorical usages of the same word, unlike many other words
        - Both global and local contexts perform significantly better than all of the baselines
    - Negatives
        - Predefining set of metaphor candidates words means less generalization, lacks confirmation of value of methods used on unrestricted texts
        - Local features drowned out global features in effectiveness rather than work synergistically
        - Low performance of lexical chains
            - ELKB uses thesaurus, which might be very vulnerable to noise
- Takeaways
    - This is the paper whose methods I will implement on VUAMC
    - Three classes of features
        - Selectional preferences
        - Abstractness and concreteness
        - Lexical cohesion
    - Low performance of lexical chains just like my implementation
    - This determines whether posts are metaphorical or not, therefore we would have to modify the unigram model for our purposes to not include the word itself when building the features

Jang 2016 - Metaphor Detection with Topic Transition, Emotion and Cognition in Context

https://www.semanticscholar.org/paper/Metaphor-Detection-with-Topic-Transition%2C-Emotion-Jang-Jo/0d356807cd84ab3767d80e31350a815b82283563

- What
    - Examine sentence-level topic transitions
        - Topic similarity between sentence and neighbouring sentences (i.e. surrounding context), measured by Sentence LDA
    - Model the motivation to express emotions and cognition words by using metaphors
- Why

- Using semantic incohesion between a metaphor and the dominant topic of the surrounding text to detect metaphors tend to overclassify as metaphor
  - High recall but low precision
- Topic transition patterns between sentences containing metaphors and their contexts are different from that of sentences without metaphors
- Metaphor often used to express emotions

- How
  - Dataset
    - Online breast cancer discussion forum
      - Features metaphors in conversational texts
    - Same as Jang et al. (2015)
  - System
    - Built on top of Jang et al. (2015)
    - Topic transition
      - Even if a text can technically be interpreted literally (semantically cohesive), not all words within the text are semantically related
      - Models context at sentence level
      - First assign topic using Sentence Latent Dirichlet Allocation (LDA) (Jo and Oh, 2011)
        - All words in same sentence assigned the same topic
      - Five features
        - Target sentence topic (TargetTopic)
          - T-dimensional binary feature, where T is the number of topics
          - Indicates topic assigned to the sentence containing the target word
        - Topic difference (TopicDiff)
          - Two-dimensional binary feature that indicates whether topic assigned to target sentence is differefnt from that of the sentences before and after it
        - Topic similarity (TopicSim)
          - Two dimensional, continuous values between 0 and 1
          - Cosine similarity between word distribution of the target sentence's topic to that of sentences before and after it
        - Topic transition (TopicTrans)
          - 2T-dimensional feature, whereT is the number of topics
          - Indicates the topics of the nearest sen-tences that have a different topic from the target sentence
        - Topic transition similarity (Topic-TransSim)

            - Two-dimensional continuous feature
            - Cosine similarity between the topic of the target sentence and the topics of the nearest sentences that have a different topic before and after the target sentence
        - Emotion and cognition
            - Paper uses Linguistic Inquiry Word Count (LIWC) (Tausczik and Pennebaker, 2010), which counts word use in certain psychologically relevant categories
            - Words indicative of emotion and cognition should fall into one of 12 categories chosen by the paper, and are expected to appear more frequently in metaphorical caes
        - Multi-level modeling
            - Accounts for specificity of target words
            - Attempts to pair each feature with every target word while keeping one set of features independent of the target words
- Discussion
    - Strengths
        - Significant reduction in overclassification of metaphors
    - Takeaways
        - Metaphors are more likely in personal topics and emotional contexts
        - Personal topics and certain patterns in topic transition can indicate metaphor

Klebanov et al. 2016 - Semantic classifications for detection of verb metaphors

https://www.semanticscholar.org/paper/Semantic-classifications-for-detection-of-verb-Klebanov-Leong/f708ad5789b721ab534d0341a762a6d8572e9de7

- What
    - Use semantic generalization/classes to classify verbs in running text as metaphor/non-metaphor
- Why
    - Conceptual metaphor theory: Metaphor is property of concepts in a particular context of use, rather than of specific words (Lakoff and Johnson, 1980)
        - Words which very general semantic meaning man be used metaphorically in similar contexts
    - Most previous supervised metaphor detection for verbs are evaluated on selected and/or small examples rather than naturally occurring running text
- How
    - Data
        - VUAMC

            - All verbs besides have, be, do
    - Features
        - Grammar-based
            - Lemma unigrams
        - Resource-based (VerbNet frames and WordNet classes)
            - VN-Raw: syntactic frames/VerbNet classes
            - VN-Pred: Predictive meaning of verb
            - VN-Role Thematic role
            - VN-RoRe: Thematic role restrictions
        - Corpus-based
            - Automatically generated verb clusters as classes
            - Spectral clustering with SCF and the verb's nominal arguments are features
            - 150 clusters
    - Method
        - Tried Logistic Regression, Random Forest, and Linear Support Vector Classifier
        - During training, each class is weighted in inverse proportion to its frequency. Optimized for F1
        - Calculated consistency of metaphoricity behaviour of semantic classes across genre by calculating correlations between weights as assigned by UL+WN model to the 15 WordNet features
- Results
    - Discussion
        - Knowledge-lean system is relatively easy to adapt to a new domain or language
        - Logistic regression better for unigram features, random forest better for WordNet and VerbNet classifications, whereas corpus-based features yielded similar performance across classifiers
        - Only grammatical classification feature whos improvement over unigram baseline with no detriment for any of the genres
        - Distributional clusters (corpus) generally perform worse than resource-based classifications, but better when combined with lemma unigrams
    - Positives
        - Generalization from orthographic unigrams to lemmas is effective
        - Combination of lemma unigram + clusters automatically generated from large corpus yielded competitive yet resource-learn model
        - Better than SOA Klebanov 15 system
    - Negatives
        - Only good for detecting metaphors in verbs
        - Still doesn't consider context metaphors appear in
- Takeaways

    - Since distributional clusters seem to perform better with logistic regression classifier
    - One technique to improve results was training on all genres and prediction on one
- Thoughts/Questions

Jang et al. 2017 - Finding Structure in Figurative Language: Metaphor Detection with Topic-based Frames

https://www.semanticscholar.org/paper/Finding-Structure-in-Figurative-Language%3A-Metaphor-Jang-Maki/962fd9edff76d99f4a4882c3de4b7cbb2ecb00ed

- Why
    - Frame and topic information enable more accurate metaphor detection
    - Work on metaphor has largely focused on detection within individual sentences for identifying literal meaning
    - Broader conceptualization of metaphor is needed in detecting them in extended discourse
- What
    - Frame-based metaphor detection
        - Frame represents conceptual domain, defined by a number of co-occurring facets
        - Contrasts with conceptualization of metaphors as violation of linguistic rules
        - Metaphor occurs when a context introduces new frame to enhance its original frame
- How
    - Dataset
        - Breast cancer dataset, same as Jang et al., (2015)
        - Focus on journey-related words
            - After filtering out words not related to “journey” frame, training dataset contains 1119 instances, and testing dataset of 488 instances
    - System
        - Constructs metaphor frame from unannotated text by inferring implicit facets of given frame through template induction, using semi-supervised bootstrapping approach
            - Seed words
            - Collect lexico-grammar patterns
            - Cluster lexico-grammar patterns
            - Identify representative facet instances

        - Use templates on top of Jang (2015) features to label words as metaphor or literal
        - Features
            - Vector of binary features for each target word
            - Indicates which of the facets of the “journey” frame appear in its immediate context
    - Classifier
        - SVM classifier provided in LightSIDE toolkit
            - Sequential minimal optimization (SMO)
            - Polynomial kernel of exponent 2
        - 10-fold cross-validation for each feature
- Discussion
    - Strengths
    - Weaknesses
        - Limitations in scalability b/c we need to know which frame target words belong to in advance, which would require some sort of supervision
    - Takeaways
        - Frame switching can occur not just for metaphors but for other reasons, e.g. simply topic a switch

Koper and Walde 2017 - Improving Verb Metaphor Detection by Propagating Abstractness to Words, Phrases and Individual Senses

https://www.semanticscholar.org/paper/Improving-Verb-Metaphor-Detection-by-Propagating-to-K%C3%B6per-Walde/0d7fd73a29ed28e17931bafd9c663f0bc01730f0

- What
    - Compare supervised techniques to learn abstractness ratings
    - Extend abstractness to phrases (verb-noun pairs) and individual word senses of 3 million English words
- Why
    - Multisense abstractness ratings are potentially useful for metaphor detection
    - Methods to learn abstractness lack comparison in literature
- How
    - Dataset
        - VU Amsterdam Metaphor Corpus, looking only at verb metaphors
        - Comparison of abstractness norm creation approaches
            - Turney et al. (2011) first to automatically create abstractness norms
                - Greedy forward search to learn “paradigm words”
            - Another method based on low-dimensional word embeddings and linear regression classifier (Tsvetkov et al., 2013; Tsvetkov et al., 2014)

            - Comparison of resources
                - Two manually annotated resources
                    - MRC Psycholinguistic Database
                    - Brysbaert Concreteness database (Brysbaert et al., 2014)
                - One automatically created resource
                    - Turney et al. (2011)
- Discussion
    - Strengths
        - Automatically created norm-based abstractness ratings can easily cover huge dictionaries
    - Weaknesses
        - Automatically created norm-based abstractness ratings are less reliable
        - Like many other papers, focuses on verb metaphors, giving it narrower scope of application
    - Takeaways
        - Neural network outperforms other methods when learning abstractness ratings
        - Norms for multi-word phrases can be beneficial for type-based metaphor detection
        - Sense-specific norms improve token-based verb metaphor detection

Rei et al. 2017 - Grasping the Finer Point: A Supervised Similarity Network for Metaphor Detection

https://www.semanticscholar.org/paper/Grasping-the-Finer-Point%3A-A-Supervised-Similarity-Rei-Bulat/f88e97c5e5e64fa8f6273c54d6ad7645c837adef

- Why
    - Majority of metaphor processing systems to date rely on hand-engineered features
    - No consensus on which features are optimal
    - Corpus-driven lexical representations already have enough info to learn metaphor patterns
        - Sparse distributional features (Shutova et al. 2010, Shutova and Sun 2013)
        - Dense neural word embeddings (Bracewell et al. 2014, Shutova et al. 2016)
    - Deep learning methods shown be be successful in many other semantic tasks
- What
    - First deep learning architecture for metaphors which contains
        - Gating function to model source-target interaction

        - Word embedding/representation mapped to metaphor-specific space
        - Optimization with hinge loss function
    - Quantifies metaphoricity via weighted similarity function that automatically selects relevant dimensions of similarity
- How
    - Data
        - Mohammad et al. (2016) dataset (MOH)
            - WordNet is used to find verbs and annotated for metaphoricity by 10 annotators
            - Verb-subject, and verb-direct object
        - Tsvetkov et al. (2014) dataset (TSV)
            - Adjective-noun
            - Duplicates and metaphorical phrases that depend on wider context are removed
    - Method
        - Supervised similarity network
            - Inspired by Shutova et al. (2016)
                - Cosine similarity between two neural word embeddings is indicative of metaphoricity
            - Word representation gating
                - Word meanings can vary dependent on context
                - Intuitively, word representation domain can interact and give more context for interpreting a metaphor
                - Gating function modulates representation of one word based on other
            - Vector space mapping
            - Weighted cosine
            - Prediction and optimization
        - Word representations
            - Experimented with 100-dimensional skip-gram word embedding (Efficient Estimation of Word Representations in Vector Space, Mikolov et al. 2013) and 2526-dimensional attribute-based vectors trained by Bulat et al. (2017) (Fagarasan et al., 2015)
            - Also tried combination of two
- Discussion
    - Results
        - Supervised similarity network learns phrase representations with a very clear boundary for metaphoricity, in contrast to traditional compositional methods?
        - Performance on TSV dataset is higher than on MOH dataset
            - Former has more larger training set
    - Strengths
        - Outperforms metaphor-agnostic baseline (feed-forward neural network)

        - Can outperform metaphor identification systems that are based on hand-coded lexical knowledge
    - Weaknesses
        - Needs big training set to outperform hand-coded knowledge-based systems
        - Types of metaphor constructions restricted
        - Explicitly ignores context in one dataset (Tsvetkov)
    - Takeaways
        - Introduction has good overview of metaphor
        - Metaphor focused neural network could be very effective with enough data

Leong et al. 2018 - A Report on the 2018 VUA Metaphor Detection Shared Task

https://www.semanticscholar.org/paper/A-Report-on-the-2018-VUA-Metaphor-Detection-Shared-Leong-Klebanov/caaaeecfb559be3be20d5bbb91522e827dcb6c4c

- Why
    - Community working on computational approaches to figurative language is growing and as methods and data become increasingly diverse
    - One way of creating shared empirical knowledge of system performance in range of contexts is by benchmarking multiple systems on a common dataset
- What
    - Report on the shared task on metaphor identification on the VU Amsterdam Metaphor Corpus conducted at the NAACL 2018 Workshop on Figurative Language Processing
- How
    - Dataset
        - VU Amsterdam Metaphor Corpus
    - Systems
        - Bot.zen (Stemle and Onysko, 2018)
            - Encode each word into multiple vector-based embeddings
            - Embeddings fed into LSTM RNN network
            - Backpropagation uses weightings computed based on logarithmic function of inverse of count of metaphors and non-metaphors
        - DeepReader (Swarnkar and Singh, 2018)
            - NN that concatenates hidden states of forward/backward LSTMs
            - Also includes with linguistic features which improve performance further
        - MAP (Pramanick et al., 2018)
            - Bi-directioanl LSTm and Conditional Random Fields (CRF) hybrid
        - nsu_ai (Mosolova etal., 2018)

            - Linguistic features trained on Conditional Random Field (CRF) with gradient descent using L-BFGS method to generate predictions
        - OCOTA (Bizzoni and Ghanimifard, 2018)
            - Bi-LSTM combined with explicit features
            - Shows that ensemble of two types of NN works well
        - Samsung_RD_PL (Skurniak et al., 2018)
            - Three features
                - Brysbaert concreteness score
                - Intermediate hidden vector representing word in neural translation framework
                - Generated logits of CRF sequence tagging model trained using word embeddings and contextual information
        - THU NGN (Wu et al., 2018)
            - Word embeddings created using pre-trained word2vec model and other features
            - CNN and Bi-LSTM used to capture local and long-range dependencies
        - ZIL IPIPAN (Mykowiecka et al., 2018)
            - word2vec embeddings over orthographical word forms with lemmatization
            - LSTM network used for prediction
- Discussion
    - Results
        - Metaphor detection seems to be easier for verbs than other POS
        - Big discrepancy in performance between genres
            - 0.7s for Academic
            - 0.5s for Conversation
        - Most systems employed deep learning architecture effectively, but traditional feature-engineering techniques were not far behind
    - Takeaways
        - Combination of DNN and linguistic features can be promising direction for future work

Gao 2018 - Neural Metaphor Detection in Context

https://www.semanticscholar.org/paper/Neural-Metaphor-Detection-in-Context-Gao-Choi/a9f2797aeede946c14457e73b7633d282cf29f69

- Why
    - Detecting metaphors often require reasoning about whether situation in the context can actually happen

    - Most previous approaches focused on limited forms of linguistic context, for example by only providing SVO triples
- What
    - Using neural models for detecting metaphorical word use with contextual information in running text
        - Sequence labeling (metaphoricity of each word)
        - Classification (verb is metaphor/literal)
- How
    - Dataset
        - VUAMC (Steen et al., 2010)
        - MOH-X dataset (Mohammad et al., 2016)
    - System
        - Model uses Bi-LSTM to encode sentence
            - Fedforward NN for classification
            - Optimized over log-likelihood of gold labels
        - Sentence encoding
            - Vectors used shown to be useful for word sense disambiguation, task closely related to metaphor detection (Birke and Sarkar, 2006)
        - Classification model
            - Compared to sequence labelling model, another attention layer is added at the end and representation *c* is weighted sum of LSTM output states
            - Representation *c* is fed to feedforward network for label scores for target verb
- Discussion
    - Weaknesses
        - Broader context is needed to understand many metaphors, but this paper only looks at sentence containing the target word
    - Takeaways
        - Relatively standard BiLSTM models with contextualized word embeddings perform well to detect metaphors in running text of complete sentences

Zayed et al. 2018 - Phrase-Level Metaphor Identification Using Distributed Representations of Word Meaning

https://www.semanticscholar.org/paper/Phrase-Level-Metaphor-Identification-Using-of-Word-Zayed-McCrae/891acfb31ab1b674dd75fd4ca78e27839b0c49fe

- What
    - Semi-supervised approach which uses distributed representations of word meaning to identify metaphors on the phrase-level
    - Investigate the use of different word embeddings models

- How
    - Extract verb-noun grammar relations using Stanford parser
    - Pre-trained word embeddings model used to measure semantic similarity between candidate and predefined seed set of metaphors
        - Similarity threshold is used to classify candidate
    - Dataset
        - Verb-noun pairs
- Discussion
    - Strengths
        - Uses fewer lexical resources and does not require annotated datasets or highly-engineered features
    - Weaknesses
        - Seed set may greatly influence performance of system
    - Takeaways
        - Introduction has good writeup regarding challenges in metaphor detection

Klebanov et al. 2018 - A Corpus of Non-Native Written English Annotated for Metaphor

https://www.semanticscholar.org/paper/A-Corpus-of-Non-Native-Written-English-Annotated-Klebanov-Leong/faaeb60b663403a3b91c9df4f854dd9fd9beab00

- What
    - Publicly available corpus of 240 argumentative essays written by non-native speakers of English annotated for metaphor
    - Benchmark performance of state-of-the-art feature set systems on this new corpus
    - Explore relationship between writing proficiency and metaphor use
- Why
    - Most current work in supervised metaphor detection uses data from British National Corpus (BNC)
- How
    - Dataset
        - Sampled from ETS Corpus of Non-Native Written English
        - Five essays for each of the eight prompt questions
        - Two proficiency levels - medium and high
    - Annotation
        - Protocol taken from Beigman Klebanov et al. (2013)
        - Developed for analyzing argumentative writing, emphasizing identification of argumentation-relevant metaphors, which are metaphors that help the author advance her argument

    - System
        - Two feature sets:
            - v-16 from Beigman Klebanov et al. (2016)
                - Classifies verbs
            - All-15 from Beigman Klebanov et al. (2015)
                - Classifies all content words
    - Writing proficiency
        - Paper quantifies the extent of metaphor use in an essay as the logarithm of metaphors per 1000 words
- Discussion
    - Takeaways
        - Use of argumentation-relevant metaphor is far more significant indicator of essay quality than essay length

Dinh et al. 2018 - Killing Four Birds with Two Stones: Multi-Task Learning for Non-Literal Language Detection

https://www.semanticscholar.org/paper/Killing-Four-Birds-with-Two-Stones%3A-Multi-Task-for-Dinh-Eger/db8335595aa18f0b6eb849f15523266438558f1d

- What
    - View the detection problem as a generalized non-literal language (e.g. idioms, metaphors) classification problem, as they share similar features and definitions
    - Use multi-task (MTL) learning for related non-literal language phenomena
    - Investigate when soft parameter sharing and learned information flow can be beneficial by comparing two state-of-the-art multi-task learning architectures
- Why
    - Boundaries between types of non-literality are not clear
    - Many datasets using differing definitions, even when addressing same type of non-literality
    - Training data is sparse due to diverging definitions
    - Data quality can be hard to maintain due to disagreement among workers
    - MTL learning can be beneficial due to spill-over effects
- How
    - Datasets
        - Two datasets are English, other two are German
        - One task for each of the four datasets
            - Metaphor detection in content tokens
            - Classification of metaphorical adjective-noun constructions
            - Detection of idiomatic use of infinitive-verb compounds
            - Non-literal usage of particle verb

- Methods
    - MTL sequence tagging framework using hard-parameter sharing (Kahse, 2017)
        - Has to make use of all available data
    - Sluice Networks (Ruder et al., 2017)
        - Can decide if other tasks contain relevant information and if so, how much to share
- Discussion
  - Results
    - Soft parameter sharing and learning architecture outperforms hard parameter sharing
        - Former performs best in mono-lingual setting while latter can benefit from more information
  - Strengths
    - First to abstract to more general model for non-literal language detection using multi-task learning
    - Improves on single-task learning and learning on merged training data
  - Takeaways
    - In contrast to simply joining the data of multiple tasks, multi-task learning consistently improves upon four metaphor and idiom detection tasks in two languages, English and German

Mu et al. 2019 - Learning Outside the Box: Discourse-level Features Improve Metaphor Identification

https://www.semanticscholar.org/paper/Learning-Outside-the-Box%3A-Discourse-level-Features-Mu-Yannakoudakis/bbb0d8f80c294a39494b1aa9b0fd815e5b2c2d80

- Context
  - Most current approaches to metaphor identification use restricted linguistic context
    - e.g. by considering only a verb’s arguments or the sentence containing a phrase (Gao et al. 2018)
  - But broader discourse features are crucial for better metaphor identification, often extending beyond immediate sentence
- What
  - Simple gradient boosting classifiers on representations of an utterance and its surrounding discourse learned with a variety of document embedding methods,
- How
  - Dataset
    - Verbs subset of VUAMC

        - 17240 training and 5873 test samples across 4 genres
    - System
        - For each word, learn generic representation of verb lemma, syntactic arguments, and broader discourse context
        - Features concatenated into single vector and fed to gradient boosting decision tree classifier
        - Word embeddings
            - GloVe
                - 300-dimensional pre-trained GloVe embeddings
                - Pennington et al. (2014)
            - Doc2vec
                - 300-dimensional pretrained paragraph vectors learned with distributed bag-of-words method
                - Le and Mikolov (2014)
            - Skip-thought
                - 4800-dimensional pretrained vectors learned from training encoder-decoder model to reconstruct surrounding sentences of an input sentence from Toronto BooksCorpus
                - Kiros et al. (2015)
            - ELMo
                - 1024-dimensional vector from last layer of stacked Bi-LSTM model trained on Wikipedia and monolingual news data from WMT 2008-2012
                - Peters et al. (2018)
- Discussion
    - Results
        - Incorporating broader discourse features on top of immediate context improves metaphor detection performance
    - Strengths
        - Near state-of-the-art results on the 2018 VU Amsterdam metaphor identification task without the complex metaphoric specific features or deep neural architectures
    - Takeaways
        - Very relevant to project
        - According to paper, first work to incorporate broader discourse features

Glossary

- Distributional semantics
    - Semantics is the study of meaning
    - Research area that develops and studies theories and methods for quantifying and categorizing semantic similarities between linguistic items based on their distributional properties in large samples of language data
    - Relies on a specific view of meaning, i.e. that meaning comes from usage, or in other words, that the meaning of ‘cat’ comes from the way we use the word in everyday life
- Textual entailment
    - Directional relation between text fragments
- Conceptual metaphor
    - Generalization of a metaphor with source-target domain mapping
- Lexical cohesion
    - A text is said to be lexically cohesive when the words in the text describe a single coherent topic, either directly or indirectly
- Lexicon
    - A collection of individual lexemes
- Lexeme
    - **Basic unit** of lexical meaning consisting of words or several words, which is an **abstract** unit that represents the set of **forms** or “senses” taken by a single morpheme
- Morpheme
    - A meaningful morphological unit of a language that cannot be further divided
- Vector space models (VSMs)/Vector space word representations
    - They represent (embed) words in a continuous vector space where semantically similar words are mapped to nearby points
- MRW
    - Stands for: metaphor related word
- WIDLII
    - Stands for: when in doubt, leave it in
- PP
    - “Possible personification” resulting in a metaphor related word
- MFlag
    - A metaphorical signal
- Lexical chain
    -
- Semantic category
    - Generalized, underlying concept shared by words/phrases
    - For example, “human” can be semantic category for “man” and “woman”
- Latent Dirichlet Allocation (LDA)

    - A type of **generative** statistical model for discovering the abstract "topics" that occur in a collection of documents
    - It posits that each document is a mixture of a small number of topics and that each word's presence is attributable to one of the document's topics
    - The topic distribution is assumed to have a **sparse Dirichlet prior distribution**
- Grammatical relations
    - Functional relationships between constituents in a clause
- (Grammatical) Dependency relations
    - A one-to-one correspondence: for every element (word or morph) in a sentence, there is just one node in the syntactic structure
- Constituency relation
    - Sentence structure is viewed in terms of the constituency relation
    - Branching of constituents all the way down from sentence root
    - Often called context-free grammar
- Selectional restrictions/Selectional Preferences (SPS)
    - The tendency for a word to **semantically** “select”/constrain which other words may appear in a direct syntactic relation with it
    - Verbs tend to have semantic preferences of their arguments and violation of these preferences are strong indicators of metaphorical language use
    - Expressed in binary term (allowed/not-allowed)
    - Example: a verb like “eat” requires that its subject refers to an animate entity and its object to something concrete
- Metonymy
    - Figure of speech in which a thing or concept is referred to by the name of something closely associated with that thing or concept
- Verb Subcategorization Frame (SCF)
    - Subcategorization is the tendency of verbs to have restrictions on the arguments that they can take
    - For example, some verbs do not take a noun-phrase object, while some verbs do take an object, or two objects (direct and indirect)
    - Syntactic as opposed to semantic
    - The subcategorization frame gives you information about the syntactic category(s) a verb (or, predicate in general) combines with
- Spectral Clustering (SPEC)
    - A clustering algorithm
    - Has elements of randomness
- British National Corpus (BNC)
    - 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English
- RASP
    - Parsing system with multiple uses. Results provide basis for many text classification purposes

- Supersense
    - A general semantic taxonomy
- Jensen-Shannon Divergence (JSD)
- Stop words
    - Words which are filtered out before processing of natural language data
- Word-sense disambiguation (WSD)
    - Determining which of the meanings of a word is used in a sentence
- http://ucrel.lancs.ac.uk/bnc2/bnc2guide.htm