TDT4310: Exam Questions 2018 [maximum points: 80]

(Together with some things that could be mentioned when answering the questions.)

1) Give some reasons for processing human languages.

a) Allow computer agents to **communicate** with people
b) Allow agents to **acquire information** from (written) language
c) Make it easier for people to communicate with people

2) Suppose you wanted to create an artificial human language (such as Esperanto and Interlingua) that could be used for communication by everybody on the planet Earth regardless of mother tongue. Which would be most important issues to consider?

Argue which of the following would be most important to try to restrict, and why: the lexicon, the grammar, ambiguity or redundancy. Important issues would be the ease of learning the language, the choice of words, the generality of the language (regardless of domains and cultures), and expressiveness vs grammar restrictions.

3) What is the Distributional Hypothesis?

a) Words with similar usage have similar meanings
b) Similarity = share contexts: Distributional data used to model similarity
c) Distributional data used to model similarity
d) "you shall know a word by the company it keeps"

4) Describe Frege's Principle.

a) Meaning ultimately flows from the lexicon
b) Meanings are combined by syntactic information
c) The meaning of the whole is a function of the meaning of its parts
d) ('parts' = the substructure given by syntax)

5) What is Lambda Calculus (and lambda reduction)?

a) Notational extension of first order logic
b) Variable binding by an operator λ ("lambda")
c) Variables bound by λ are 'placeholders' (for missing information)
d) "lambda reduction" performs the substitutions

6) What is "Grounding"? How does it relate to the "Principle of closure" (Clark)?

a) Common ground = set of things mutually believed by both speaker and hearer
b) The hearer must ground or acknowledge the speaker's utterances
c) ***Principle of closure =*** Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it

7) When evaluating NLP systems, we use precision, recall and F-score. What are those?

a) Precision = Num correct items retrieved / Num items retrieved
b) Recall = Num correct items retrieved / Num items in data
c) F-score = weighted harmonic mean of precision and recall: $F_a = [(a^2+1)*P*R] / [a^2*P + R]$

8) What is cross-validation and how is it used?

a) Break up data into *k* folds (equal positive and negative inside each fold?)
b) Choose one fold as a temporary test set; train on the other folds; average performance over *k* runs

9) Discuss the influence of the size of the available data on the choice of text classification methods.

a) No training data: Manually written rules, careful crafting
b) Very little data: Use Naïve Bayes; get more labeled data (e.g., active learning); semi-supervised methods
c) Reasonable amount of data: SVM, Regularized Logistic Regression, Decision Trees
d) Huge amount of data: Can achieve high accuracy, but at a cost: SVMs (train time) or kNN (test time); Regularized logistic regression can be somewhat better; Deep learning
e) With enough data, classifier may not matter

10) A crucial factor for the coverage and performance of a part-of-speech tagger is its ability to handle words unknown to the system. Suggest two different methods that make it possible to assign tags even to those words.

i) Assign the most common tag to all unknown words.
ii) Assign the most likely tag based on the previous words (using the Markov assumption).
iii) Use the same probability as for the (known) word with lowest frequency in the dataset.
iv) Laplacian smoothing: add 1 when calculating the word probabilities.

11) What are the basic components of an opinion?

a) **Opinion holder** : The person or organization that holds a specific opinion on a particular object
b) **Object**: on which an opinion is expressed
c) **Opinion**: a view, attitude, or appraisal on an object from an opinion holder

12) Briefly describe how you could build a tool to determine if a movie review is positive or negative.

i) Scraping a movie review website to create a labelled corpus for training and test sets
ii) Decide on a feature set (e.g., binary features for the most frequent words; present or not in a document)
iii) (Possibly use or create a sentiment / prior polarity lexicon)
iv) Train a classifier on the data (e.g., some on-line MaxEnt or Naïve Bayes classifier, or build one yourself)
v) Make the classifier output (e.g.) 1 for positive and 0 for negative reviews

13) Statistical Machine Translation systems basically consist of two language models. Briefly describe the difference between them and how they relate to Bayes' Theorem.

In SMT, translation is solved by coupling a translation model (t_m) with a language model (l_m). A generative model can be used to predict P(T|S), which following Bayes' Theorem is: P_{tm} (S|T) * P_{lm} (T), since we want to maximize P(T|S) = [P(T) * P(S|T)] / P(S), but the denominator P(S) can be left out when finding argmax.

14) What is the noisy channel model? Describe how it is utilized in statistical machine translation.

We assume that the Target Language string TL really is a Source Language string SL. The TL string has been passed through a noisy channel and got out in a garbled way. We model this by modeling the "Source" language (i.e our target) P(T) and the channel P(S|T). Feed it our source, which is the "target" in the model.

15) The term FAHQGPMT came about partially as a parody of all the abbreviations in the computer society. It does, however, introduce three important constraints on a machine translation system. Select *one* of them and briefly describe how relaxing it can make MT feasible today.

i) Fully Automatic vs human intervention (pre- and/or post-editing, MT vs TM, translation aids)
ii) High Quality vs grasp content
iii) General Purpose vs specific (domain restricted or controlled language)

16) Describe the differences between a machine translation system using an interlingua approach and a system using a transfer-based approach. Discuss some advantages and problems with each strategy.

a) Interlingua = translation in two stages:
   (1) SL reduced to language-independent intermediate representation
   (2) TL generated directly from this representation
   ii) intermediate representation language = *interlingua*
   iii) (alt.: using a human language = *pivot* language)
b) Transfer = translation in three stages:
   (1) SL text transformed into syntactic representation
   (2) this is then *transferred* into TL counterpart
   (3) TL text generated from syntactic representation
c) Knowledge representation? Transfer level?

17) Describe the bag-of-words model and n-gram language models. How do they relate to each other?

Bag-of-words = (word) unigrams = count single word occurrences
Bigrams and trigrams count word pairs and triplets, respectively. And so on.

18) Why is the independence assumption useful for language models?

Assuming independence allows for smaller, simpler and more general models.

19) A common statement concerning natural language grammars is that "all grammars leak".
A related statement is "natural languages are inherently ambiguous".
Discuss different techniques to suppress ungrammatical interpretations in a grammar, and techniques to give ambiguous sentences different interpretations.

a) Feature bindings, restrictive grammar, restrictive lexicon.
b) Ambiguous grammar rules, alternative rules for the same construct.

20) The company *Skrivefeil AS* wants you to build a spelling correction system for Norwegian using a lexicon-based strategy, i.e., an approach that assumes that all words can be found in the lexicon. Why is this not a good idea? Briefly outline a better strategy.

A lexicon can never contain all words: language evolves and is used differently at different times and in different contexts and domains (it would be even worse is Skrivefeil's assumption entailed that they wanted all possible spelling errors to be included in the lexicon…).

A better approach would be to collect a large corpus of current (written) Norwegian and train a statistical model on it, possibly including rules for morphology and compound splitting. This could then be complemented with a lexicon. Alternatively, a model could be trained on common spelling mistakes and/or the Norwegian keyboard layout, to cover common errors.