

Final examination
CSC 205: Computational linguistics

1. Using the Stuttgart Finite State Transducer tools, define and test a transducer that takes any English syllable as input and echoes it with a slash after its onset (the consonant or consonant cluster, possibly empty, preceding the first vowel) and another slash after its nucleus (the vowel or vowel cluster immediately following the onset). Count *y* as a consonant if it is the initial letter of the syllable and is followed by a vowel, and as a vowel otherwise. Count *u* as a consonant if it occurs immediately after *q*, as a vowel otherwise.

2. The file

`/home/stone/courses/computational-linguistics/white.txt`

on MathLAN contains about twenty thousand words of English prose—the first two chapters of a book written by an American historian and college president and published in 1896 (*A history of the warfare between science and religion in Christendom*, by Andrew Dickson White).

(a) Tokenize the text, compute its unigram frequencies (that is, the number of occurrences of each word) and bigram frequencies, and prepare a table of the “frequencies of frequencies” of bigrams (as described in the third paragraph of section 4.5.2 of our textbook).

(b) Estimate the probability that a word taken at random from a subsequent chapter of the same book is a new word, that is, one not occurring in the first two chapters.

3. In our discussions of parsing, we often used \mathcal{L}_1 (Figure 13.1, page 428) as an instance of a context-free grammar for a subset of English.

(a) Extend the grammar of \mathcal{L}_1 to include the two **wh**-structure rules introduced in section 12.3.1 of the text, and extend the grammar and lexicon so that \mathcal{L}_1 can generate all five of the sentences illustrating **wh**-constructions (page 393, bottom half). You will need to add new lexical categories for some of the words in those examples.

(b) For each of the five illustrative sentences, give a parse tree that shows how your extended grammar generates it.

(c) Determine which, if any, of the five sentences are ambiguous in your extended grammar.

4. Suggest some features that one could add to the lexical items of the extended grammar you constructed in the previous exercise and some unification constraints that could add to the new productions in the context-free grammar. Describe the effect of integrating these constraints into an Earley parser for that grammar. How would the presence of the unification constraints improve the performance of the parser?

5. In this problem, we’ll use first-order logic to represent basic knowledge about the chemical elements. The name of each element (*Hydrogen*, *Helium*, etc.) will be a constant, and we’ll have unary predicates like *Metallic*, *Halogen*, and *Inert* and binary predicates like *DenserThan* and *FormsStableCompoundsWith*.

(a) Express the following propositions in the notation of first-order logic: (1) Oxygen is not metallic. (2) Uranium is denser than every element that is a halogen, but

some metallic elements are denser than uranium. (3) Inert elements form stable compounds only with halogen elements.

(b) The *DenserThan* relation is transitive. Express this in a formula of first-order logic, which we could adopt as a meaning postulate.

6. Suppose next that we want to represent the same basic knowledge of chemical elements in an OWL ontology.

(a) Write OWL elements for *ChemicalElement* as a class, and *Metal*, *Halogen* and *Inert* as subclasses of *ChemicalElement*.

(b) Write OWL elements for *oxygen* and *uranium* as individual chemical elements. Include the information that uranium is a metal.

(c) Write OWL elements for *denserThan*, specifying that it is a transitive relation, and for *formsStableCompoundsWith* specifying that it is a symmetric relation.

(d) Express the following three proposition as statements in OWL: (1) Uranium is not a halogen. (2) Every metallic element is denser than oxygen, but oxygen is denser than some non-metallic elements. uranium. (3) An element is inert if, and only if, it does not form stable compounds with elements that are not halogens.