Computation linguistic

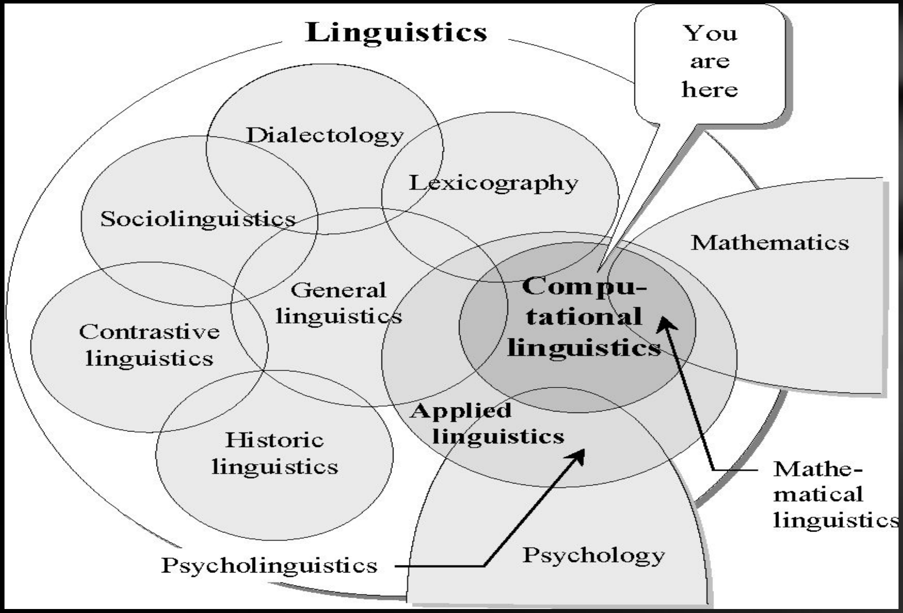
SI – 4

Daria Startseva & Alyona Gordeichuk

What is computational linguistics?

The Association for Computational Linguistics (ACL) describes computational linguistics as the scientific study of language from a computational perspective.

Computational linguistics (CL) combines resources from linguistics and computer science to discover how human language works.



Computational linguists create ***tools*** for important practical tasks such as Machine translation, Natural language interfaces to computer systems, Speech recognition, Text to speech generation, Automatic summarization, E-mail filtering, Intelligent search engines .

Computational Linguistics

- encoding/production: speech synthesis, word processing help, production side of an expert system, generation of sentences in the target language in machine translation.
- decoding/understanding: speech recognition, parsing, disambiguation via a network of semantic relations.

Language Production

- thinking: cannot be simulated

- speech/writing: computer simulation of speech sounds is possible to some extent. Computer can help this process with a grammar checker, an input system and a word breaker (in a language like Japanese). But these tasks do not simulate what people actually do when they talk.

Language Production (2)

- Though not part of the natural production process, turning speech into written text has some practical applications.
- This is very useful because speaking is usually quicker than writing. It would be like having a personal secretary.
- This is also useful for someone who cannot write because of disability or injury.

Language Understanding

- speech recognition: difficult but possible if the domain is restricted (e.g. speaker and/or expected input types)
- syntactic analysis: "parsing" (syntactic analysis by computer) is possible but needs semantic/pragmatic information for disambiguating instances of structural ambiguity.
- Interpretation (truth conditions): unclear as to how to simulate this; usually done via semantic representations (in some machine translation systems).

Corpus Linguistics

- ❑This is a generic name for various computer applications that make use of large language databases (called corpora)
- ❑Having access to a large database enabled us to process linguistic data in a statistical way, rather than in an analytical way.
- ❑This conflict of two opposing views (statistical vs. analytical) is very apparent in machine translation.

Machine Translation (1)

- text-to-text translation (great need for translation at UN, EC (European Community)
- Works best when two languages in question are similar in structure
- Usually, pre-editing and/or post-editing by a human translator is required — machine-assisted translation.

Machine Translation (2)

- Traditionally, MT required parsing, possibly some semantic analysis, then mapping to a syntactic tree of the sentence in the target language.
- An alternative is appeal to statistical means of mapping a surface string in the source language to a surface string in the target language.

Computational Semantics

- The study of how to automate the process of constructing and reasoning with meaning representations of natural language expressions.
- This could play an important role in such application areas as machine translation when two typologically distinct languages are involved (e.g. English and Japanese).

Text Summarization

- We need to be able to select the right information from the electronic documents available (esp. on the web).
- Automatic text summarization is a technique that can help people to quickly grasp the concepts presented in a document by creating an abstract or summary of the original text.

Semantic Web

- ❑ Some people are trying to classify contents of web pages so that they are meaningful to computers. But this is not an easy task since the categories must presumably be pre-selected by people.
- ❑ **The semantic Web** provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.

Speech Recognition/Synthesis

- ❑ actually being used on personal computers (on a limited basis), automated telephone answering system, etc.
- ❑ Application of acoustic phonetics, phonology

Computational linguistic students study subjects such as :

- semantic
- computational semantics
- syntax
- models in cognitive science
- natural language processing systems and applications
- morphology
- linguistic phonetics
- phonology.

Also study: sociolinguistics, psycholinguistics, corpus linguistics, machine learning, applied text analysis, grounded models of meaning, data-intensive computing for text analysis, and information retrieval.

Why are the results so poor?

- Language understanding is complicated
- The necessary knowledge is enormous
- Most stages of the process involve ambiguity
- Many of the algorithms are computationally intractable

Companies

- Alelo
- Apple
- Expert System
- Facebook
- Google
- Intel
- Lingsoft
- Lionbridge
- Microsoft
- North Side
- Nuance
- Oracle
- SDL
- S