Disambiguation

Tim Dobert

Fachbereich Informatik
Fakultät für Mathematik, Informatik und Naturwissenschaften
Universität Hamburg

June 23, 2016

Overview

Introduction

Ambiguous Cases

Ambiguities concerning a single word:

- Homographs (speech generation)
- Homophones (speech understanding)
- Homonyms (language understanding)
- Reference resolution

Multiple words

- Syntactic ambiguity

Examples

"She lives in New York."

"There is a wooden table."

"There is a bat in my garage."

"He saw that gas can explode." [1]

"Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo." [1]

Resolution Techniques

Resolution Overview

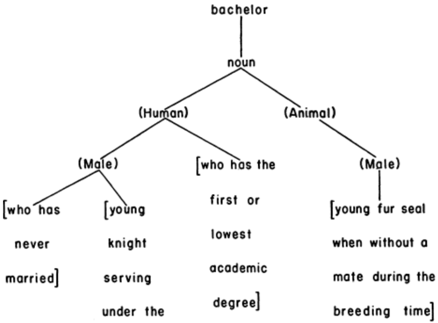
Disambiguation approaches:

- Part of speech and grammar
- N-grams
- Selection restriction

Selection Restriction 1

Using connected words to choose the correct meaning:

- Every sense has semantic restrictions
- Linguistic groundwork by Katz and Fodor (1963)



[5]

Selection Restriction 2

Limitation: Deliberate violations, "When pigs fly"

Possible solutions

- No hard restrictions, more probability
- Resnik's algorithm checks association to hypernyms

Input

First the input format has to be figured out:

- Feature Vector
- Words and their part of speech
- Co-occurrence vector

Supervised approaches

Classification can be done in two different ways:

- Naive-Bayes classifier
- Decision lists

Rule		Sense
fish within window	⇒	$bass^1$
striped bass	⇒	$bass^1$
guitar within window	⇒	$bass^2$
bass player	⇒	$bass^2$
piano within window	⇒	$bass^2$
tenor within window	⇒	$bass^2$
sea bass	⇒	$bass^1$
play/V bass	⇒	$bass^2$
river within window	⇒	$bass^1$
violin within window	⇒	$bass^2$
salmon within window	⇒	$bass^1$
on bass	⇒	$bass^2$
bass are	⇒	$bass^1$

[5]

Unsupervised

Compared to supervised approaches

- Less preparation work
- Lack of knowledge to label clusters
- Hybrid bootstrapping approach:
  - Start supervised with small labeled training data
  - Add disambiguated text with high confidence to training data
  - Usually one sense per discourse and collocation [4]

Reference Resolution

In General

Reference resolution concerns:

- Pronouns
- Definite noun phrases
- Inferrables

The following slides will focus on pronoun resolution.

Agreements

A valid reference has to agree in these aspects:

- Number
- Person
- Gender

Selection restriction can also apply.

Priorities

If multiple entities are valid, the one with the highest priority is the most likely candidate.

Priority depends on

- Recency
- Grammatical role
- Verb semantics

Algorithms

Intuitive algorithm [2]:

- Keep track of entities and their salience
- Add references to entity when they appear
- Update Salience

Other algorithms:

- Tree search
- Centering algorithm

Summary

- Single words can be ambiguous in a lot of cases
- Rule based systems can disambiguate with grammar and semantics
- Machine learning approaches can also be effective
- Reference resolution is important in a discourse
- Agreements and priorities can help narrow down the reference

References and Literature

- https://en.wikipedia.org/wiki/List_of_linguistic_example_sentences
- Speech and Language Processing, Daniel Jurafsky and James H. Martin, 1999
- A Semantic Process for Syntactic Disambiguation, Graeme Hirst, 1984
- Unsupervised Word Sense Disambiguation Rivaling Supervised Methods, David Yarowsky, University of Pennsylvania
- The Structure of a Semantic Theory, Jerrold J. Katz; Jerry A. Fodor, 1963