

NLP/MT
Principles

EBMT Principles
and Solution

EBMT & Rule-based
MT

EBMT & Knowledge-
based MT

EBMT & Stat.;
Evaluation

U+H

Example-based Machine translation with semantic information

Cristina Vertan

University of Hamburg • Informatics Department

Natural Language Systems Group

WWW: <http://nats-www.informatik.uni-hamburg.de/~cri/>

E-Mail: vertan@informatik.uni-hamburg.de

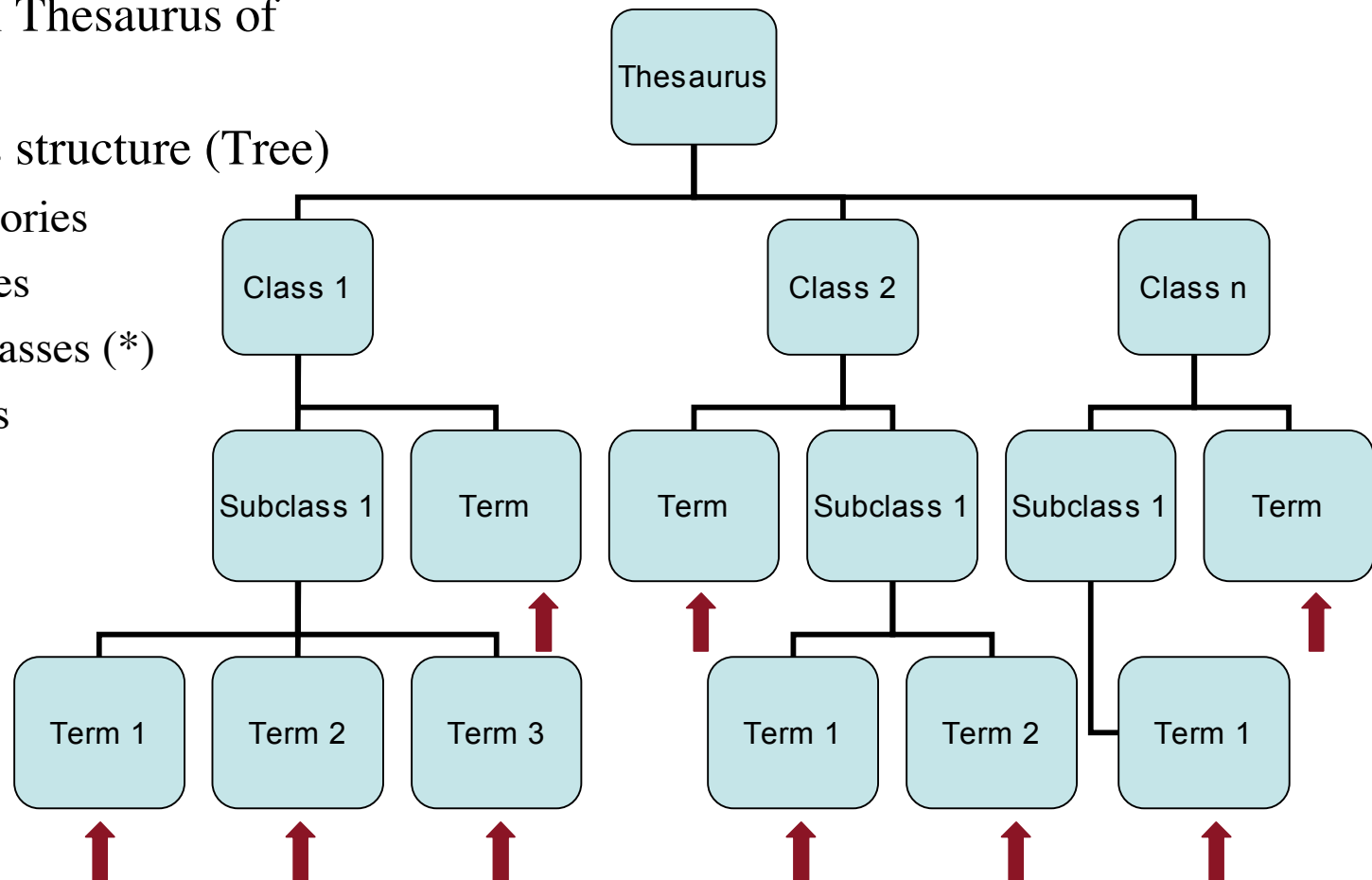
Word-based matching - 1 -

- A “semantic network“ is used additionally; in this semantical network the distances between words express semantic similarity.
- For example for the following entries in the DB:
 - *Der **Abstand** zwischen den Kontrollen soll 2 Jahre nicht überschreiten*
↔ *The **interval** between 2 general checks should not exceed 2 years.*
 - *Der **Abstand** zwischen den Nebelleuchten ist x cm.*
↔ *The normal distance between fog-lights is x cm.*
- The input : *Wo finde ich den Abstand zwischen den Rädern?*
 - *Räder* in the semantic network is closer to *Nebelleuchten*, therefore *Abstand* is translated by *distance*,
although the edit distance between *Räder* and *Kontrolle* is smaller than the edit distance between *Räder* and *Nebelleuchte*.

Construction of the Semantic Network (I)

- Bilingual Thesaurus of NOUNS.
- Elements structure (Tree)
 - Categories
 - Classes
 - Subclasses (*)
 - Terms

NOUNS



Construction of the Semantic Network (II)

- Spanish Culture
 - Entertainment
 - Fashion
 - Sports
 - Religion
 - Dietary Habits
 - Mediterranean Diet
 - Typical Food
 - Tapas
 - Art
 - Monuments
 - Mosque
 - Museum
 - Monastery
 - ...
- Spanish Geography
 - Territories (“map”)
 - Autonomous Region
 - City
 - Province
 - Town
 - ...
 - Geographical Quirks (“geo”)
 - Mount
 - Mountain
 - Mountain Range
 - River
 - Ocean
 - ...
 - Cardinal Points

Construction of the Semantic Network (III)

- Preferred Terms / Non Preferred Terms

- scopeNote

- Synonyms

- BT, NT, RT

- UF, USE

- Translation

Word Net
Structure

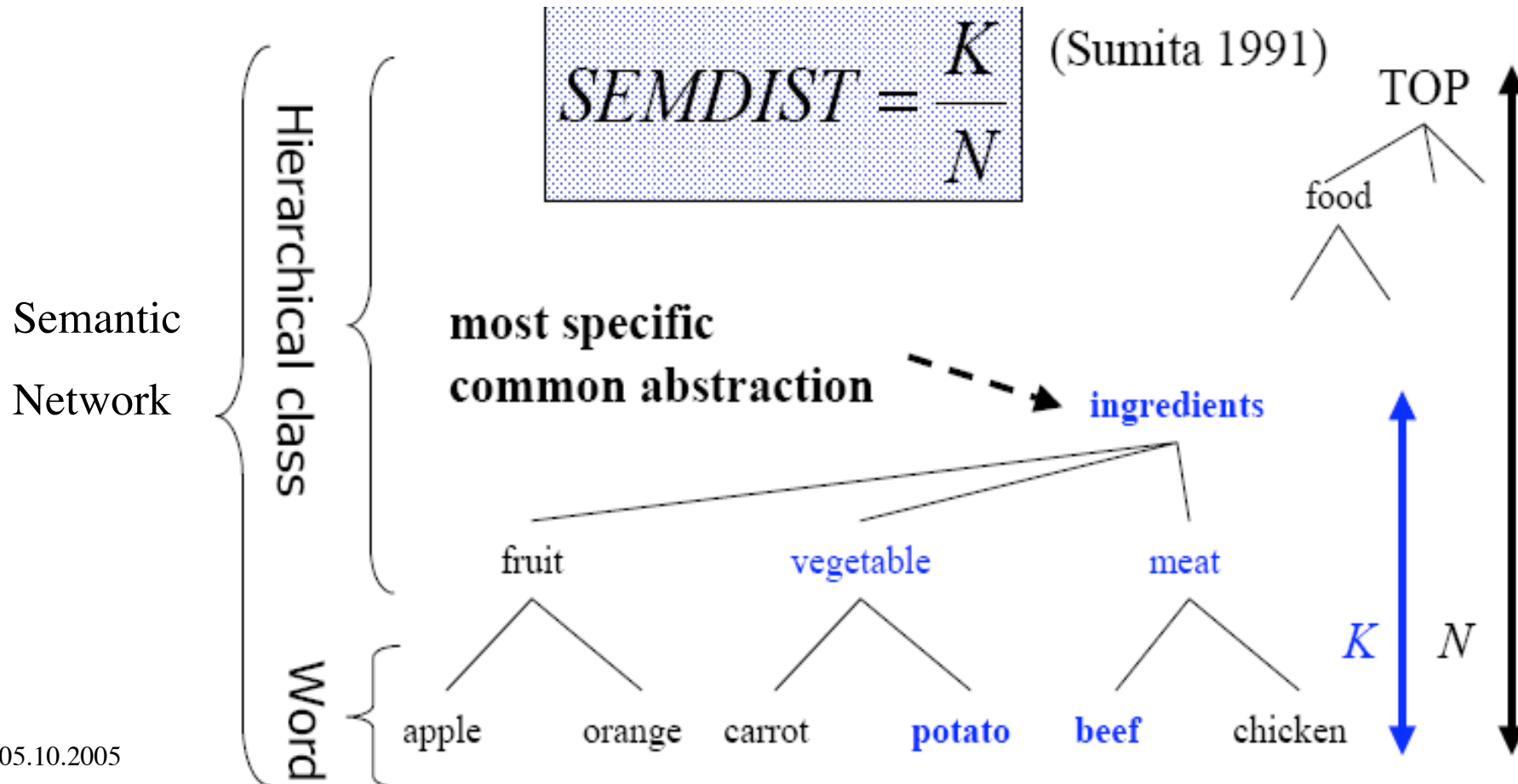


XML Format

Measuring the Distance (I)

$$dist = \frac{I + D + 2 \sum semdist}{L_{input} + L_{example}}$$

- Distance



Measuring the Distance (II)

- Semantic Distance
 - If two words are in the same subclass -> Semantic Distance = 0. Maximal Similarity.
 - Sea – Mountain -> SD = 0
 - If they are in different categories -> Semantic Distance = 1. Completely Dissimilar.
 - Sea – Museum -> SD = 1

Measuring the Distance. Sample

- Initial sentence manipulation (lexicon):
 - INPUT: “I have seen the Alhambra of Granada”

↓
“see the monum of map”

- CORPUS : “You will see the Mosque of Cordoba”

↓
“see the monum of map”

0 insertions 0 deletions 0 substitutions
dist = 0

Measuring the Distance. Sample

- Initial sentence manipulation (lexicon):
 - INPUT: “The autonomous region of Andalusia lies in the south of Spain”



“The region of map lie in the cardinal point of map”

- CORPUS : “The gulf of Almeria lies in the east of Andalusia”



“The gulf of map lie in the cardinal point of map”

**0 insertions 0 deletions 1 substitutions
semdist (region, gulf) = 0.5
dist = (0+0+2*0.5) / (11+11)**

How we could use knowledge in our system

- Organize small semantic networks for geographic regions and time events
- Make a thesaurus in which you insert the proper names referring to geographic regions (also other than those appearing in the database) and assign them concepts in the semantic network
- Before measuring the Edit distance, replace such proper names by their concepts and digits by time events
- Measure the Edit distance with these modified Strings