



Employing GermaNet in Virtual Courses of Computational Linguistics

Claudia Kunze, Lothar Lemnitzer

Seminar für Sprachwissenschaft

Universität Tübingen

{kunze,lothar}@sfs.uni-tuebingen.de

LREC Workshop "Language Resources: Integration & Development in e-learning & in Teaching Computational Linguistics"

Lisbon, 24 May 2004



Topics of the talk

- Computational Lexicography within the *MiLCA* project
- The virtual learning scenario
- Student projects centered around GermaNet
 - Modelling of linguistic contents
 - Data structure and presentation
 - Tools for accessing and visualizing wordnet structures
 - Wordnets in NLP applications
- Advantages and challenges of virtual learning scenarios



Aims of the *MiLCA* project

MiLCA – Medien-intensive Lehrmodule in der Computerlinguistik-Ausbildung: Joint BMBF project of the universities Bonn, Gießen, Osnabrück, Saarbrücken, Tübingen

The aims of the project are

- Extending the range of topics which can be offered to students at various sites
- Developing core modules for Computational Linguistics and Language Engineering curricula
- Testing and implementing of various media-based learning scenarios
- Disseminating of the results and managing the rights on these modules



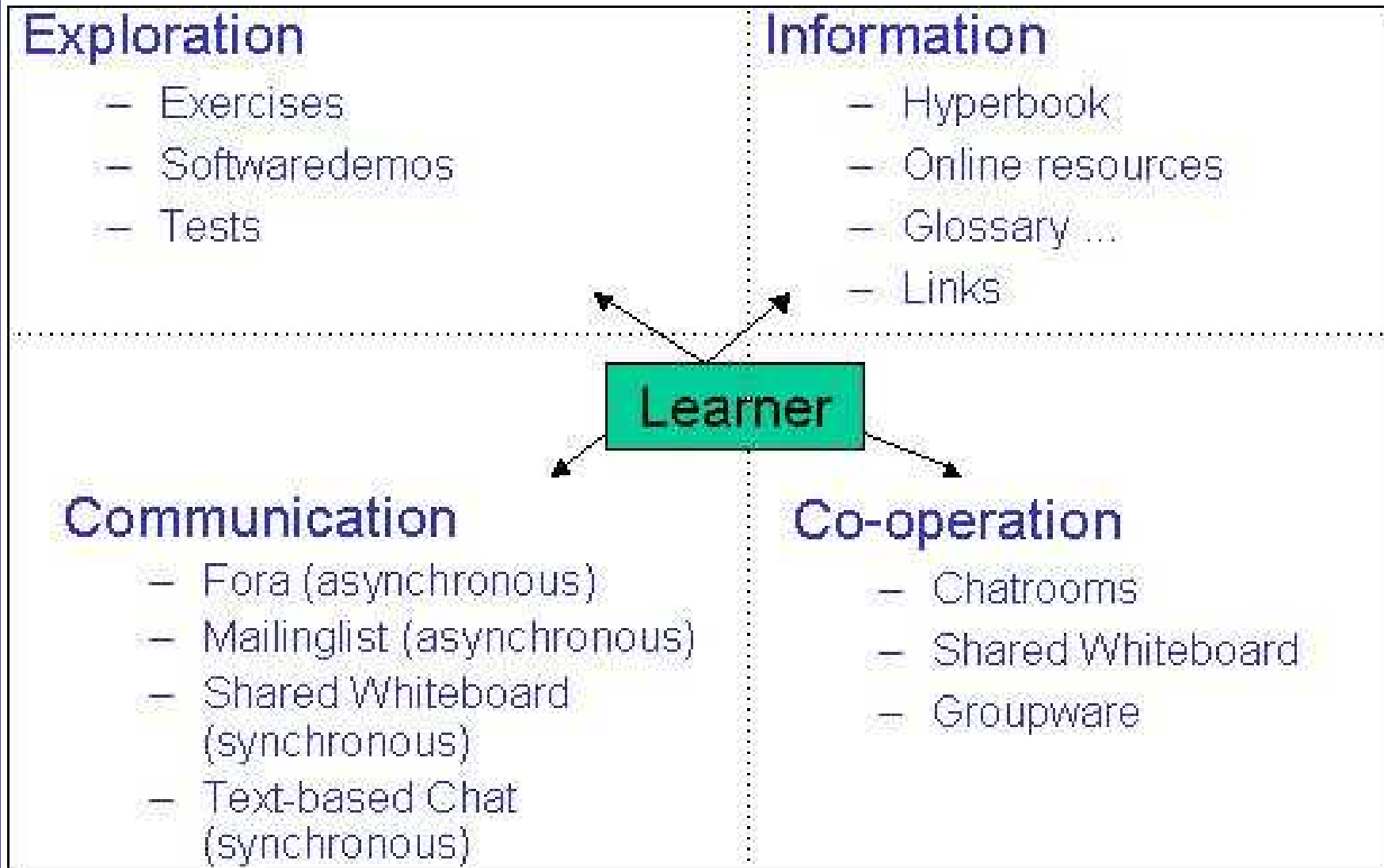
CoLex within the *MiLCA* project

Topics: Machine-readable lexical resources, their generation, management and use in NLP

- Lexicon Theory in linguistics
- Lexicology and Lexical Semantics
- Treatment of multi-word units
- (Meta-)Lexicography
- Lexical Rules
- Lexical Databases
- Lexical Acquisition from text (corpora)
- Computational Linguistics and Text Technology - Tagging, Parsing etc.



The virtual learning scenario





The virtual learning scenario

The learning process in the virtual course is as follows:

- Students prepare a topic in advance, with a textbook
- They discuss the topic, in plenary and group sessions, with the help of synchronous computer-mediated communication (text based chat)
- They solve exercises
- They work on their own or in groups on a larger project to earn a certificate



The virtual learning scenario

GermaNet plays a key role in this learning process

- GermaNet represents a prominent type of lexical database
- GermaNet qualifies as a reference resource and case study
- GermaNet yields appropriate material for exercises in different fields



Student projects centered around GN

Student projects which have been assigned so far, in relation to GermaNet or wordnets in general, encompass:

- Modelling of linguistic contents
- Data structure and presentation
- Tools for accessing and visualizing wordnet structures
- Wordnets in NLP applications



Modelling of linguistic contents

Analysis of the meronymy / holonymy relation

- Question: given the German lexical data, is it justified to subdivide this relation into a set of (sub)relations?
- Question: given that, is the subdivision implemented in the Princeton WordNet (3 different pointers for PART, MEMBER, SUBSTANCE) adequate for GermaNet?
- Or is the subdivision realized in EuroWordNet with one generic meronymy pointer and five subpointers (PART, MEMBER, PORTION, MADEOF, LOCATION) even more adequate?
- What about other classifications like the one of Chaffin with 7 categories?



Analysis of the antonymy relation

- Question: given the German lexical data, is it justified to subdivide this relation into a set of (sub)relations (CONTRAST, COMPLEMENTARITY, CONVERSE, DIRECTIONAL OPPOSITION)?
- If so, develop an empirically justified subdivision of this relation



Applicability of regular polysemy in wordnets

- Question: which are the advantages and disadvantages of implementing rules which handle regular polysemy?
- Question: is there an appropriate hierarchical level for the implementation of such rules (e.g. the basic concept level)?

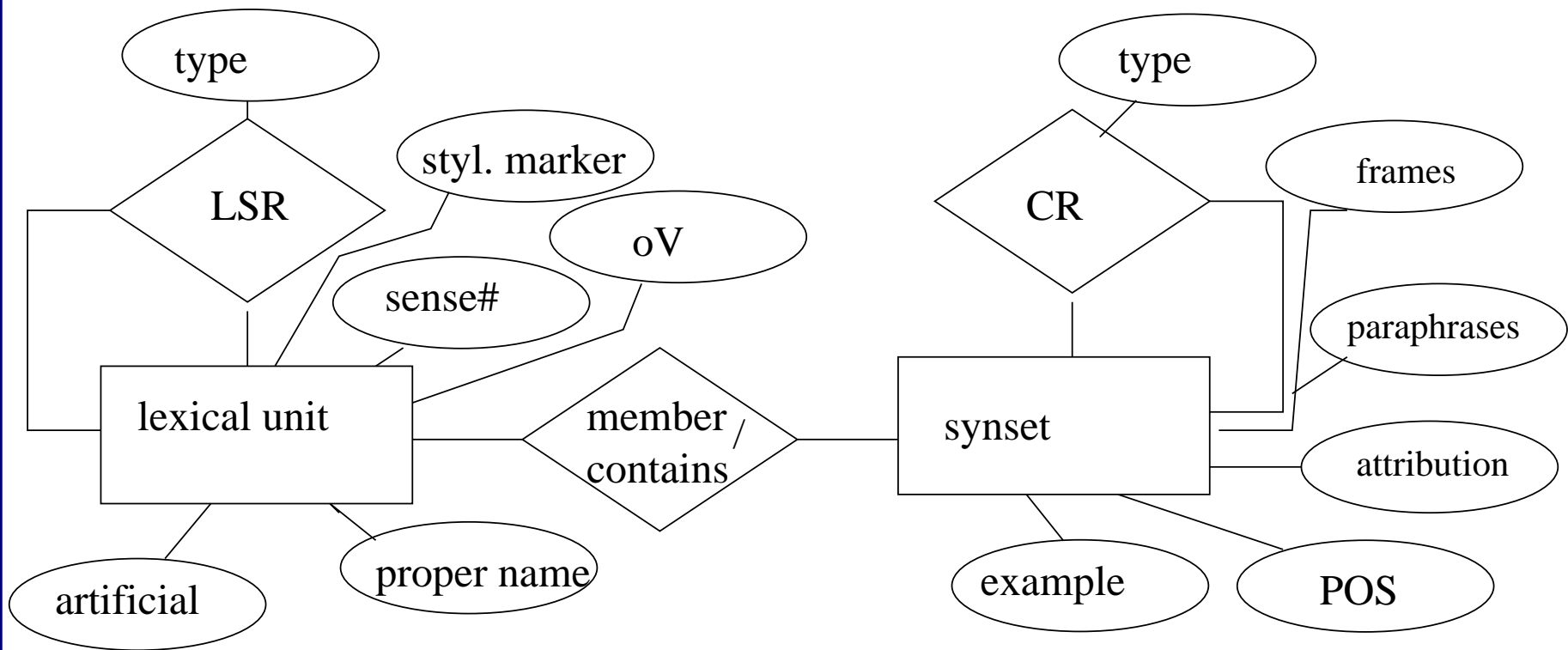


Data structure and representation

- Convert the lexicographer's files of GermaNet into an XML compliant format
- The conversion has been based on an Entity Relationship model of the GermaNet data structure
 - Extended Links (XLink) have been used to model the relations between synsets and literals ('lexical units')
 - The XML version of GermaNet is the basis for data exchange, visualization and further conversions



ER-graph of the GN data model



CR=conceptual relation; LSR=lexical-semantic relation; oV=orthographic variant



Data structure and representation

Explore the possibility of integrating GermaNet data into the framework of an ontology language (OWL)

- An experiment with a data sample showed the feasibility of this approach
- The use of OWL as modelling language seems to be superior to the use of RDF(S)
- Some problems still remain to be solved (e.g. classes as instances, blocking of undesired inferences)



Accessing and visualizing WN structures

Develop a tool for the extraction of relational neighbours of any synset or lexical unit

- two approaches have been tested. One based on GermaNet as a relational database, one on the XML encoded version of the GermaNet data
- Both approaches turned out to be reliable, with the RDMBS approach being slightly faster



GermaNet extraction tool

The screenshot shows the GermaNet Browser window with the following details:

- Menu: System Buffer SQL Help
- Navigation: Synsets (selected), LexUnits, Words
- Filters: WordClass: **nomen**, LexGroup: **Artefakt**, SynsetId: 222
- Buttons: Search, hyperonymy (selected), Chains
- Results: Result#1, Result#2, Result#3

```
<results>
  <result>
    <synset id="nArtefakt.222" wordClass="nomen" lexGroup="Artefakt">
      <lexUnit id="nArtefakt.222.Stütze" stilMarkierung="nein" sense="1" orthVar="nein">
        <orthForm>Stütze</orthForm>
      </lexUnit>
    </synset>
    <con_rel name="hyperonymy" dir="->">
      <synset id="nArtefakt.221" wordClass="nomen" lexGroup="Artefakt">
        <lexUnit id="nArtefakt.221.Stützkonstruktion" stilMarkierung="nein" sense="1">
          <orthForm>Stützkonstruktion</orthForm>
        </lexUnit>
      </synset>
      <con_rel name="hyperonymy" dir="->">
        <synset id="nArtefakt.27" wordClass="nomen" lexGroup="Artefakt">
          <lexUnit id="nArtefakt.27.Konstruktion" stilMarkierung="nein" sense="1">
            <orthForm>Konstruktion</orthForm>
          </lexUnit>
          <lexUnit id="nArtefakt.27.Vorrichtung" stilMarkierung="nein" sense="1">
            <orthForm>Vorrichtung</orthForm>
          </lexUnit>
        </synset>
        <con_rel name="hyperonymy" dir="->">
          <synset id="nArtefakt.1" wordClass="nomen" lexGroup="Artefakt">
            <lexUnit id="nArtefakt.1.Artefakt" stilMarkierung="nein" sense="1">
              <orthForm>Artefakt</orthForm>
            </lexUnit>
            <lexUnit id="nArtefakt.1.Werk" stilMarkierung="nein" sense="1">
              <orthForm>Werk</orthForm>
            </lexUnit>
          </synset>
        </con_rel>
      </con_rel>
    </con_rel>
  </result>
</results>
```

Conceptual chains(hyperonymy) for synset ("nomen", "Artefakt", "222")



Accessing and visualizing WN structures

Develop a wordnet visualization tool built on the XML version of the data

- Result: the implementation (using Java) does not work very well
- Consequence: we will try to convert GermaNet structures into SVG and look for 'off-the-shelve' visualization tools



GermaNet visualization tool

The screenshot displays the GermaNet visualization tool interface. On the left is a file explorer showing a directory structure with folders like 'files', 'adj.Bewegung', 'adj.Geist', 'nomen.Mensch', 'nomen.Tops', and 'nTops.1' through 'nTops.13'. The 'nTops.6' folder is selected, containing 'geschlecht', 'Parasit', and 'Schmarotzer'. At the bottom left are buttons for 'Start XML' and 'Start REL'. The main area shows a network diagram with nodes and edges. A central node 'nTops.6' is connected to a cluster of nodes including 'Ding', 'Sache', 'Gegenstand', and 'Gebilde'. Another cluster contains 'Parasit' and 'Schmarotzer'. A third cluster contains 'Lebewesen', 'Kreatur', and 'Wesen'. A 'Subjekt' node is connected to the 'Lebewesen' cluster. The top of the window has a 'top' button. On the right is a control panel with a 'Show' column and three columns for filtering: 'for All', 'for C', and 'for O'. The panel includes checkboxes for 'Hyponyms', 'Hyperonyms', 'Meronyms', 'Holonyms', 'Entailments', 'Causations', 'Associations', 'Antonyms', 'Pertoynms', and 'ParticipleOf'. Below these is a 'LexUnit:' section with the following values: 'Eigenname: nein', 'artificial: nein', 'id: nTops.2.Lebewesen', 'orthVar: nein', 'sense: 0', and 'stilMarkierung: nein'.



Wordnets in NLP applications

A vocabulary trainer has been developed in a joint virtual seminar and project of Osnabrück, Edinburgh and Tübingen

- Development of a network-like platform for collaborative work (MAPA)
- Framework for constructing networks of knowledge which enables users to link words and data of any kind such as images, videos, e-mails and documents in a cognitively adequate way



Wordnets in NLP applications

- Development of a vocabulary trainer as reference application: GN data help the language learner to understand and learn German words in a wider semantic context
- User profile overviews the user's individual learning strategy
- GermaNet data have successfully been used for the training module



Advantages of virtual learning scenarios

- A virtual computer-mediated learning environment enables access to various language resources, including lexicons and corpora
- Hypotheses can be tested against large bodies of data, using sophisticated CL techniques
- Virtual learning scenarios broaden the access to linguistic contents
- Special interest groups can be established with learners and experts from different sites and fields. Thus, interdisciplinary work is promoted
- Perspective of contributing to an ongoing research activity increases the intrinsic motivation of the students



Challenges of virtual learning scenarios

Challenges for teachers who want to use existing language resources are

- to define projects of manageable size
- to prepare the data so that students work only on those samples which are relevant to the task
- to give an account of the state of the art as a context for the project
- to provide for the relevant background literature