
Lego Bricks for Lexicons

Peter Wittenburg

Max-Planck-Institute for Psycholinguistics

Problems

Syntax Issues

- lexicon formats are different
- lexicon structures are different (language and theory dependent)
- will remain - but can we create a flexible framework for all needs

Gains

- increase of interoperability
- increase of reusability
- simpler merging for searching etc

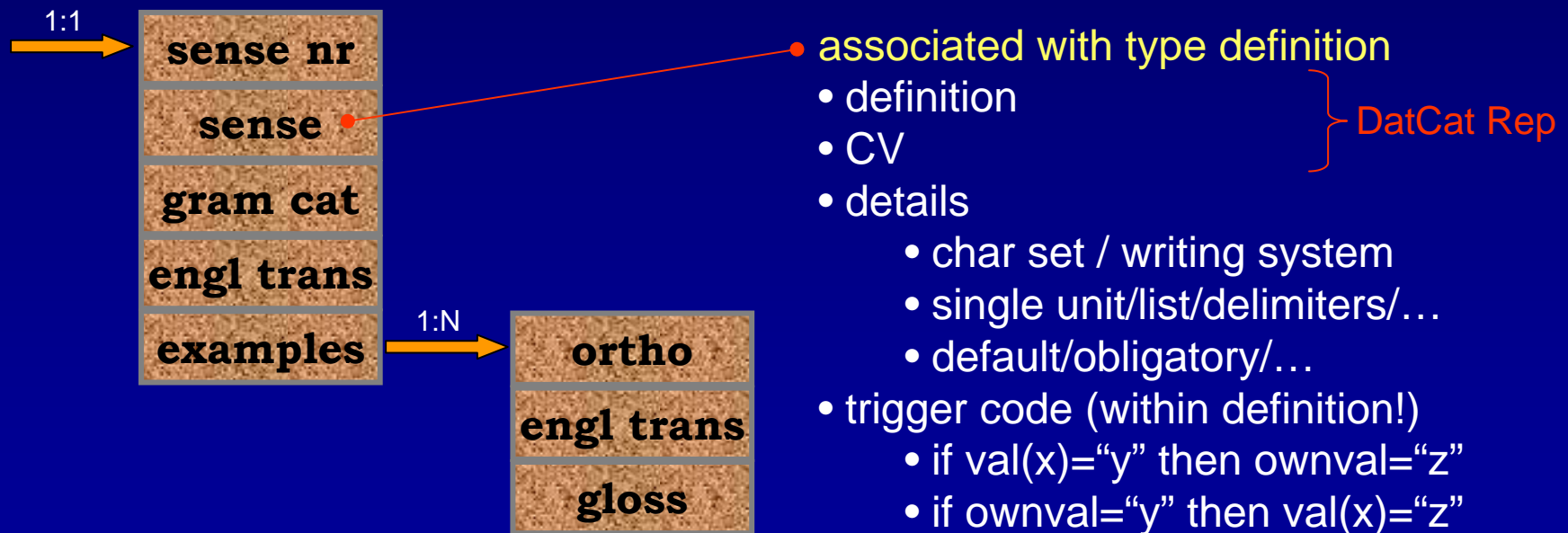
Need a framework that can be called an Abstract Lexicon Model

What is it?

- An exchange model?
- A workable model?
- A blueprint for a flexible tool?

Lexicon Lego Bricks

- concrete lexicon schema is basically seen as lexical attributes grouped with others and embedded in a tree structure.
- it is assumed that a scenario is available as indicated in the first talk



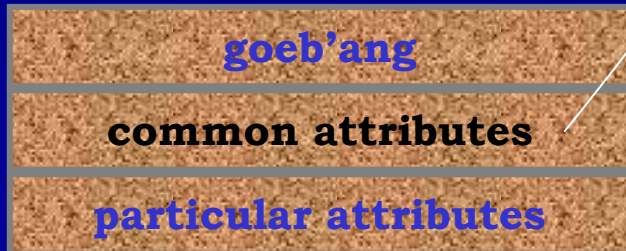
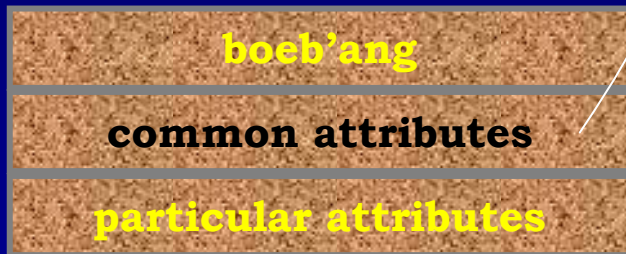
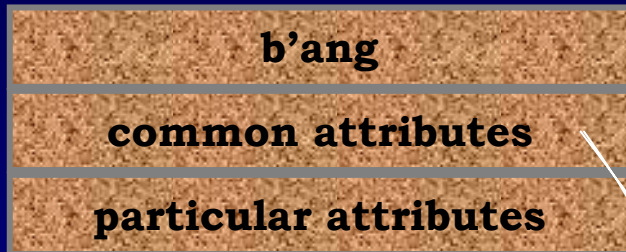
Relations between Bricks



- need various type of relations between attributes and units in value strings
- each relation can be associated with features, i.e. relations can be seen as attributes in its own
- can be units within a string that need to be related to others

Inheritance between Bricks

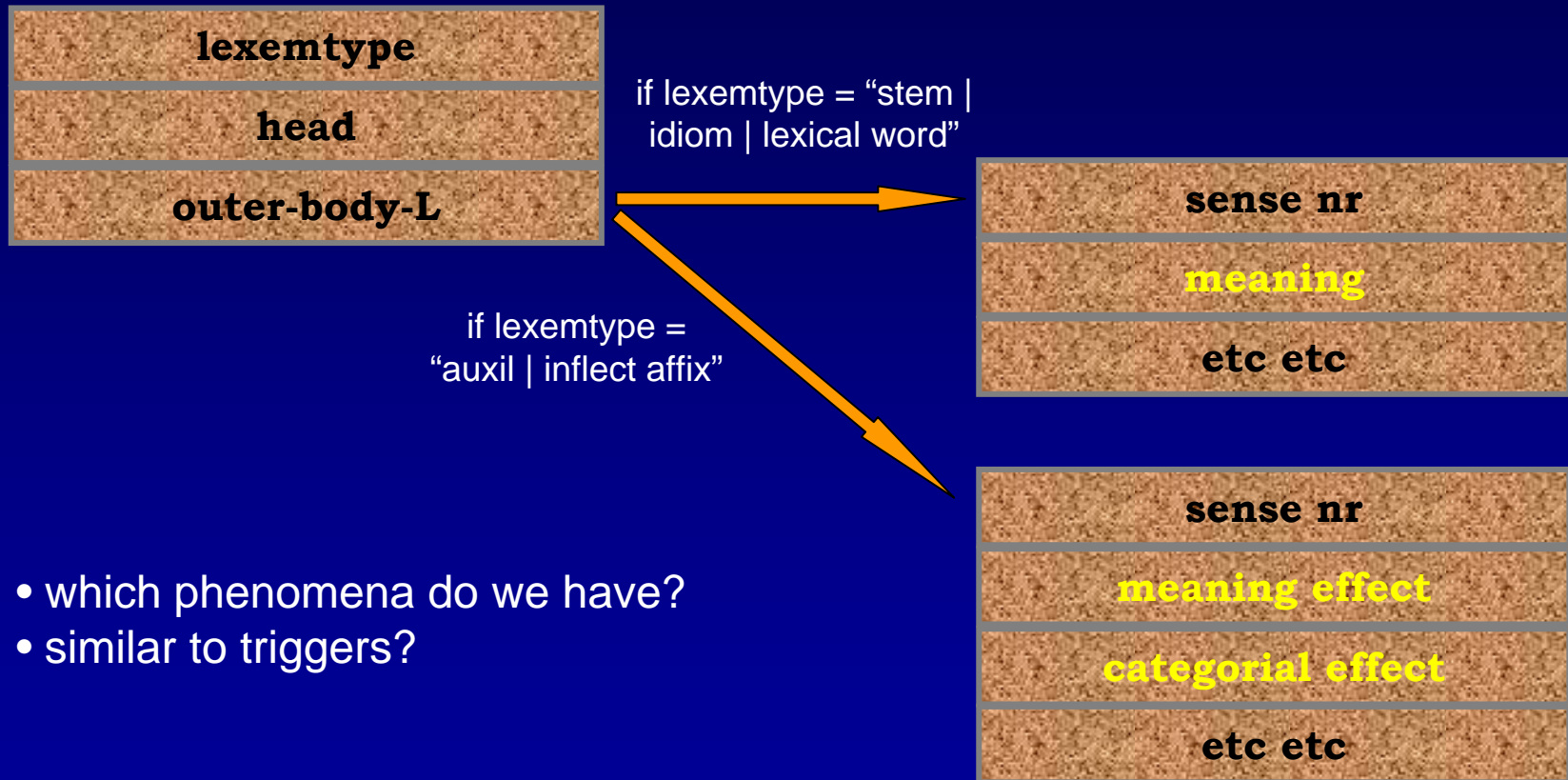
just one example to reduce typing



- which phenomena do we have?
- discussion in paper by Ide/Romary
 - cumulative features
 - overwriting features
 - local exceptions

Conditions within Bricks

just one example from DOBES



ISO TC37/SC4 Tasks

- what do we need?
 - summary of all requirements for lexicon brick
 - create an abstract model (incl. relations, ...) probably in UML
 - design a persistent format for abstract lexicon model
 - recommend framework for re-usage of elements and schemas
 - requirements for registration services