

Managing multilinguality in embeddable language processing

Atro Voutilainen
Connexor



Connexor

U Helsinki

Conexor 1997

Machineese sdk

~150 customers worldwide

IR, IE, interaction, MT,

education, proofing, speech

challenge

the apps developer
wants to use different NLs
but knows only C!

interpreter design

computationally and linguistically
adequate formalisms

language-generic interface (API)

case: Machineese

**Phrase Tagger, Syntax,
Semantics, Metadata**

10 European languages
multiple platforms & technologies

language models

finite state morphology

constraint grammar

functional dependency grammar

text

representation

uniform, systematic

focus on useful information

baseforms, compound structure

word class, inflection, derivation

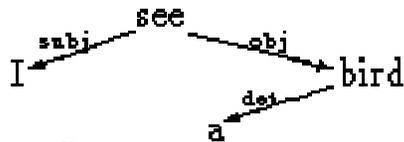
phrases, terms, keywords

names, "things"

representation (contd)

relations

- as functional dependencies



- core functions
- circumstantial functions

coreference



Thank you!