INSTITUTO DE LINGUÍSTICA E COMPUTACIONAL

MorDebe & NeoTrack

Using Neologism Research for Lexical Resources

Maarten Janssen















Loss of Work

iltec -

- All candidates are checked manually
 Trained linguist
- Many words are non-neologistic
 - Will not be stored in the neologism database

Pity to throw them away

- They will reappear on the candidate list
- They should be in the dictionary
- Dictionary should be structured















Regular Manual Check Weekly check of new additions Ideally - check for errors

- Daily backups
 - No roll-back extensive change log files
- Check by use
 - Multiple linguistic projects

User feedback

- Mostly irrelevant questions
- Sometimes indications of errors
- New feature: user words

Inherent Inflection

iltec —

- Between inflection and derivation
 Different entries with strong ties
- Modeled by "Lexical Functions"
 - Inflectional Functions
 - s0a, s0v, adv0, a2, able2, max, min, aum, dim, female
- Orthographic variation
 - alt, pe
 - doirar / dourar 2 different paradigms
 - "same" word

Special Cases

iltec —

Defective Paradigms

- Non-existing forms marked by "-" (nevo I snow)
- Adundant Paradigms
 - Two entries for same lemma with same inflection code
 - Waaide / woei

Homonymy

- Not semantically driven
- Formally driven
- redar redou re+dar / redo rede+ar

Part III : OSLIN

Open Source Lexical Information Network

Open Source - Modular Network All Lexical resources need a lemma list

- MorDebe could provide thatOpen source useable for everybody
- MorDebe modular design
 - Word forms separate database
 - Grammar yet another
 - Extendible quasi indefinitely

Web Based

ilteC -

- Different parts can be maintained by different people
- Cooperation rather than competition





- 0 (0



ilteC •

Reusable Resources

Hype Concept

- Many old resources unusable by now
- EAGLES, ISLE-MILE, etc.
- Reusability by standardization
- Build new sources upon old ones

OSLIN different

- Reusable by extension
- The lexicon can be used live rather than once
- Politically difficult but conceptually better