

What did STEVIN do for HLT Research?

Walter Daelemans

CLiPS Research Group, University of Antwerp
STEVIN Final Event, Rotterdam, Nov 28 2011



U | STEVIN

Overview

- Priorities in research
- Highlights
- Impact
- One view on the results of STEVIN
 - Disclaimer
 - Parochial (point of view of research)
 - Personal (CL-oriented)

Highlights Written

- Research priorities written language:
 1. Semantic analysis (tagging, integration with syntax and morphology)
 2. Text pre-processing (tokenization, spelling correction, named entity recognition, ...)
 3. Morphological analysis (compounding and derivation)
 4. Syntactic analysis: a robust parser for Dutch

Achievements

- D-COI/SONAR, COREA, IRME, DPC, CORNETTO
 - Analysis tools associated with annotation layers (NER, COREF, ...), lexicons (MWE), alignment tools (1,2,4)
- LASSY
 - Robust dependency parser + treebank + tools (4)
- DAESO
 - Semantic relations, paraphrases (1)
- Task-oriented projects
 - DAISY (summarization), DUOMAN (sentiment), PACO-MT (translation) (1,2,4)

Highlights Spoken

- Research priorities spoken language:
 1. Robustness of speech recognition
 2. Output treatment (inverse text normalization)
 3. Confidence measures
 4. Adaptation
 5. Lattices

Achievements

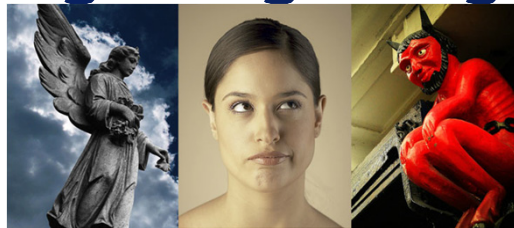
- JASMIN, MIDAS, N-BEST
 - Robustness with different types of speakers, robustness to noise and changing acoustic environments, evaluation data and protocols (1, 4)
- AUTONOMATA, AUTONOMATA TOO
 - Tools and data for name recognition and synthesis (1)
- SPRAAK
 - Complete SR system (1,2,3,4,5)
- DISCO
 - Robustness in CALL context (1)
- STEVIN can Praat
 - Additional functionality

Impact

- Fair number of publications, majority in workshops and conferences (LREC)
- Impressive legacy of tools and data thanks to HLT central model (that should be adopted at a European level)
 - Many Dutch / Flemish projects in other funding streams (IWT, NWO, FWO, DH) are starting to use these tools
- Citations limited (still)
- Interesting research, but
 - Incremental
 - Application / resource creation oriented
 - “Doing it for Dutch”
 - There are exceptions!

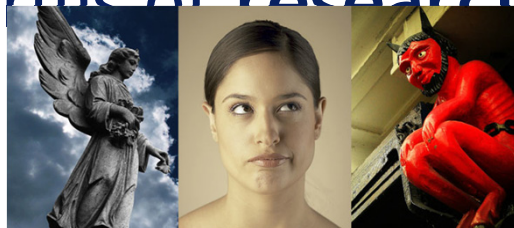
So what has STEVIN done for research?

- Education of HLT researchers
 - But: no (immediate) continuation (again)
- Necessary resources for research
 - But: predominantly for Dutch (lower impact)
- Tight(er) links between Dutch and Flemish researchers
 - But: we were getting along very well before too!



So what has STEVIN done for research?

- Resources are high on the European HLT agenda (Stevin is a pioneer!)
 - But:
 - IBM WATSON and Apple SIRI are not “*made in USA*” by accident
 - Google is doing Dutch SR and MT
 - (focus on Dutch / multilingualism costs us in Europe in terms of research impact)



Did we have the right priorities?

- We missed the “Text Understanding” revival (Watson, DARPA MR program, open information extraction) and corresponding resources, we missed SIRI and dialogue in general
 - But IMIX
- The priorities were a bit conservative and unadventurous
 - They were state of the art, but the state of the art changes quickly
 - Allow redefinition in future long-term programs (also by the external advisory panel)

Did we have the right organizational model?

- CGN (Spoken Dutch Corpus) model
 - Decide what has to be done and divide the work among research teams with proven past performance and experience
 - Has the advantage of efficiency and goal directedness
- STEVIN model
 - Set priorities (and a few tenders) but select proposals bottom up
 - Has the advantage of involving new teams

Conclusion

- STEVIN has produced a valuable set of resources that many other languages (will) envy us and that will help produce state of the art research in *Dutch language* HLT research and development
 - This is already starting
- But please give us also a STEVIN now for **basic** research
 - Inference / reasoning / structured learning
 - World and background knowledge mining
 - Infrastructure projects
 - Discourse
 - Deep(er) text understanding
 - Multimodality (image processing, robotics)
 - Adaptation (including multilinguality)
 - ...



Thanks for your attention !

- Contact: walter.daelemans@ua.ac.be