Natural Language Applications in an RRG framework

Brian Nolan
Ricardo Mairal Usón
Carlos Periñán
www.lexicom.es
www.fungramkb.com

OUTLINE

- Challenges for NLP; the semantic web and intelligent agents.
- The historical background.
- A research computational agenda for RRG
- UniLing.
- Future prospects.

Challenges for NLP

- The Semantic Web.
- Need for language aware software applications.
- Intelligent agents: Crosslinguistic information retrieval.
- Make information accesible within a globalised multicultural and multilingual world.

Social need

- Drive our research towards real social needs.
- Migration of people and communities has reinforced the need for a shared understanding of information relevant and important to different people.
- UniLing Semantic Web application.

The historical background

- Computational RRG applications:
  - Rule-based lexicalist interlingua bridge MT.
  - Ontological engineering: CLS machine tractable.
  - The implementation of a unified lexical metalanguage in software.
  - The parsing of complex sentences.

Consequences

- Enrichment of RRG lexicon: Ontological semantics.
- Building of frame-based RRG applications in processing.
- Conclusion:

RRG has a positive and crucial role in NLP.

Credible and realistic linguistic model.

A computational RRG agenda

- Interlingua engine to provide automatic translation from Arabic to English, Spanish and German: UNILING.
- Library cataloguing on the web.
- Extending the concept ontology into terminological subontologies: factoid questions.
- Speech recognition engine to convert Arabic speech into a machine readable text.
- Computational forensics around RRG.

RRG computational implementation

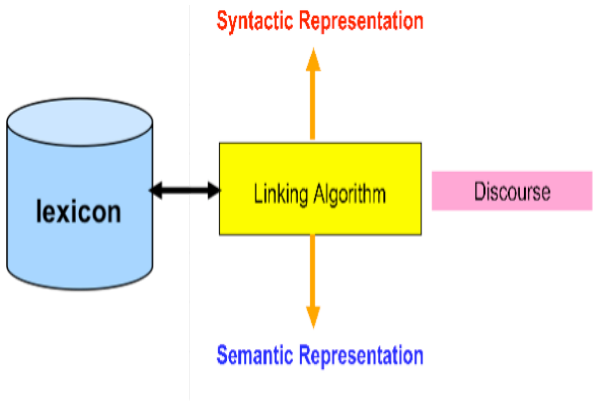
- Germany: Van;
  - Expertise in RRG.
- United Kingdom: Elizabeth Guest:
  - Automatic summarizer.
- Ireland: Brian Nolan:
  - Machine Translation Engine.
- Spain: Ricardo Mairal:
  - Lexical Conceptual Knowledge Base.

Nolan, B,; R. Mairal; C. Periñán “The UniLing Project”. *ITB Journal*.

UniLing: A computational implementation of RRG

- A follow-on project to two other projects: UnirArab and FungramKB.
- MT: Based on statistical methods of equivalent translation probability (Google, Microsoft etc.). Deep vs. Surface semantics.
- UniLing: to build linguistically based langauge-aware applications and across the Arabic (MSA), English, Spanish, German and Italian.

Components



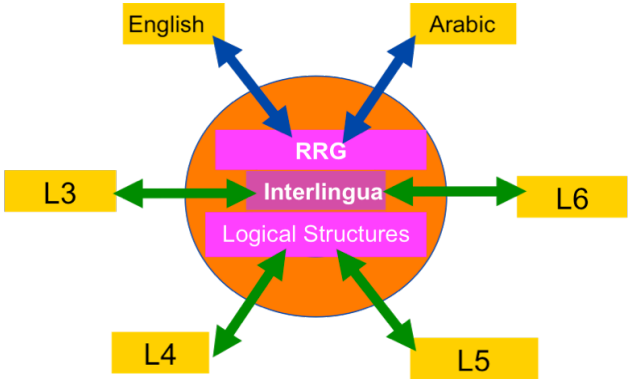
- UniArab: A machine translation system written in Java and XML.
- RRG and an Interlingua.
- FunGramKB (www.fungramkb.com): a multipurpose lexical conceptual knowledge base for NLP systems.

Advantages of UniLing

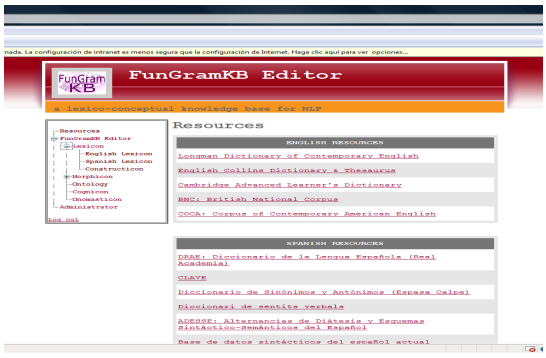
- Deployment of a linguistically motivated Interlingua Architecture for Machine Translation.
- Use of a strong and vibrant functional linguistic model rather than a statistical model in the MT kernel: RRG.
- Use of a lexico-conceptual Knowledge Base linked to the linguistic model.
- The quality of outputs in the domain of the language aware applications
- The range of coverage of human language within the linguistic model and the UniLing software.
- UniLing is by design Internet aware and Internet facing as a multilingual, language aware semantic web application

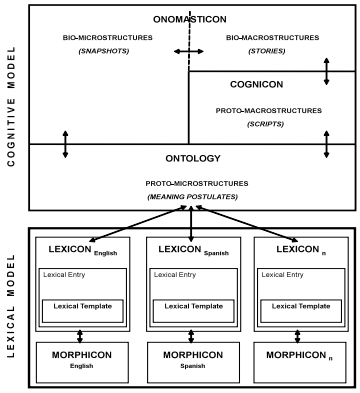
Evaluation I

- UniArab: Better results than statistical based approaches (e.g. Google, Microsoft).
- A proof-of-concept system supporting the fundamental aspects of Modern Standard Arabic. (Nolan and Salem, 2009).



Evaluation II

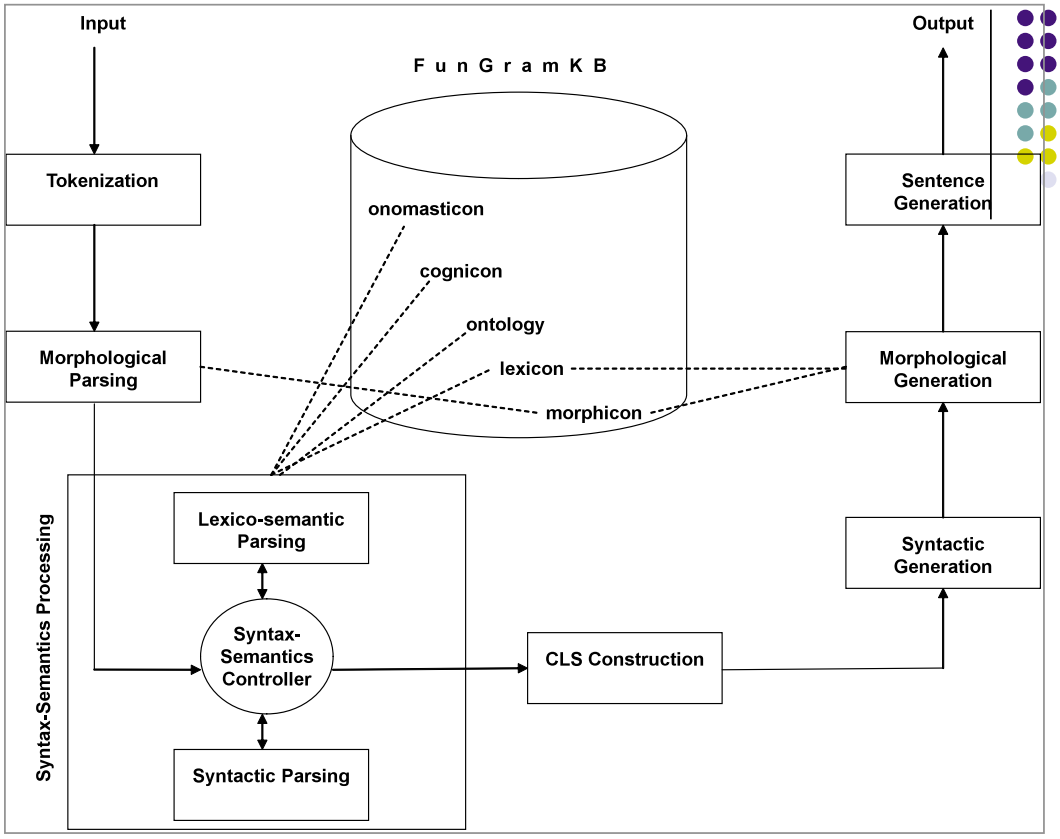




- FungramKB.
- Ontological Semantics.
- Cognitive Level: Ontology, Cognicon and Onomasticon.
- Lexical Conceptual Linkage: CLS.
- Information retrieval: MicroKnowing.

Methodology

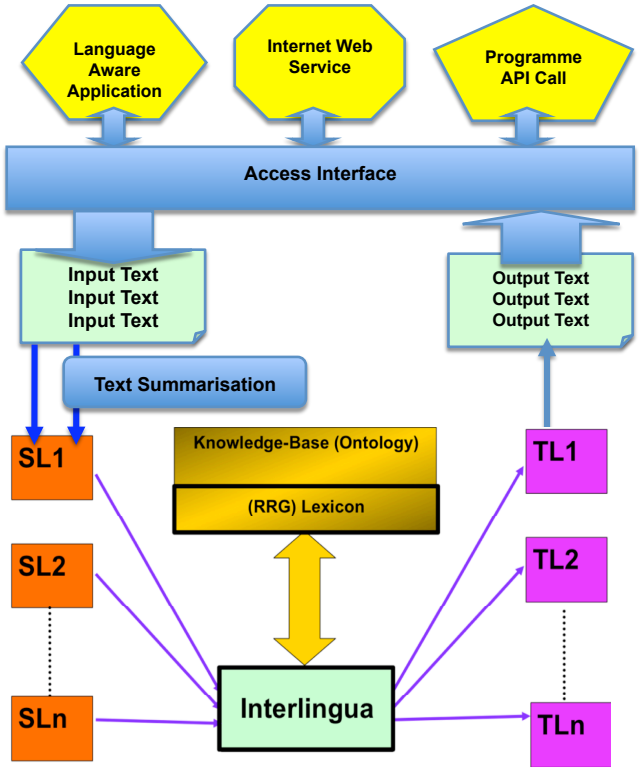
- Task (1): Tokenization.
- Task (2) Morphological parsing.
- Task (3) Syntax-Semantics Processing.
- Task (4) CLS Construction.
- Task (5) Syntactic Generation.
- Task (6) Morphological Generation.
- Task (7) Sentence Generation.



Linguistic domains

- Legal language incorporating family and inmigration.
- Summarisation and visualisation of textual information from other legal jurisdictions.
- Areas of intelectual property law, trade and commerce.

The system architecture and interfaces of UniLing



Deliverables

- Language applications for government and security use.
- Linked and wrapped with speech recognition and text to speech synthesis.
- Dissemination of activities.
- Make communication easier in a globalised multicultural and multilingual context.

Thanks a lot for your attention