



Towards HLT Evaluation at the University of Malta

Michael Rosner

`mike.rosner@um.edu.mt`

ELRA Workshop, Malta

December 1st 2005



University of Malta

- Founded 1592
- 10,000 students
- 500 academics
- CSAI – approx 400 students
- Students specialising in HLT: 10-20



HLT @ University of Malta

CS & AI	Inst of Ling	Comp. Eng.
M. Rosner A. Dalli P. Farrugia S. Calleja D. Attard C. Scicluna	R. Fabri A. Gatt A. Vella	P. Micallef



The Language

- A minority language (but not in Malta!)
- Maltese: 1M native speakers
- Became an official European language in 2003
- Translation issue
- A 'mixed' language with Semitic; Romance; English substrates
- Rich morphology
- Mainly Latin alphabet



The Language

- Inflectional Complexity
- Derivational Complexity
- Typical Semitic language features
 - root and template morphology
 - “construct state” for expressing inalienable possession (kelb il-baħar)
- Mostly Latin alphabet



Inflectional Complexity

gh - m - l

a - e



ghamel

nagħmluhila

n + a + ghml + u + hi + la

(“we make it for her”)



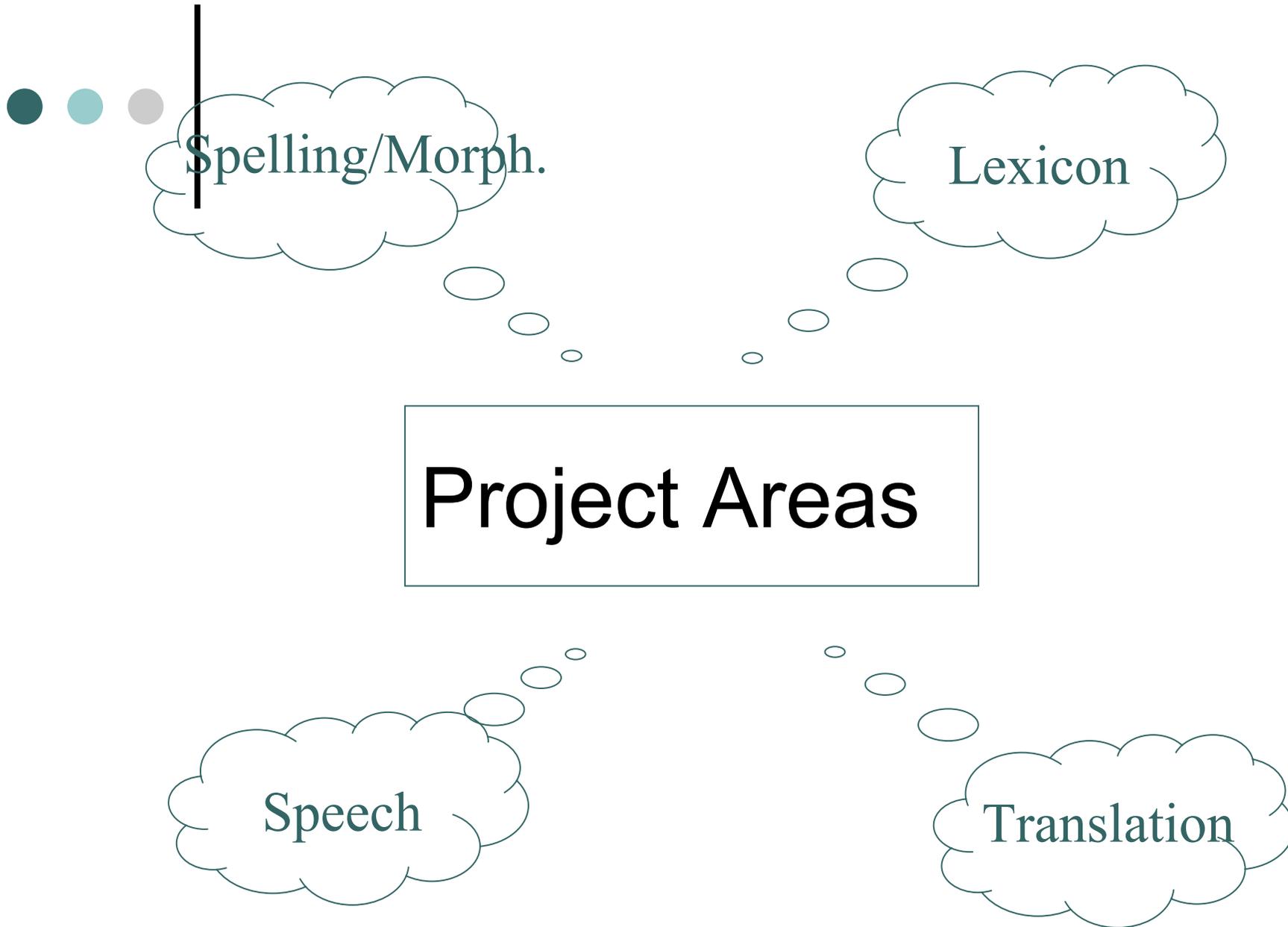
Issues for Handling Maltese

- Need for basic computational infrastructure
 - Language Resources
 - Spell/Style checking
 - Keyboard
- Mixed language input



Obstacles

- Cultural resistance: preference for using English or mixed language in computer-based communications (e.g. email) and in interfaces.
- Lack of formalised linguistic knowledge as input for language-sensitive support
- Difficulty of obtaining representative corpora.
- Lack of research continuity



Spelling/Morph.

Lexicon

Project Areas

Speech

Translation



Spell Checking/Morphology

- Prolog model of verb morphology (Galea 1996)
- Rule-based spelling error detection model (Mangion 1999 BSc)
- Probabilistic spelling error detection (Mizzi 2000 BSc)
- Initial development of xfst model (Rosner 2003)
preposition + article + noun: fil + L + skola
fl-iskola



Lexicon: Maltilex

- Broad coverage computational lexicon for Maltese
- Internet-based lexicographers interface
- Wordlist extraction from text
- LST – Lexical Structuring Technique (Dalli 2002) – semi automatic headword discovery.
- MLRS



Speech

- **Text-to-Speech**

- P. Micallef (1996) – TTS
- Paulseph Farrugia (2006). Mixed language TTS for SMS messaging;

- **Speech-to-Text**

- for numerical expressions, Sinclair Calleja (2004)



Translation

- Weather Forecasts (Farrugia, 2000) – based on PATR-like transfer rules.
- Alignment methods for legalistic bitexts (J. Bajada, 2005).



Current Developments

○ **National**

- MLRS: RTDI National Project (2005-2007)
- Statistical Translation
- National Council for the Maltese Language

○ **European**

- LT4EL (kicking off) – multilingual information retrieval for eLearning



National Council for the Maltese Language

COUNCIL (11 members)



TECHNICAL COMMISSION (heads of committees)



TECHNICAL COMMITTEES (currently 5)



VARIOUS SECTORS OF SOCIETY



NCML: Technical Committes

- Linguistic Research (Prof. Albert Borg)
- Terminology (Prof. Manwel Mifsud)
- Literature (Dr. Adrian Grima)
- Media (Mr. Charlot Bonnici)
- IT (Dr. Ray Fabri)



IT Technical Committee

- Provide advice/know how
- IT terminology
- Projects
 - National Conference
 - Workshop
 - Booklets for Maltese IT in schools
 - Ensure availability of basic tools



Evaluation

- **Mixed language systems**
- **Watch this space...**



DINNER THURSDAY – MEDINA RESTAURANT, MDINA

Terrine of Parma Ham with chicken liver pate with a hint of mango chutney, garnished with carved fruit

or

Traditional Maltese **Fish Soup**

or

Penne served with aubergines, mushrooms, bell peppers, basil and toasted pine nuts

~~o~~

Awrat (**sea bream**) with herbs and white wine

or

Fillet of **Beef** in Port wine jus

or

Stuffed **Avocado** with nuts, cheese and tomato concasse

French Beans rolled in Bacon

&

Baked Potatoes

~~o~~

Amaretto **Gateau**

or

Carpaccio of **Apples** with a hint of cinnamon & vanilla ice-cream

~~o~~

Coffee

Bus leaves at 19.45 outside Hotel – return c. 23.30