A computational morphology for Maltese
and using it to build a full-form lexicon

John J. Camilleri
Chalmers | University of Gothenburg

The Maltese language
- Language of Malta, one of 23 official E.U. languages
- Approx 400,000 native speakers
- At risk of "digital extinction" [1]

Make-up
- Classified as Semitic → Siculo-Arabic
- Latin alphabet
- Heavy Romance influence (~50% of vocabulary)
- More recent introduction of English words (~10%)

Typical sentence

Def article + N Pl
From Sicilian **cassata** "cheesecake"
Gloss "the cheesecakes"

Prep + Def article + N Pl
From English **fridge**
Gloss "from the fridges"

Oħroġli l-qassatat mill-friġġijiet jekk jogħġbok!

V Imperative P2 Sg + I.O. P1 Sg
Semitic root **ħ-r-ġ** "go out"
Pronominal suffix **-li** "for me"
Gloss "take out for me"

Prep + V Imperfect P3 Masc Sg + D.O. P2 Sg
Semitic root **għ-ġ-b** "like"
Pronominal suffix **-ok** "you (Sg)"
Gloss "if it is liked by you" ("please")

Verb morphology
Combination of root-and-pattern and affix morphology
Verb classes
- Semitic (Tri/Quadri-consonantal), e.g. *ħarab* (**ħ-r-b**) "escape"
- Integrated Romance verbs, e.g. *kanta* (**k-n-t-j**) "sing"
- Loan verbs, e.g. *pparkja* "park"

Computational morphology
- Built using the Grammatical Framework (GF), to be eventually included in GF Resource Grammar Library
- 9 root-and-pattern paradigms, 1 loan paradigm
- The verb inflects for:
  - Tense/aspect (Perfective, Imperfective, Imperative)
  - Subject (7/2 cases)
  - Direct object (7 cases)
  - Indirect object (7 cases)
  - Direct+Indirect object (3×7 cases)
  - Polarity (2 cases)
- Total size of single table: **952** forms (excludes derived verbs, which convey transitivity, reflexivity etc.)
- 73 manually-checked treebanks for testing

Inflection table excerpt
GF command: `mkV "ħareġ"`

```
s . VPerf (AgP3Sg Masc) => VSuffixNone => Pos => ħareġ
s . VPerf (AgP3Sg Masc) => VSuffixNone => Neg => ħariġx
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP1 Sg) => Pos => ħariġni
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP1 Sg) => Neg => ħariġnix
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP1 Pl) => Pos => ħariġna
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP1 Pl) => Neg => ħariġniex
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP2 Sg) => Pos => ħarġek
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP2 Sg) => Neg => ħarġekx
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP2 Pl) => Pos => ħariġkom
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP2 Pl) => Neg => ħariġkomx
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP3Sg Masc) => Pos => ħarġu
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP3Sg Masc) => Neg => ħarġux
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP3Sg Fem) => Pos => ħariġha
s . VPerf (AgP3Sg Masc) => VSuffixDir (AgP3Sg Fem) => Neg => ħariġhiex
s . VPerf (AgP3Sg Masc) => VSuffixDir AgP3Pl => Pos => ħariġhom
s . VPerf (AgP3Sg Masc) => VSuffixDir AgP3Pl => Neg => ħariġhomx
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP1 Sg) => Pos => ħariġli
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP1 Sg) => Neg => ħariġlix
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP1 Pl) => Pos => ħarġilna
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP1 Pl) => Neg => ħarġilniex
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP2 Sg) => Pos => ħariġlek
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP2 Sg) => Neg => ħariġlekx
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP2 Pl) => Pos => ħarġilkom
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP2 Pl) => Neg => ħarġilkomx
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP3Sg Masc) => Pos => ħariġlu
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP3Sg Masc) => Neg => ħariġlux
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP3Sg Fem) => Pos => ħarġilha
s . VPerf (AgP3Sg Masc) => VSuffixInd (AgP3Sg Fem) => Neg => ħarġilhiex
s . VPerf (AgP3Sg Masc) => VSuffixInd AgP3Pl => Pos => ħarġilhom
s . VPerf (AgP3Sg Masc) => VSuffixInd AgP3Pl => Neg => ħarġilhomx
s . VPerf (AgP3Sg Masc) => VSuffixDirInd (GSg Masc) (AgP1 Sg) => Pos => ħariġhuli
s . VPerf (AgP3Sg Masc) => VSuffixDirInd (GSg Masc) (AgP1 Sg) => Neg => ħariġhulix
```

Existing LT resources for Maltese

META-NET	Quantity	Availability	Quality	Coverage	Maturity	Sustainability	Adaptability
Language Technology: Tools, Technologies and Applications							
Speech Recognition	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Speech Synthesis	2.4	0.8	3.2	3.2	2.4	2.4	2.4
Grammatical analysis	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Semantic analysis	0	0	0	0	0	0	0
Text generation	0	0	0	0	0	0	0
Machine translation	1.6	1.6	1.6	1.6	1.6	1.6	1.6
Language Resources: Resources, Data and Knowledge Bases							
Text corpora	3.2	3.2	2.4	2.4	2.4	3.2	3.2
Speech corpora	2.4	0.8	2.4	1.6	2.4	2.4	2.4
Parallel corpora	3.2	3.2	2.4	1.6	1.6	1.6	1.6
Lexical resources	2.4	2.4	1.6	2.4	2.4	2.4	2.4
Grammars	0	0	0	0	0	0	0

State of language technology support for Maltese (scale from 0-9) [1]

Bootstrapping new linguistic resources
- No full-form lexicon exists for Maltese, not even in printed form (unfeasible by non-automatic means)
- Even native speakers have trouble with orthography

Idea: combine compiled list of root-and-pattern verbs (closed list of ~4,000 verb forms) [2] with GF grammar to produce full-form lexicon of ~4,000,000 forms
- Morphological analyser, spelling correction

A collaborative resource
- Rule-based morphology may over-generate
- At best an approximation of real-world language use
- How to keep up with newly imported words

Idea: create an entirely open online resource, allowing user feedback for marking dubious forms and adding new [loan] verbs to lexicon
- Crowd-sourced evaluation of morphological grammar, and overall maturity of resource itself

Current stage: exploring options for formats and standards, others' experiences with collaborative lexica

Sources
[1] M. Rosner and J. Joachimsen, *The Maltese Language in the Digital Age*. META-NET White Paper Series. Springer, 2012.
[2] M. Spagnol, *A Tale of Two Morphologies: Verb structure and argument alternations in Maltese*. Germany: University of Konstanz dissertation. 2011.

MOLTO: Multilingual On-Line Translation
This work has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. FP7-ICT-247914.