



The Irish & Welsh Languages in the Digital Age

John Judge/Jeremy Evas

CNGL, Ireland/School of Welsh, Cardiff University, Wales

jjudge@computing.dcu.ie/evasj@cf.ac.uk

Unlocking the Potential of Technology to Promote CRSS Language
Usage

Cardiff, Wales – January 23, 2014



Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

Outline

- ❑ Main Findings of the LWPs
- ❑ How does this compare?
- ❑ Current and Future Developments

The Irish & Welsh Languages in the Digital Age

Some Key Findings

<http://www.meta-net.eu>

LT Landscape in Ireland

- ❑ Historically strong in Translation and Localisation industry
- ❑ Home to several internationally recognised research centres
 - NCLT
 - DERI
 - CNGL
 - INSIGHT
- ❑ Government funding for research has been consistent despite worsening economic conditions

LT for Irish and Welsh

- ❑ Many of the basics are covered
 - Spell checker
 - Grammar checking
 - T9 predictive text, smartphone predictive text (through additional software, and by agreement of manufacturer)
 - Localisation of open source software, and many major applications
 - Translation technology used (but could be used more)

- ❑ Some of the more advanced stuff
 - Speech synthesis
 - Machine translation
 - Basic speech recognition
 - Part-of-Speech Tagger
 - (Dependency Parser)

LT for Irish & Welsh

- ❑ But we need more
- ❑ And what's already there needs more use at the micro level
- ❑ Availability of text corpora, speech corpora, parallel texts, wordnets and other LT building blocks is limited
- ❑ Some resources exist – small, narrow coverage, restricted availability
- ❑ Lack of linguistic resources is stifling development of modern language processing technologies for our languages
- ❑ Irony: Ireland's own research centres are producing world leading LT for other languages!



The Irish & Welsh Languages in the Digital Age

How Do Irish and Welsh Measure Up? Cross Language Comparison



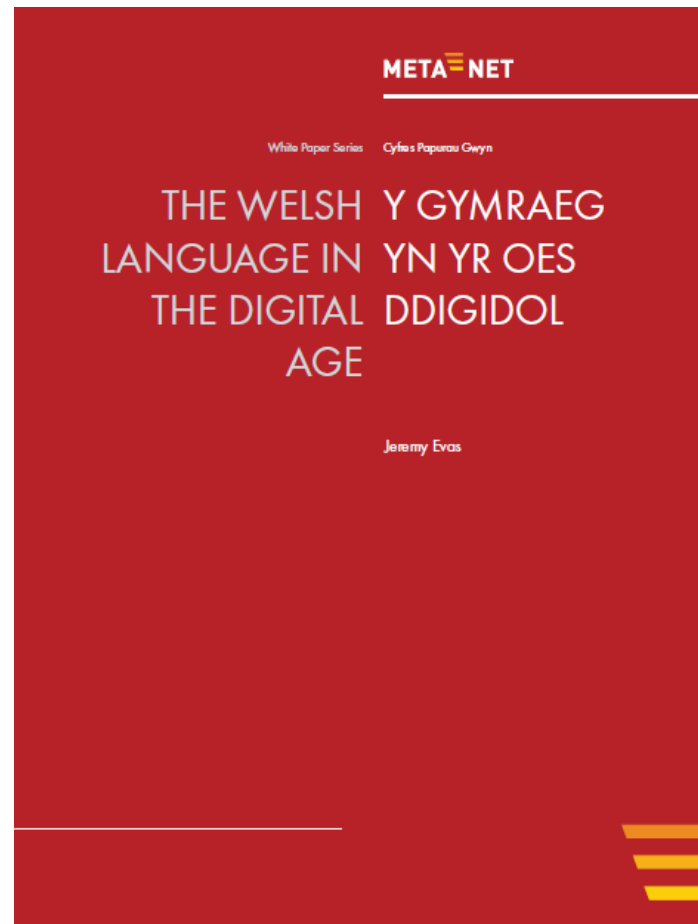
Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

State of LT support for Irish

	Quantity	Availability	Quality	Coverage	Maturity	Sustainability	Adaptability
Language Technology: Tools, Technologies and Applications							
Speech Recognition	0	0	0	0	0	0	0
Speech Synthesis	3	3	3	3	2	1	2
Grammatical analysis	4	4	3	3	4	4	4
Semantic analysis	2	2	0	0	2	2	1
Text generation	1	0	1	1	1	1	1
Machine translation	2	3	1	1	1	1	1
Language Resources: Resources, Data and Knowledge Bases							
Text corpora	3	4	3	3	4	3	3
Speech corpora	1	2	3	1	3	3	3
Parallel corpora	2	3	2	2	3	2	3
Lexical resources	4	3	4	3	3	3	3
Grammars	1	3	3	2	2	2	2

State of LT support for Welsh

META[≡]NET



Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

MT

excellent	good	moderate	fragmentary	weak or no support
	English	French, Spanish	Catalan, Dutch, German, Hungarian, Italian, Polish, Romanian	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Welsh, Irish , Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Serbian, Slovak, Slovene, Swedish

Text Analysis

excellent	good	moderate	fragmentary	weak or no support
	English	Dutch, French, German, Italian, Spanish	Basque, Bulgarian, Catalan, Czech, Danish, Finnish, Galician, Greek, Hungarian, Norwegian, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish	Croatian, Welsh, Irish , Estonian, Icelandic, Latvian, Lithuanian, Maltese, Serbian

Speech

excellent	good	moderate	fragmentary	weak or no support
	English	Czech, Dutch, Finnish, French, German, Italian, Portuguese, Spanish	Basque, Bulgarian, Catalan, Danish, Estonian, Galician, Greek, Hungarian, Irish , Norwegian, Polish, Serbian, Slovak, Slovene, Swedish	Croatian, Welsh , Icelandic, Latvian, Lithuanian, Maltese, Romanian

Resources

excellent	good	moderate	fragmentary	weak/no support
	English	Czech, Dutch, French, German, Hungarian, Italian, Polish, Spanish, Swedish	Basque, Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Galician, Greek, Norwegian, Portuguese, Romanian, Serbian, Slovak, Slovene	Icelandic, Welsh, Irish , Latvian, Lithuanian, Maltese



Europe's Languages and LT



English

Dutch
French
German
Italian
Spanish

Catalan
Czech
Finnish
Hungarian
Polish
Portuguese
Swedish

Basque
Bulgarian
Danish
Galician
Greek
Norwegian
Romanian
Slovak
Slovene

Croatian
Estonian
Icelandic
Irish
Welsh
Latvian
Lithuanian
Maltese
Serbian

*good support through
Language Technology*



*weak or
no support*



The Irish & Welsh Languages in the Digital Age

Current and Future Developments



Recent Developments

- White Paper Publication +12 Months
 - Significant media response to publication (online, offline, broadcast, print, social)
 - Increased profile for LT for Irish & Welsh in the public domain
 - Ministerial endorsement
- Further investment in R&D centres in Ireland developing LT as well as targeted projects
 - CNGL
 - INSIGHT
- CIGILT Group and the Celtic Language Technologies Workshop

Future

- ❑ Clear demand for LT for Irish and Welsh
 - Official Languages Act (Ireland)
 - Forthcoming Codes of Practice for implementation of language standards under Welsh Language (Wales) Measure 2011
 - Welsh Government Language Strategies
 - Mainstreaming of language policy into all fields
 - Education
 - Driving language revitalisation

- ❑ Scientific expertise and strong industry links



Future Barriers to progress

- ❑ Current barriers to R&D include
 - Knowledge transfer
 - Continuity of research
 - Lack of coordination with what is happening at EU level
 - Language expertise in research sector

Future – Moving Forward

- ❑ To overcome these we must
 - Support Irish & Welsh LT research through dedicated mechanisms
 - Foster an attitude of open sharing of data and resources for research and implementation of projects
 - Look to EU instruments like SRA and H2020/CEF to secure same
 - Encourage greater uptake and engagement with technology in education at 3rd level in our speakers



Thank You!

Acknowledgements: This work would not have been possible without the dedication and commitment of our colleagues Georg Rehm, Aljoscha Burchardt, Kathrin Eichler, Tina Klüwer, Arle Lommel, Felix Sasaki and Hans Uszkoreit (all DFKI), the 60 member organisations of the META-NET network of excellence, the ca. 70 members of the Vision Groups, the ca. 30 members of the META Technology Council, the more than 200 authors of and contributors to the META-NET Language White Paper Series and the ca. 200 representatives from industry and research who contributed to the META-NET Strategic Research Agenda.

jjudge@computing.dcu.ie

ie.linkedin.com/in/judgejohn/

<http://www.cngl.ie>

evasj@Cardiff.ac.uk

@jeremyevas

<http://www.meta-net.eu>

<http://www.facebook.com/META.Alliance>

