

lemonGAWN

WordNet Gaeilge as Linked Data

Jim O'Regan, Kevin Scannell, Elaine Uí Dhonnchadha

Trinity College, Dublin
Saint Louis University, Missouri, USA

24 May 2016

Outline

- 1 Why?
- 2 What?
- 3 Some Numbers
- 4 What's next?

Why? – Honest reason

- 1 I didn't have wifi at TCD for my first week
- 2 I did have the data

Why?
What?
Some Numbers
What's next?

LINKING



memegenerator.net

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

Why?

What?

Some Numbers

What's next?

Why? – Retrospective reasons

Irish–English machine translation needs semantic classes

Why?

What?

Some Numbers

What's next?

Why? – Retrospective reasons

Irish–English machine translation needs semantic classes
English to Irish: numbers

Why? – Retrospective reasons

Irish–English machine translation needs semantic classes
English to Irish: numbers different for humans

Why? – Retrospective reasons

Irish–English machine translation needs semantic classes
English to Irish: numbers different for humans
Irish to English: structural ambiguity

Why? – Retrospective reasons

Irish–English machine translation needs semantic classes

English to Irish: numbers different for humans

Irish to English: structural ambiguity

Tá sé ina theach

He is in his house

Tá sé ina dhochtúir

He is a doctor

Why? – Retrospective reasons

Subword modelling for MT: lexicon expansion through affixes

Outline

- 1 Why?
- 2 What?
- 3 Some Numbers
- 4 What's next?

Why?

What?

Some Numbers

What's next?

What? – WordNet Gaeilge

- Created by Kevin Scannell

What? – WordNet Gaeilge

- Created by Kevin Scannell
- Irish synonyms attached to English sense IDs

What? – WordNet Gaeilge

- Created by Kevin Scannell
- Irish synonyms attached to English sense IDs
- **NOT** synsets! – semi-lexical senses

What? – WordNet Gaeilge

- Created by Kevin Scannell
- Irish synonyms attached to English sense IDs
- **NOT** synsets! – semi-lexical senses
- Includes items which have no PWN equivalent

Outline

- 1 Why?
- 2 What?
- 3 Some Numbers**
- 4 What's next?

Some Numbers

Wordnet Gaeilge “synsets”	77814
Missing from PWN	28356
Missing Irish label	5936
Nouns	49889
Verbs	11548
Adjectives	15250
Adverbs	1127

Table: WordNet Gaeilge “synsets”

has_hyperonym	14965
has_hyponym	14965
has_mero_madeof	151
has_mero_member	295
has_mero_part	1617
has_subevent	155
is_caused_by	55
is_derived_from	360
is_subevent_of	155
near_antonym	1818
near_synonym	2946

Table: Synset relations obtained by projection from EuroWordNet (MCR).

region	155
region_term	155
related_to	16590
see_also_wn15	1684
state_of	472
sumo_at	1364
sumo_equal	749
sumo_plus	25503
topConcept	79757
usage	160
usage_term	160
verb_group	250

Table: Synset relations obtained by projection from EuroWordNet (MCR), pt.

2

Outline

- 1 Why?
- 2 What?
- 3 Some Numbers
- 4 What's next?

What's next

- 1 Hosting the data!

What's next

- 1 Hosting the data!
- 2 Irish to other languages (via Open Multilingual Wordnet)

What's next

- 1 Hosting the data!
- 2 Irish to other languages (via Open Multilingual Wordnet)
- 3 Integration in CALL (Ní Chiaráin & Ní Chasaide, 2016)