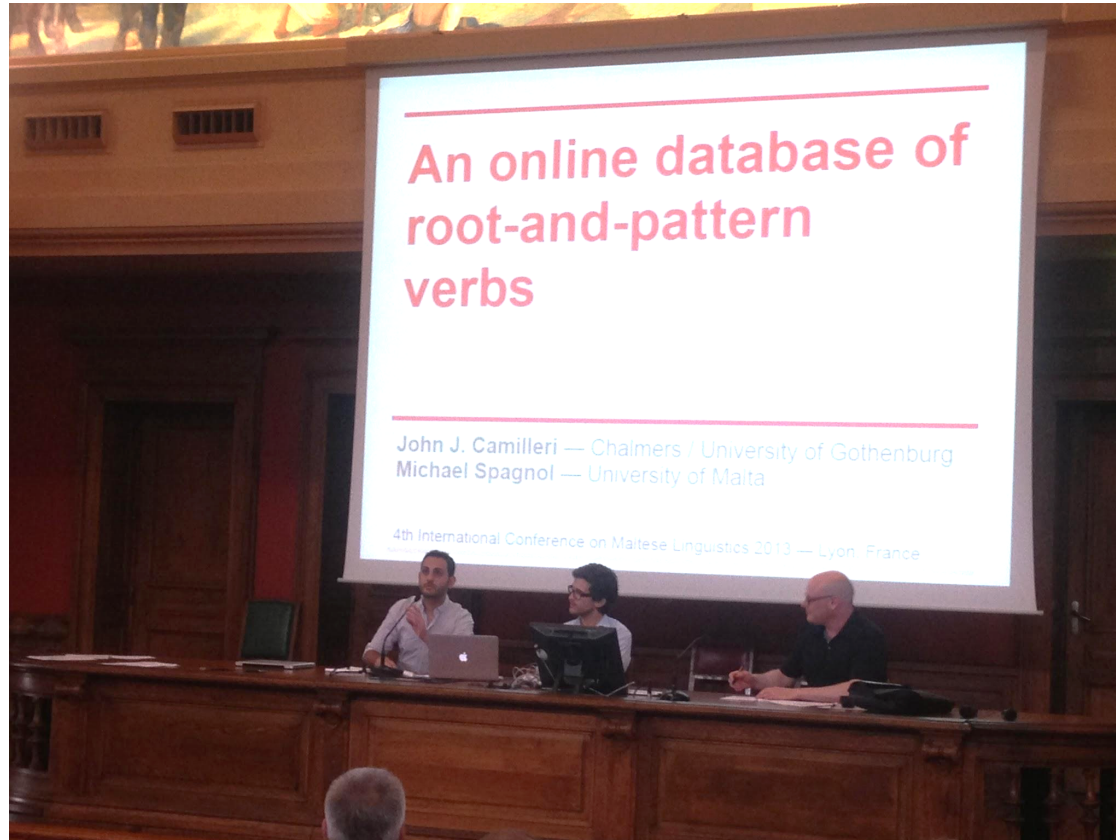

Ġabra: an open, online collection of Maltese lexical resources

John J. Camilleri

Chalmers & University of Gothenburg, Sweden

5th International Conference on Maltese Linguistics 2015 — Turin, Italy

Previously...



Digital lexical resources

1. Root-and-pattern verbs (Spagnol 2011)
 2. English-Maltese Dictionary (Falzon 1997/2013)
 3. Broken plurals (Mayer 2013)
 4. Verbal nouns (Ellul 2013?)
 5. Morphological generator (Camilleri 2013)
-

Il-Ġabra (the Collection)

A collection of Maltese Lexical Resources

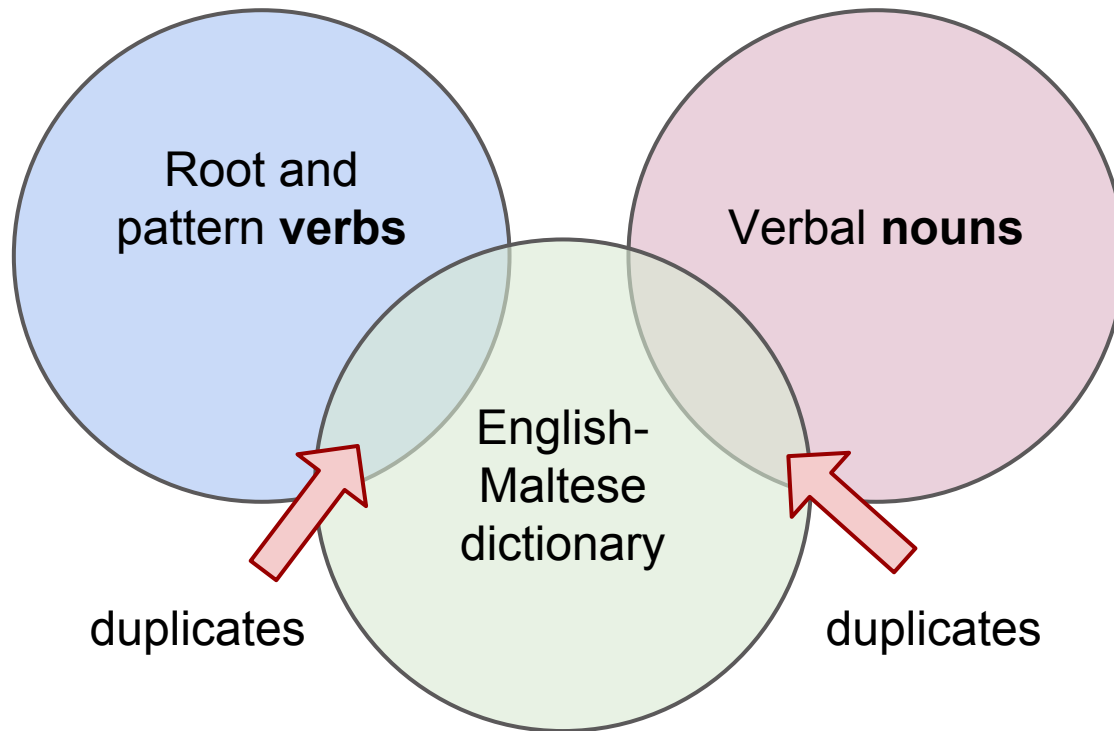
<http://mlrs.research.um.edu.mt/resources/gabra/>

- **Singular**: one database for all resources
 - **Online**: available, searchable
 - **Flexible**: homogenous, extensible
 - **Open**: accessible, usable
-

Putting it together

- Digital sources are processed automatically
 - Each has a different **format**
 - e.g. Excel, CSV, XML
 - Customised script for each source
 - Each has a different **structure** (schema)
 - Different fields
 - Need general structure which is flexible
-

Merging sources



Flexible, schema-less database

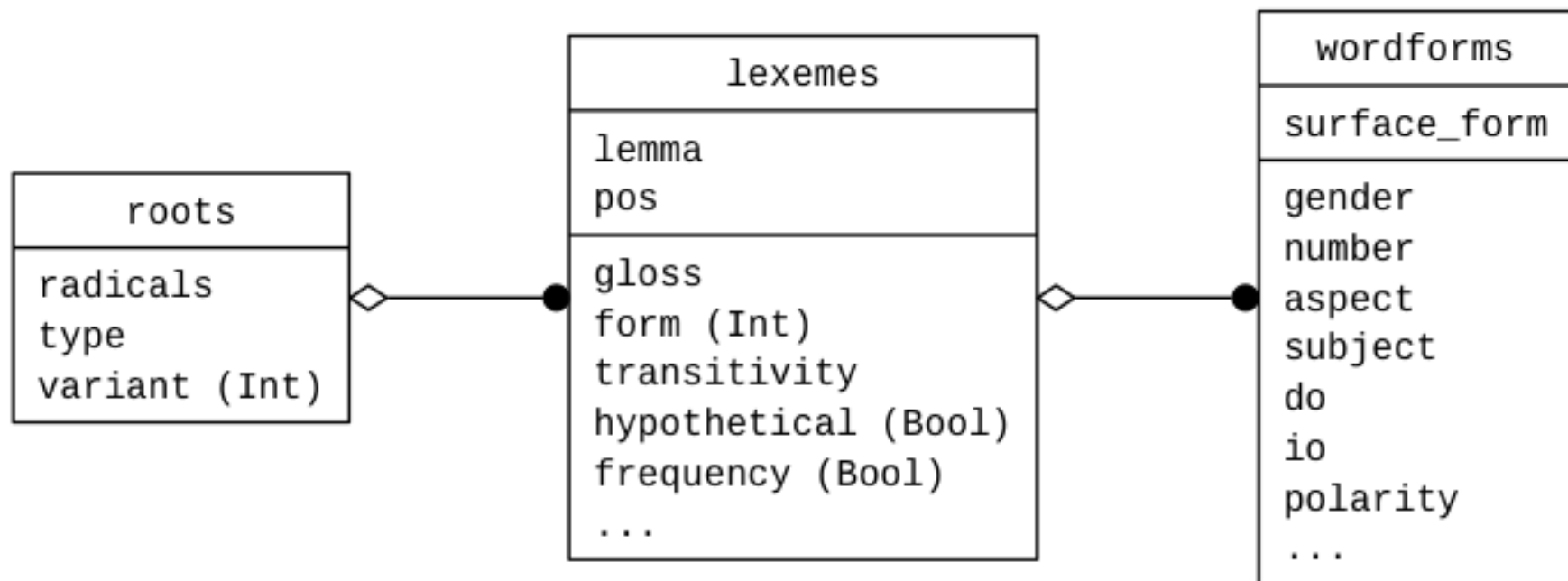
Relational database

| Lemma | POS | Root | Gender | Derv. |
|---------|-----|-------|--------|-------|
| ħarga | N | ħ-r-ġ | f | - |
| qabbel | V | q-b-l | - | 2 |
| pparkja | V | - | - | - |

Schema-less database (JSON)

```
{ lemma : ħarga ,  
  pos   : N ,  
  root  : ħ-r-ġ ,  
  gender : f  
} ,  
{ lemma : qabbel ,  
  pos   : V ,  
  root  : q-b-l ,  
  derv  : 2  
} ,  
{ lemma : pparkja ,  
  pos   : V  
}
```

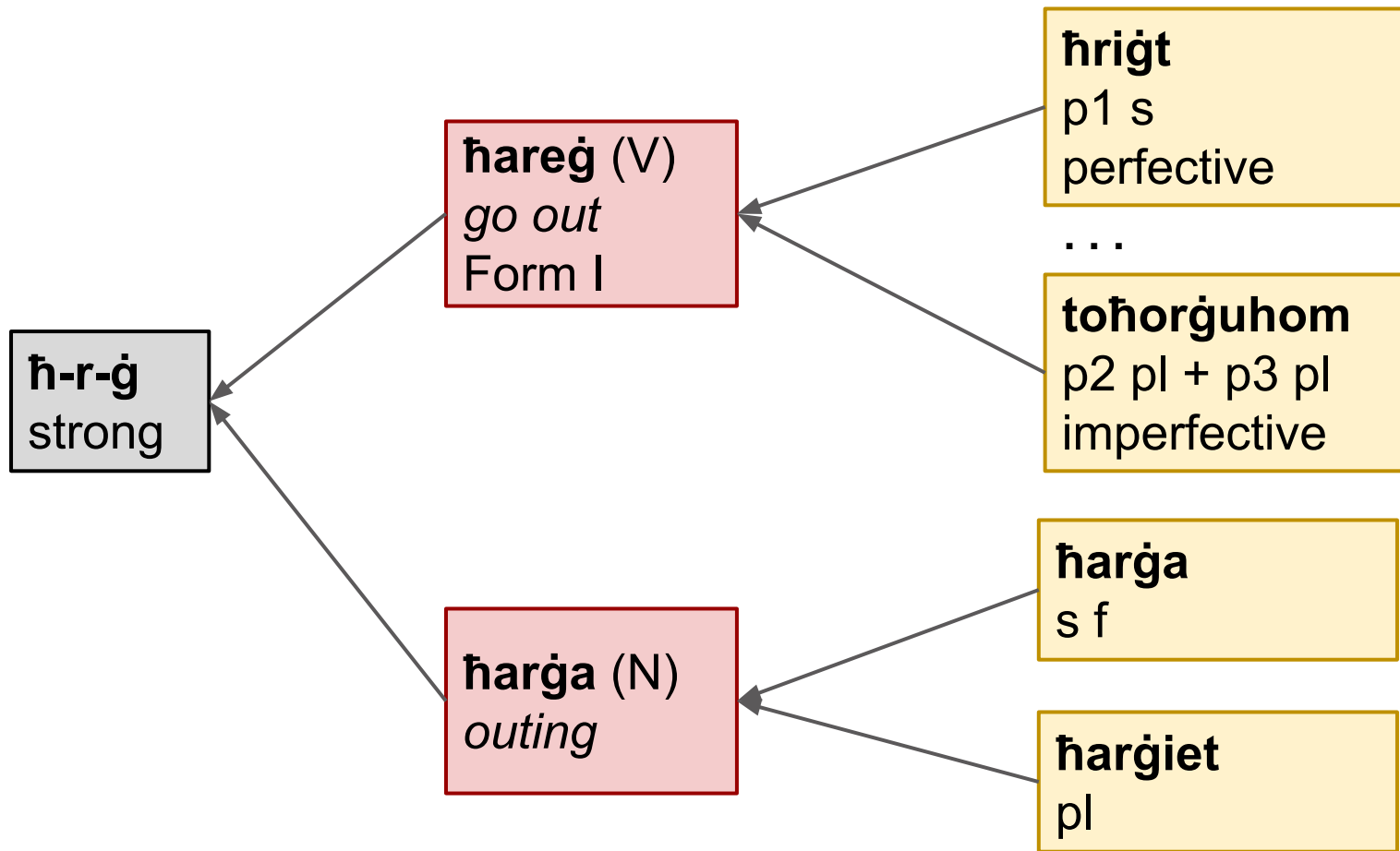
Top-level structure



Root

Lexeme

Word form

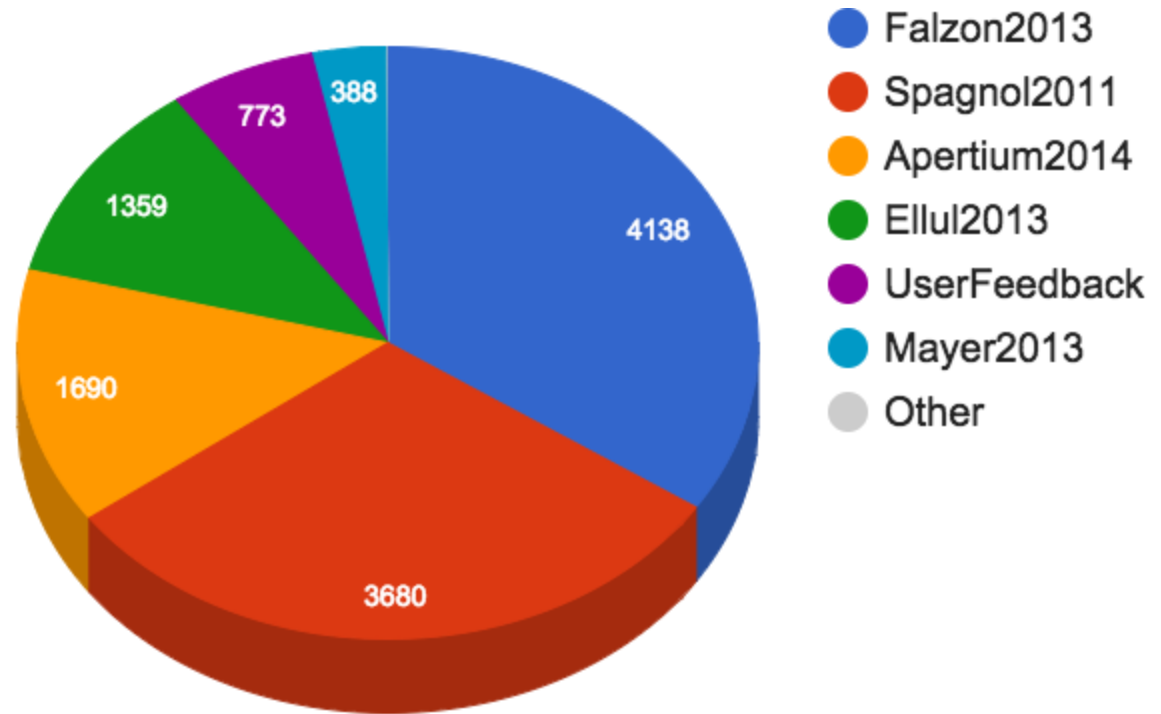


Demo!

Stats: lexemes

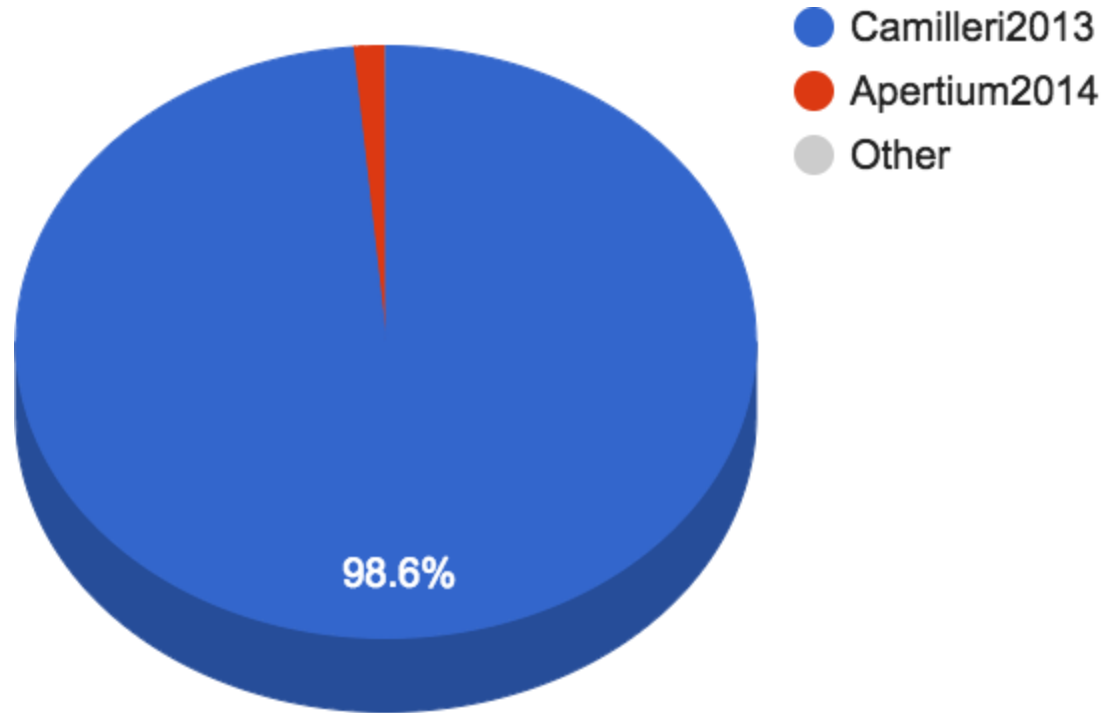
Total: 12,863

Glossed: 60%



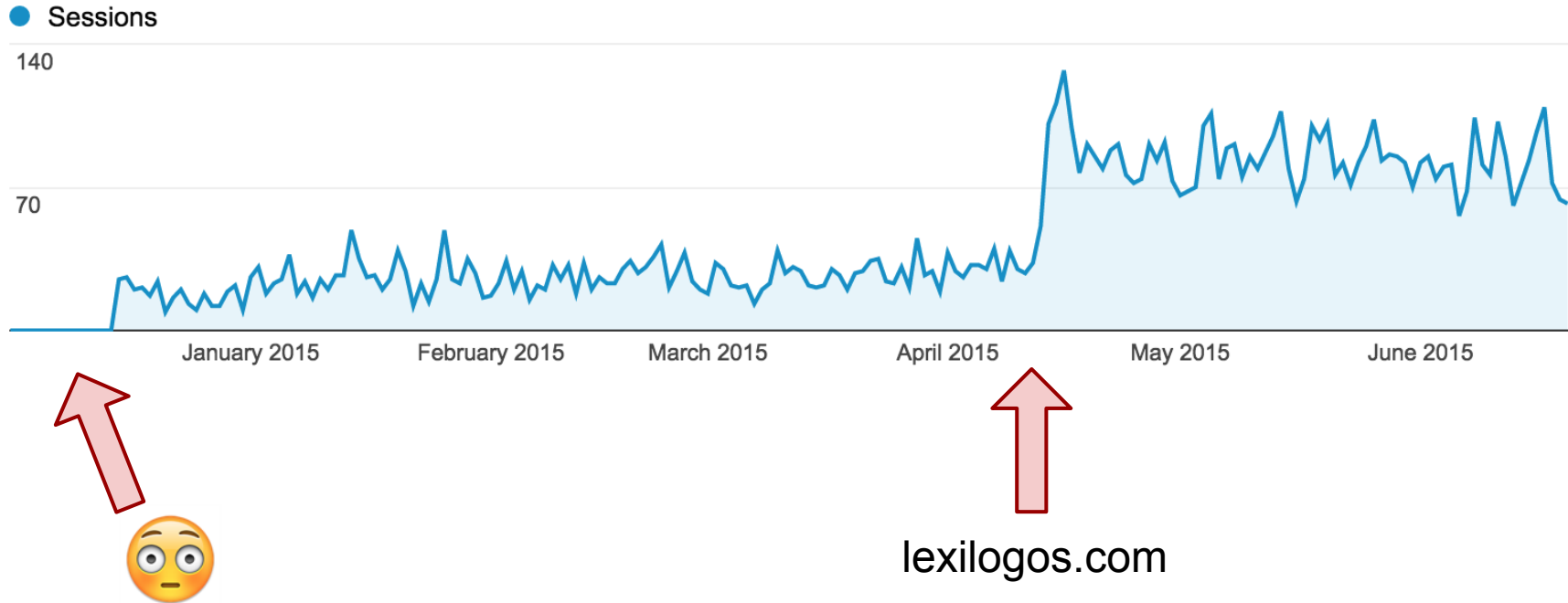
Stats: word forms

Total: 4,836,720













Website analytics: sessions


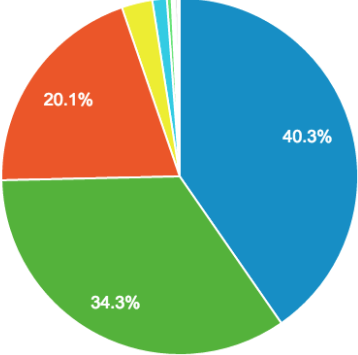









Approx 85 sessions per day



Website analytics: location

| Country ? | Sessions ? ↓ | % New Sessions ? | New Users ? | Bounce Rate ? | Pages/Session ? | Avg. Session Duration ? |
|---|--------------------------------------|--|--------------------------------------|--|------------------------------------|--|
| | 8,820 % of Total: 100.00% (8,820) | 55.36% Avg for View: 55.31% (0.10%) | 4,883 % of Total: 100.10% (4,878) | 39.85% Avg for View: 39.85% (0.00%) | 5.98 Avg for View: 5.98 (0.00%) | 00:06:48 Avg for View: 00:06:48 (0.00%) |
| 1.  Malta | 5,576 (63.22%) | 60.51% | 3,374 (69.10%) | 42.09% | 4.96 | 00:04:50 |
| 2.  United Kingdom | 731 (8.29%) | 31.74% | 232 (4.75%) | 36.80% | 4.37 | 00:06:00 |
| 3.  Australia | 605 (6.86%) | 46.61% | 282 (5.78%) | 30.58% | 12.03 | 00:21:18 |
| 4.  United States | 282 (3.20%) | 56.38% | 159 (3.26%) | 34.75% | 7.14 | 00:07:30 |
| 5.  Germany | 218 (2.47%) | 35.32% | 77 (1.58%) | 19.72% | 12.32 | 00:15:35 |
| 6.  France | 164 (1.86%) | 42.68% | 70 (1.43%) | 26.22% | 5.76 | 00:06:55 |
| 7.  Switzerland | 102 (1.16%) | 12.75% | 13 (0.27%) | 31.37% | 6.03 | 00:08:08 |
| 8.  Belgium | 99 (1.12%) | 47.47% | 47 (0.96%) | 35.35% | 3.24 | 00:02:49 |
| 9.  Sweden | 92 (1.04%) | 20.65% | 19 (0.39%) | 29.35% | 5.86 | 00:03:32 |
| 10.  Greece | 77 (0.87%) | 10.39% | 8 (0.16%) | 14.29% | 15.75 | 00:25:22 |

Website analytics: sources

| <input type="checkbox"/> | Source/Medium | Sessions <input type="text" value="Sessions"/> | Sessions | contribution to total: <input type="text" value="Sessions"/> |
|--------------------------|--|--|--------------------------------------|---|
| | | 8,972 % of Total: 100.00% (8,972) | 8,972 % of Total: 100.00% (8,972) | |
| <input type="checkbox"/> | 1.  lexilogos.com / referral | 3,620 | 40.35% |  |
| <input type="checkbox"/> | 2.  google / organic | 3,078 | 34.31% | |
| <input type="checkbox"/> | 3.  (direct) / (none) | 1,802 | 20.08% | |
| <input type="checkbox"/> | 4.  l.facebook.com / referral | 250 | 2.79% | |
| <input type="checkbox"/> | 5.  facebook.com / referral | 118 | 1.32% | |
| <input type="checkbox"/> | 6.  m.facebook.com / referral | 39 | 0.43% | |
| <input type="checkbox"/> | 7.  bing / organic | 14 | 0.16% | |
| <input type="checkbox"/> | 8.  cse.chalmers.se / referral | 13 | 0.14% | |
| <input type="checkbox"/> | 9.  yandex / organic | 9 | 0.10% | |
| <input type="checkbox"/> | 10.  normancborg.net / referral | 7 | 0.08% | |

Ğabra is OPEN

Open license

- Creative Commons Attribution (CC BY 4.0)
<https://creativecommons.org/licenses/by/4.0/>
 - You are free to:
 - **Share** — copy and redistribute in any format
 - **Adapt** — remix, transform, build upon the materialfor any purpose, even commercial
 - **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
-

Open access

- API
 - Direct access to data
 - For developers of websites, mobile apps
 - Download
 - Snapshot of entire database
 - For local use, batch processing
 - Requires knowledge of JSON/MongoDB
 - Schema is documented
-

Open to crowd-sourcing

- Gābra contains errors
 - Users can flag mistakes
 - Gābra is incomplete
 - Users can add new entries
 - Both require manual moderation
-

Open to new input

- Built for flexibility from day one
- Platform for **future** lexical resources
 - Wiktionary
 - Dictionary of old Maltese
 - Aquilina's dictionary?

Ultimate goal: online dictionary for Maltese...

The future of Ġabra

- Talks between:
 - ICT Committee (Kunsill tal-Malti)
 - University of Malta
 - Malta Communications Authority
 - Vodafone Foundation
 - Raised €15,000 from MCA & VF to invest in both **content** and **interface**
-

Content

- 2 RA's to be hired this summer at UoM
 - Fixing errors
 - Adding English glosses
 - New entries from social media
 - Moderating crowd feedback
-

Interface

- New web site design
 - Mobile apps
 - Promotion & exposure
 - Branding...
-

Dizzjunarju tal-Malti



Anton Buttigieg
1912 - 1983

Anton Buttigieg was a Maltese political figure and poet. He served as the second President of Malta from 1976 until 1981.



Search in Maltese or English



ADVANCED SEARCH

“Lexicon” vs. “Dictionary”

- Lexicon: a list of words
 - Useful for NLP applications
 - Full-forms practical in digital setting
 - Dictionary
 - Maltese-English
 - Maltese-Maltese
 - Etymologies
 - Examples of usage
-

Conclusion

- Ġabra was built quickly but is actively used
 - Flexible, open platform for lexical resources
 - Exciting new funding: **Dizzjunarju tal-Malti**
 - Ġabra will live on as the back end of DM
 - Always open to new data & contributions
 - Student projects?
-

Special thanks to Albert Gatt
