smg

Two Decades of Open Data for Language Diversity at the Surrey Morphology Group

UNIVERSITY OF SURREY

A collective work. Poster by Sacha Beniamine & Erich Round

2002 · SYNCRETISM — The Syntax–Morphology Interface · A Study of Syncretism · Matthew Baerman, Dunstan Brown, Greville G. Corbett

2011 · SAANICH VERB

2019 · SKOLT SAAMI PARADIGMS

2009 · MORPHOSYNTACTIC CHANGE

2019 · ENDANGERMENT IN SIBERIA

2003 · SUPPLETION · GO WENT

2002 · PERSON SYNCRETISM · -KIMBI -KIBI

2019 · CHICHIMEC PARADIGMS

2021 · LEXICAL SPLITS

2007 · DICTIONARY OF ARCHI

2019 · PROMINENT INTERNAL POSSESSORS

2015 · OTO-MANGUEAN INFLECTIONAL CLASS

2015 · MIAN & KILIVILA COLLECTION

2020 · NUER LEXICON

2009 · DEFECTIVENESS · Paradigms · Missing forms and what they tell us · ANDREW BAERMAN · R. G. CORBETT

2006 · DEPONENCY · Deponency and Morphological Mismatches

2008 · GRAMMATICAL FEATURES · Features · Greville G. Corbett

2010 · PERIPHRASIS · Periphrasis · The Role of Syntax and Morphology Paradigms

2002 · AGREEMENT · Agreement · Greville G. Corbett

2008 · OWNERS INTO ACTORS

Background
- In the 1990s most linguistic research was based on language data which were not publicly available.
- This made it impossible to verify theoretical claims.
- **SMG** committed instead to high-value databases that were freely accessible.

Methods
- Open data supporting replicability.
- Interactive visualizations of complex relationships.
- Dynamic engagement with language communities.
- DOIs and cite buttons promoting good citation practices.

Results
- 20 public databases created over 20 years.
- They document 325 languages around the world, across 100 families.
- Around 200 of these languages are vulnerable, endangered or extinct.
- Colleagues internationally report usage in teaching, research, and research design.
- Hundreds of citations.

Conclusions
- Early adoption of open data was one of the keys to SMG's success.
- Ensuring technical continuity is a challenge still awaiting a stable solution.
- We are always improving our open research practices, creating new standards and methodologies.

Many SMG members and collaborators have contributed to the databases: Matthew Baerman, Julia Barron, András Bárány, Oliver Bond, Dunstan Brown, Lydia Byrne, Lynne Cahill, Marina Chumakina, Greville Corbett, Gerrit Jan Dimmendaal, Nick Evans, Roger Evans, Sebastian Fedden, Timothy Feist, Edward Garrett, R. J. Hayward, Andrew Hippisley, Nicholas Jenkins, Elene Kalinina, Anna Kibort, Alexander Krasovitsky, Alison Long, Marianne Mithun, Irina Monich, Irina Nikolaeva, Enrique Palancar, William Palmer, Harley Quilliam, Tatiana Reid, Sandy Ritchie, Erich Round, Helen Sims-Williams, Carole Tiberius, Claire Turner.

https://www.smg.surrey.ac.uk/databases/