Computational approaches to morphological typology

Micha Elsner[1] *and Sacha Beniamine*[2]
[1] The Ohio State University
[2] University of Surrey

INTRODUCTION 1

Theories of morphology pertain to the lexicons of languages: what forms of words exist, how they relate to one another, and what they mean. To refine and test such theories, morphologists require high-quality information about lexicons, and where they posit particular learning mechanisms, these naturally operate on lexical knowledge to make their predictions. The size of a natural language lexicon, with its various quirks and irregularities in form and frequency, lends itself naturally to a databasing approach, and morphologists have a long history of productive engagement with computation.

The classification of languages into morphological types constitutes one of the earliest attempts to linguistic typology (von Schlegel 1818). As soon as 1960, Greenberg sought to objectivise these types by calculating indexes on corpora. In the past two decades, different strands of multi-variate morphological typology have converged to set the scene for scaling up morphological typology. The program of Canonical Typology (see among others Corbett 2005; Brown *et al.* 2012; Corbett 2023) has contributed to map out the space of typological variation in morphology and at its interfaces. Simultaneously, the program of Autotypology (see among others Bickel and Nichols 2002; Bickel *et al.* 2022; Witzlack-Makarevich *et al.* 2022) has supported the creation of large, interconnected typological databases, flexible enough to support diverse typological investigations. In inflection, the conversation on morphological complexity shifted gradually from the search of natural limits on morphological complexity (such as the Paradigm Economy Principle or the No Blur Principle, see Carstairs 1987; Cameron-Faulkner and Carstairs-McCarthy 2000), to the careful measure of this complexity, accompanied with a general turn towards Word & Paradigm approaches (Stump and Finkel 2013). Relying on quantitative analysis, Ackerman and Malouf (2013) describe two kinds of morphological complexity: enumerative (E-complexity) measuring how 'large' the system is and integrative (I-complexity) measuring its inter-predictability. Cotterell *et al.* (2019) conjecture that E- and I-complexity trade off against one another, so that languages with larger paradigms are easier to predict, and finds support for this proposal in a dataset of 36 Unimorph languages.

Two great endeavours underpin computational approaches to morphological typology: the elaboration of computational databases and the modelling of morphological systems based on this data. Constructing a computational database for a single language is a serious undertaking, so early studies often restricted themselves to a single language or a handful of related ones. Typological surveys, on the other hand, might be biased in the regions or language families they were able to cover, or might be forced to rely on unstandardised descriptions of different languages in which underlying similarities might be concealed by choices in analysis. Recent trends in morphological typology are striving to close this gap. Larger databases, representing more languages and phenomena, or connected together through standardisation and linked data, allow researchers to scale their modelling studies beyond the best-studied European languages. At the same time, modelling contributes to the standardization of typological description, by defining replicable measurements of theoretical constructs like 'zero markers', 'number of inflection classes' or 'inflection vs. derivation.' Thus, database construction and modelling are potentially synergistic activities which can feed one another, expanding our coverage of human languages while ensuring that our analytical constructs are valid.

While early morphological projects used small ad-hoc datasets or larger resources covering only one or two languages, recent projects have drawn on larger standardised resources. On the one hand, databases of inflected or derived forms document entire un-analyzed morphological systems. For example, Batsuren *et al.* (2022) provides structured lexical data for 169 languages in a unified format, and the Paralex standard (Beniamine *et al.* 2023) provides conventions to encode rich linguistic information concerning such inflectional resources. These databases of forms allow researchers to test conjectures about the statistics of lexicons at scale. On the other hand, databases of languages provide coded examples of a single phenomenon across many languages (Haspelmath *et al.* 2013; Bickel and Nichols 2002; Skirgård *et al.* 2023).

THE PAPERS IN THIS ISSUE 2

The first paper of this volume describes a novel cross-linguistic database, following the Autotyp approach. The three subsequent papers follow in the tradition of Ackerman and Malouf (2013) by proposing new models.

Inman & al: Alignment everywhere all at once: Applying the late aggregation principle to a typological database of argument marking 2.1

Inman *et al.* (2024) introduce the ATLAs Alignment Module, a typological database of argument marking at the morpho-syntactic interface, for languages of North and South America. The database is meant to capture the considerable language-internal variation in argument marking. It focuses on main declarative clauses with verbal predication and positive polarity. To a large extent, it conforms to the principes of Autotyp: it is *modular*, with each module covering a specific typological domain; variables and their values were kept open throughout coding (*autotypology*), ensuring detailed and faithful encoding. It enables *late aggregation*, where generalisations are not primary, but instead derived from data encoded at a granular level. Finally, it relies on *exemplars*. The database documents three argument roles (S, A, P) defined by semantics. Across languages, these roles can align together in various fashion, leading to basic alignment types.

For example, in a nominative-accusative alignment, roles S and A are aligned together, and distinctly from P, whereas in ergative-absolutive alignments, S and P are aligned together and contrast with A. Argument selectors are the devices by which arguments can be treated identically or differently, through either morphological marking or syntactic behaviour. Inman *et al.* (2024) focus on two types of selectors: flagging, which pertains to case marking and adposition within NPs, and indexing, which concerns verbal marking and agreement within clauses. The database is distributed in CLDF format, as a set of `csv` tables. It documents specific alignment contexts, the selectors involved, as well as the languages documented, the database source, and information aggregated automatically concerning references and alignment.

In short, Inman *et al.* (2024) present a wealth of precise data on alignment which can be aggregated at any documented level. It will enable testing numerous typological hypotheses, definitions, and operationalisations, much beyond those which were considered by the database authors.

2.2 *Becker: Zero marking in inflection: A token-based approach*

Becker (2024) tackles the challenge of observing the invisible. What is the typological distribution of zero markers? Do they behave like short markers, which, for reasons of coding efficiency, tend to be more frequent and predictable than longer markers (Zipf 2013; Greenberg 1966; Haspelmath 2008)? Becker surveys adjectival, nominal and verbal systems from 114 languages across six macro-areas. The data is derived from Unimorph (Kirov *et al.* 2016, 2018; McCarthy *et al.* 2020; Batsuren *et al.* 2022), with pre-processing to improve data quality and comparability, including conversion of some datasets to phonemic representations. Zero marking is unfortunately difficult to distinguish in a principled manner from the absence of a feature. Becker (2024) escapes this dilemna by adopting a Word & Paradigm perspective. She avoids morphemic segmentation altogether, and instead focuses on identifying stems automatically (following Beniamine and Guzmán Naranjo 2021; Bonami and Beniamine 2021, with some adjustments for stem allomorphy). She then defines zero-marked forms as those which consist solely of the stem. Similarly, features are not segmented, and zero markers are considered to mark the entire bundle of morpho-syntactic features for the form. To further reduce potential unfounded proliferation of zero marking, the study employs the perspective of *morphomic paradigms* (Boyé and Schalchi 2016), where any fully syncretic cells in the lexicon are merged.

Becker (2024) finds that overall, zero marking is uncommon. Yet, she observes a lot of variation across languages. Careful statistical analysis reveals this variation to be largely idiosyncrastic. A few trends emerge however: zero-marking is avoided in cells with many values; adjectives and verbs are more likely than nouns to avoid zero marking altogether. Some feature values are comparatively more likely to be zero marked across languages: IMP, SG, 3 and PRS in verbs, NOM, SG and INDF in nouns, NOM.SG in adjectives. Using the Universal Dependency corpora (Zeman *et al.* 2023) to gather frequency information, Becker (2024) confirms the Zipfian effect of frequency on length of overt markers, and finds the effect more pronounced on suffixes than other affixes. Nevertheless, this association does not hold for zero markers, which simply do not behave like short markers. Instead, she confirms the observation from Guzmán Naranjo and Becker 2021 according to which zero markers are dispreferred. This indicates that zero markers may not solely result from phonetic reduction. An alternative path to zero marking more in line with these results would be for them to arise as a distinct, contrastive strategy.

2.3 *Guzmán Naranjo: An analogical approach to the typology of inflectional complexity*

Guzmán Naranjo (2024) addresses the same conjecture as Cotterell *et al.* (2019) with a new predictive mechanism and at much larger scale. Guzmán Naranjo's model is based on explicit local segmentations of string pairs with variables. Local segmentation is both relatively fast and can be run on very small datasets, since each pair of forms produces a single pattern. Thus, while Cotterell et al. require paradigms for at least 700 lexemes to use their neural network method, Guzmán Naranjo is able to analyze on datasets of only 200. Moreover, results from 200-lexeme datasets serve as relatively reliable lower bounds on the values for larger samples, indicating that even small sets of words can yield useful information about a language.

Guzmán Naranjo (2024) concludes that Cotterell *et al.*'s results do not hold across a larger sample of 71 languages. Although there appears to be a trend relating number of paradigm cells to interpredictability, there is no significant correlation. Moreover, he argues that the most valuable measurement of E-complexity is not the number of paradigm cells, but the formal complexity of the rules used to describe them. This sort of E-complexity actually increases as predictability decreases (that is, languages with more complex paradigms are *easier* to predict).

2.4 *Haley et al: Corpus-based measures discriminate inflection and derivation cross-linguistically*

Haley *et al.* (2024) tackle another theoretical question, the division between inflection and derivation. Again, this distinction is the subject of theoretical controversy – Plank (1994) argues that the distinction is gradient rather than categorical, and Haspelmath (2024) claims that it is merely an artifact of traditional linguistic analysis, rather than a phenomenon with real explanatory power. Haley et al. propose to characterise morphological relationships by comparing the difference in orthographic form (edit distance) between the related forms, and the difference in corpus distribution (based on FasTex embeddings (Bojanowski *et al.* 2017)), as well as the variability in these measurements across lexemes. Again, while Plank (1994) is able to apply his measures to only 6 morphological relationships, all in English, Haley *et al.* can scale their analysis further, to a set of 26 languages.

Haley *et al.* find that these measurements can be used to predict the traditional divisions between inflection and derivation with relatively high accuracy (variability being more important than magnitude and distribution more important than form). The measurements can also be used to automatically categorise particular constructions as more or less canonically inflectional by ranking their distance to the decision boundary – comparatives, for example, form an intermediate class.

CHALLENGES AND FUTURE DIRECTIONS 3

Yet, the current generation of databases has not made it trivial to run morphological analyses at scale. One set of issues is evident in a comparison between Guzmán Naranjo's 71 languages and Haley et al.'s 26, most of which come from Europe: the size of available lexical databases is still closely linked to the kind of information desired. While Unimorph collects inflectional paradigms for a large number of languages, derivational relationships are accessible for far fewer, and corpus embeddings (which have to be collected separately) only for a subset of these. More broadly, there is tension between depth of analysis and typological coverage. The more information is needed, the more the analyst must fall back on scarcer resources which tend to push toward a familiar set of well-resourced European languages.

The interface in the other direction (morphophonology) is similarly problematic. Most available databases list orthographic forms gleaned from dictionaries, but these can preserve antiquated relationships, as in modern French (Baroni 2011), or obscure phonologically predictable ones. Grapheme-to-phoneme conversion is a possible solution, as in Becker 2024 and Mortensen *et al.* 2018, but again, requires resources which may not be available across a typologically diverse sample.

A final issue for lexical databases is the quality and systematicity of the data itself. Gorman *et al.* (2019) register a number of complaints about the quality of the scraped Wiktionary data underlying most Unimorph paradigm tables, including mislabeled cells and misparsed orthographic sequences. Other issues of language in use, such as overabundance (Thornton 2019) and dialectal diversity, can also lead to inconsistencies. While modelling studies like Haley *et al.* 2024 are intended to make analytical categories like 'inflection' and 'derivation' more rigorous by providing more objective ways to make the distinction, the authors acknowledge that this is to some extent undercut by the differing ways in which the database represents purported inflections and derivations in the first place. Similarly, Guzmán Naranjo's decision to include all cliticised and periphrastic forms from Unimorph within his analysis raises theoretical questions of what a word is, or whether such a notion is even cross-linguistically applicable (Dixon

Table 1: Supplementary materials

Contribution	Data and code
Inman *et al.* 2024	https://osf.io/n67mq
Becker 2024	https://osf.io/p4mkc/?view_only=5238ace9cb1d4f4d998486ebb28f4fd8
Guzmán Naranjo 2024	https://doi.org/10.5281/zenodo.11147171
Haley *et al.* 2024	https://osf.io/uztgy

et al. 2002). In practice, different Unimorph languages make different decisions on what to include within a lexical entry, and this in turn has implications for the rules produced by alignment systems.

Computational approaches to morphological typology greatly benefit from following the FAIR principles (Wilkinson *et al.* 2016), as well as those of Open Science. As shown in Table 1, each contribution in this volume makes their code and data available through open science platforms, in order to facilitate reuse and reproductibility.

Each of the papers in this volume engages with the linguistic literature by testing or sharpening earlier conjectures with reference to newer and larger datasets. In each case, although the authors' own analysis of their data makes valuable contributions, the work is primarily intended to provide resources (datasets and methods) for future investigation. We hope that the continuing trend of standardization and openness will make large-scale morphological typology more accessible to others within the field, enabling more and more hypotheses to be tested at scale.

REFERENCES

Farrell ACKERMAN and Robert MALOUF (2013), Morphological organization: The low conditional entropy conjecture, *Language*, 89:429–464.

Antonio BARONI (2011), Alphabetic vs. non-alphabetic writing: Linguistic fit and natural tendencies, *The Italian Journal of Linguistics*, 23:127–160, https://api.semanticscholar.org/CorpusID:56412403.

Khuyagbaatar BATSUREN, Omer GOLDMAN, Salam KHALIFA, Nizar HABASH, Witold KIERAŚ, Gábor BELLA, Brian LEONARD, Garrett NICOLAI, Kyle GORMAN, Yustinus Ghanggo ATE, Maria RYSKINA, Sabrina MIELKE, Elena BUDIANSKAYA, Charbel EL-KHAISSI, Tiago PIMENTEL, Michael GASSER, William Abbott LANE, Mohit RAJ, Matt COLER, Jaime Rafael Montoya SAMAME, Delio Siticonatzi CAMAITERI, Esaú Zumaeta ROJAS, Didier LÓPEZ FRANCIS, Arturo ONCEVAY, Juan LÓPEZ BAUTISTA, Gema Celeste Silva VILLEGAS, Lucas Torroba HENNIGEN, Adam EK, David GURIEL, Peter DIRIX, Jean-Philippe BERNARDY, Andrey SCHERBAKOV, Aziyana BAYYR-OOL, Antonios ANASTASOPOULOS, Roberto ZARIQUIEY, Karina SHEIFER, Sofya GANIEVA, Hilaria CRUZ, Ritván KARAHÓĞA, Stella MARKANTONATOU, George PAVLIDIS, Matvey PLUGARYOV, Elena KLYACHKO, Ali SALEHI, Candy ANGULO, Jatayu BAXI, Andrew KRIZHANOVSKY, Natalia KRIZHANOVSKAYA, Elizabeth SALESKY, Clara VANIA, Sardana IVANOVA, Jennifer WHITE, Rowan Hall MAUDSLAY, Josef VALVODA, Ran ZMIGROD, Paula CZARNOWSKA, Irene NIKKARINEN, Aelita SALCHAK, Brijesh BHATT, Christopher STRAUGHN, Zoey LIU, Jonathan North WASHINGTON, Yuval PINTER, Duygu ATAMAN, Marcin WOLIŃSKI, Totok SUHARDIJANTO, Anna YABLONSKAYA, Niklas STOEHR, Hossep DOLATIAN, Zahroh NURIAH, Shyam RATAN, Francis M. TYERS, Edoardo M. PONTI, Grant AITON, Aryaman ARORA, Richard J. HATCHER, Ritesh KUMAR, Jeremiah YOUNG, Daria RODIONOVA, Anastasia YEMELINA, Taras ANDRUSHKO, Igor MARCHENKO, Polina MASHKOVTSEVA, Alexandra SEROVA, Emily PRUD'HOMMEAUX, Maria NEPOMNIASHCHAYA, Fausto GIUNCHIGLIA, Eleanor CHODROFF, Mans HULDEN, Miikka SILFVERBERG, Arya D. MCCARTHY, David YAROWSKY, Ryan COTTERELL, Reut TSARFATY, and Ekaterina VYLOMOVA (2022), UniMorph 4.0: Universal Morphology, in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 840–855, European Language Resources Association, Marseille, France, `https://aclanthology.org/2022.lrec-1.89`.

Laura BECKER (2024), Zero marking in inflection: A token-based approach, in *Computational Approaches to Morphological Typology*, TODO.

Sacha BENIAMINE, Cormac ANDERSON, Mae CARROLL, Matías Guzmán NARANJO, Borja HERCE, Matteo PELLEGRINI, Erich ROUND, Helen SIMS-WILLIAMS, and Tiago TRESOLDI (2023), Paralex: a DeAR standard for rich lexicons of inflected forms, in *Presentation at International Symposium of Morphology*, `https://ismo2023.ovh/fichiers/abstracts/4_ISMO_2023_Paralex.pdf`, https://www.paralex-standard.org.

Sacha BENIAMINE and Matías GUZMÁN NARANJO (2021), Multiple alignments of inflectional paradigms, *Proceedings of the Society for Computation in Linguistics*, 4:216–227.

Balthasar Bickel and Johanna Nichols (2002), Autotypologizing databases and their use in fieldwork, in *Proceedings of the international LREC workshop on resources and tools in field linguistics, Las Palmas*, volume 2627, MPI for Psycholinguistics Nijmegen, doi:10.5167/UZH-76860.

Balthasar Bickel, Johanna Nichols, Taras Zakharko, Alena Witzlack-Makarevich, Kristine Hildebrandt, Michael Riessler, Lennart Bierkandt, Fernando Zúñiga, and John B. Lowe (2022), The autotyp database, doi:10.5281/ZENODO.6793367.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov (2017), Enriching word vectors with subword information, *Transactions of the Association for Computational Linguistics*, 5:135–146.

Olivier Bonami and Sacha Beniamine (2021), Leaving the stem by itself, in Sedigheh Moradi, Marcia Haag, Janie Rees-Miller, and Andrija Petrovic, editors, *All things morphology: Its independence and its interfaces*, pp. 81–98, Benjamins, doi:10.1075/cilt.353.05bon.

Gilles Boyé and Gauvin Schalchi (2016), The status of paradigms, in Andrew Hippisley and Gregory Stump, editors, *The Cambridge handbook of morphology*, pp. 206–234, Cambridge University Press.

Dunstan Brown, Marina Chumakina, and Greville G. Corbett (2012), *Canonical morphology and syntax*, Oxford University Press, ISBN 9780199604326, doi:10.1093/acprof:oso/9780199604326.001.0001, `https://doi.org/10.1093/acprof:oso/9780199604326.001.0001`.

Thea Cameron-Faulkner and Andrew Carstairs-McCarthy (2000), Stem alternants as morphological signata: Evidence from blur avoidance in Polish nouns, *Natural Language and Linguistic Theory*, 18:813–835.

A. D. Carstairs (1987), *Allomorphy in inflexion*, Croom Helm linguistics series, Croom Helm, London.

Greville G. Corbett (2005), *The canonical approach in typology*, pp. 25–49, John Benjamins Publishing Company, doi:10.1075/slcs.72.03cor.

Greville G. Corbett (2023), The typology of external splits, 99(1):108–153, ISSN 1535-0665, doi:10.1353/lan.2023.0007.

Ryan Cotterell, Christo Kirov, Mans Hulden, and Jason Eisner (2019), On the complexity and typology of inflectional morphological systems, *Transactions of the Association for Computational Linguistics*, 7:327–342.

Robert MW Dixon, Alexandra Y Aikhenvald, *et al.* (2002), *Word: A cross-linguistic typology*, Cambridge University Press.

Kyle Gorman, Arya D. McCarthy, Ryan Cotterell, Ekaterina Vylomova, Miikka Silfverberg, and Magdalena Markowska (2019), Weird inflects but OK: Making sense of morphological generation errors, in Mohit Bansal and

Aline VILLAVICENCIO, editors, *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pp. 140–151, Association for Computational Linguistics, Hong Kong, China, doi:10.18653/v1/K19-1014, `https://aclanthology.org/K19-1014`.

Joseph GREENBERG (1966), *Language universals*, number 59 in Janua Linguarum. Series Minor, De Gruyter Mouton, 3rd printing. reprint 2019 edition, ISBN 9783110802528.

Joseph H. GREENBERG (1960), A quantitative approach to the morphological typology of language, *International Journal of American Linguistics*, 26(3):178–184.

Matías GUZMÁN NARANJO (2024), An analogical approach to the typology of inflectional complexity, in *Computational Approaches to Morphological Typology*, TODO.

Matías GUZMÁN NARANJO and Laura BECKER (2021), Coding efficiency in nominal inflection: Expectedness and type frequency effects, *Linguistics Vanguard*, 7(s3):20190075, doi:10.1515/lingvan-2019-0075.

Coleman HALEY, Eduardo M. PONTI, and Sharon GOLDWATER (2024), Corpus-based measures discriminate inflection and derivation cross-linguistically, in *Computational Approaches to Morphological Typology*, TODO.

Martin HASPELMATH (2008), *8 creating economical morphosyntactic patterns in language change*, pp. 185–214, Oxford University PressOxford, ISBN 9780191711442, doi:10.1093/acprof:oso/9780199298495.003.0008.

Martin HASPELMATH (2024), Inflection and derivation as traditional comparative concepts, *Linguistics*, 62(1):43–77.

Martin HASPELMATH, Matthew S. DRYER, David GIL, and Bernard COMRIE (2013), *The world atlas of language structures online*, Max Planck Digital Library, `http://wals.info`.

David INMAN, Alena WITZLACK-MAKAREVICH, Natalia CHOUSOU-POLYDOURI, and Melvin STEIGER (2024), Alignment everywhere all at once: Applying the late aggregation principle to a typological database of argument marking, in *Computational Approaches to Morphological Typology*, TODO.

Christo KIROV, Ryan COTTERELL, John SYLAK-GLASSMAN, Géraldine WALTHER, Ekaterina VYLOMOVA, Patrick XIA, Manaal FARUQUI, Sabrina J. MIELKE, Arya MCCARTHY, Sandra KÜBLER, David YAROWSKY, Jason EISNER, and Mans HULDEN (2018), UniMorph 2.0: Universal Morphology, in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, European Language Resources Association (ELRA), Miyazaki, Japan, `https://aclanthology.org/L18-1293`.

Christo KIROV, John SYLAK-GLASSMAN, Roger QUE, and David YAROWSKY (2016), Very-large scale parsing and normalization of wiktionary morphological paradigms, in Nicoletta Calzolari (Conference CHAIR), Khalid CHOUKRI, Thierry DECLERCK, Sara GOGGI, Marko GROBELNIK, Bente MAEGAARD, Joseph MARIANI, Helene MAZO, Asuncion MORENO, Jan ODIJK, and Stelios PIPERIDIS, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, European Language Resources Association (ELRA), Paris, France, ISBN 978-2-9517408-9-1, `http://ckirov.github.io/UniMorph/`.

Arya D. MCCARTHY, Christo KIROV, Matteo GRELLA, Amrit NIDHI, Patrick XIA, Kyle GORMAN, Ekaterina VYLOMOVA, Sabrina J. MIELKE, Garrett NICOLAI, Miikka SILFVERBERG, Timofey ARKHANGELSKIY, Nataly KRIZHANOVSKY, Andrew KRIZHANOVSKY, Elena KLYACHKO, Alexey SOROKIN, John MANSFIELD, Valts ERNŠTREITS, Yuval PINTER, Cassandra L. JACOBS, Ryan COTTERELL, Mans HULDEN, and David YAROWSKY (2020), UniMorph 3.0: Universal Morphology, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 3922–3931, European Language Resources Association, Marseille, France, ISBN 979-10-95546-34-4, `https://aclanthology.org/2020.lrec-1.483`.

David R. MORTENSEN, Siddharth DALMIA, and Patrick LITTELL (2018), Epitran: Precision G2P for many languages, in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Frans PLANK (1994), Inflection and derivation, in *The Encyclopedia of Language and Linguistics*, pp. 1671–1679, Elsevier Science and Technology.

Hedvig SKIRGÅRD, Hannah J. HAYNIE, Damián E. BLASI, Harald HAMMARSTRÖM, Jeremy COLLINS, Jay J. LATARCHE, Jakob LESAGE, Tobias WEBER, Alena WITZLACK-MAKAREVICH, Sam PASSMORE, *et al.* (2023), Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss, *Science Advances*, 9(16):eadg6175.

Gregory T. STUMP and Raphael FINKEL (2013), *Morphological typology: From word to paradigm*, Cambridge University Press.

Anna M. THORNTON (2019), Overabundance: A canonical typology, *Competition in inflection and word-formation*, pp. 223–258.

A.W. VON SCHLEGEL (1818), *Observations sur la langue et la littérature provençales*, Librairie grecque-latine-allemande.

Mark D. WILKINSON, Michel DUMONTIER, IJsbrand Jan AALBERSBERG, Gabrielle APPLETON, Myles AXTON, Arie BAAK, Niklas BLOMBERG, Jan-Willem BOITEN, Luiz Bonino DA SILVA SANTOS, Philip E. BOURNE, Jildau BOUWMAN, Anthony J. BROOKES, Tim CLARK, Mercè CROSAS, Ingrid DILLO, Olivier DUMON, Scott EDMUNDS, Chris T. EVELO, Richard FINKERS, Alejandra

GONZALEZ-BELTRAN, Alasdair J. G. GRAY, Paul GROTH, Carole GOBLE, Jeffrey S. GRETHE, Jaap HERINGA, Peter A. C. 'T HOEN, Rob HOOFT, Tobias KUHN, Ruben KOK, Joost KOK, Scott J. LUSHER, Maryann E. MARTONE, Albert MONS, Abel L. PACKER, Bengt PERSSON, Philippe ROCCA-SERRA, Marco ROOS, Rene VAN SCHAIK, Susanna-Assunta SANSONE, Erik SCHULTES, Thierry SENGSTAG, Ted SLATER, George STRAWN, Morris A. SWERTZ, Mark THOMPSON, Johan VAN DER LEI, Erik VAN MULLIGEN, Jan VELTEROP, Andra WAAGMEESTER, Peter WITTENBURG, Katherine WOLSTENCROFT, Jun ZHAO, and Barend MONS (2016), The fair guiding principles for scientific data management and stewardship, *Scientific Data*, 3(1):160018, ISSN 2052-4463, doi:10.1038/sdata.2016.18.

Alena WITZLACK-MAKAREVICH, Johanna NICHOLS, Kristine A. HILDEBRANDT, Taras ZAKHARKO, and Balthasar BICKEL (2022), *Managing autotyp data: Design principles and implementation*, pp. 631–642, The MIT Press, ISBN 9780262366076, doi:10.7551/mitpress/12200.003.0061.

Daniel ZEMAN, Joakim NIVRE, Mitchell ABRAMS, Elia ACKERMANN, Noëmi AEPLI, Hamid AGHAEI, Željko AGIĆ, Amir AHMADI, Lars AHRENBERG, Chika Kennedy AJEDE, Gabrielė ALEKSANDRAVIČIŪTĖ, Ika ALFINA, Lene ANTONSEN, Katya APLONOVA, Angelina AQUINO, Carolina ARAGON, Maria Jesus ARANZABE, Bilge Nas ARICAN, Órunn ARNARDÓTTIR, Gashaw ARUTIE, Jessica Naraiswari ARWIDARASTI, Masayuki ASAHARA, Deniz Baran ASLAN, Luma ATEYAH, Furkan ATMACA, Mohammed ATTIA, Aitziber ATUTXA, Liesbeth AUGUSTINUS, Elena BADMAEVA, Keerthana BALASUBRAMANI, Miguel BALLESTEROS, Esha BANERJEE, Sebastian BANK, Verginica BARBU MITITELU, Starkaður BARKARSON, Rodolfo BASILE, Victoria BASMOV, Colin BATCHELOR, John BAUER, Seyyit Talha BEDIR, Kepa BENGOETXEA, Gözde BERK, Yevgeni BERZAK, Irshad Ahmad BHAT, Riyaz Ahmad BHAT, Erica BIAGETTI, Eckhard BICK, Agnė BIELINSKIENĖ, Kristín BJARNADÓTTIR, Rogier BLOKLAND, Victoria BOBICEV, Loïc BOIZOU, Emanuel BORGES VÖLKER, Carl BÖRSTELL, Cristina BOSCO, Gosse BOUMA, Sam BOWMAN, Adriane BOYD, Anouck BRAGGAAR, Kristina BROKAITĖ, Aljoscha BURCHARDT, Marie CANDITO, Bernard CARON, Gauthier CARON, Lauren CASSIDY, Tatiana CAVALCANTI, Gülşen CEBIROĞLU ERYIĞIT, Flavio Massimiliano CECCHINI, Giuseppe G. A. CELANO, Slavomír ČÉPLÖ, Neslihan CESUR, Savas CETIN, Özlem ÇETINOĞLU, Fabricio CHALUB, Shweta CHAUHAN, Ethan CHI, Taishi CHIKA, Yongseok CHO, Jinho CHOI, Jayeol CHUN, Juyeon CHUNG, Alessandra T. CIGNARELLA, Silvie CINKOVÁ, Aurélie COLLOMB, Çağrı ÇÖLTEKIN, Miriam CONNOR, Marine COURTIN, Mihaela CRISTESCU, Philemon DANIEL, Elizabeth DAVIDSON, Marie-Catherine DE MARNEFFE, Valeria DE PAIVA, Mehmet Oguz DERIN, Elvis DE SOUZA, Arantza DIAZ DE ILARRAZA, Carly DICKERSON, Arawinda DINAKARAMANI, Elisa DI NUOVO, Bamba DIONE, Peter DIRIX, Kaja DOBROVOLJC, Timothy DOZAT, Kira DROGANOVA, Puneet DWIVEDI, Hanne ECKHOFF, Sandra EICHE, Marhaba ELI, Ali ELKAHKY,

Binyam EPHREM, Olga ERINA, Tomaž ERJAVEC, Aline ETIENNE, Wograine EVELYN, Sidney FACUNDES, Richárd FARKAS, Jannatul FERDAOUSI, Marília FERNANDA, Hector FERNANDEZ ALCALDE, Jennifer FOSTER, Cláudia FREITAS, Kazunori FUJITA, Katarína GAJDOŠOVÁ, Daniel GALBRAITH, Marcos GARCIA, Moa GÄRDENFORS, Sebastian GARZA, Fabrício Ferraz GERARDI, Kim GERDES, Filip GINTER, Gustavo GODOY, Iakes GOENAGA, Koldo GOJENOLA, Memduh GÖKIRMAK, Yoav GOLDBERG, Xavier GÓMEZ GUINOVART, Berta GONZÁLEZ SAAVEDRA, Bernadeta GRICIŪTĖ, Matias GRIONI, Loïc GROBOL, Normunds GRŪZTIS, Bruno GUILLAUME, Céline GUILLOT-BARBANCE, Tunga GÜNGÖR, Nizar HABASH, Hinrik HAFSTEINSSON, Jan HAJIČ, Jan HAJIČ JR., Mika HÄMÄLÄINEN, Linh HÀ MỸ, Na-Rae HAN, Muhammad Yudistira HANIFMUTI, Sam HARDWICK, Kim HARRIS, Dag HAUG, Johannes HEINECKE, Oliver HELLWIG, Felix HENNIG, Barbora HLADKÁ, Jaroslava HLAVÁČOVÁ, Florinel HOCIUNG, Petter HOHLE, Eva HUBER, Jena HWANG, Takumi IKEDA, Anton Karl INGASON, Radu ION, Elena IRIMIA, Ọlájídé ISHOLA, Kaoru ITO, Siratun JANNAT, Tomáš JELÍNEK, Apoorva JHA, Anders JOHANNSEN, Hildur JÓNSDÓTTIR, Fredrik JØRGENSEN, Markus JUUTINEN, Sarveswaran K., Hüner KAŞIKARA, Andre KAASEN, Nadezhda KABAEVA, Sylvain KAHANE, Hiroshi KANAYAMA, Jenna KANERVA, Neslihan KARA, Boris KATZ, Tolga KAYADELEN, Jessica KENNEY, Václava KETTNEROVÁ, Jesse KIRCHNER, Elena KLEMENTIEVA, Elena KLYACHKO, Arne KÖHN, Abdullatif KÖKSAL, Kamil KOPACEWICZ, Timo KORKIAKANGAS, Mehmet KÖSE, Natalia KOTSYBA, Jolanta KOVALEVSKAITĖ, Simon KREK, Parameswari KRISHNAMURTHY, Sandra KÜBLER, Oğuzhan KUYRUKÇU, Aslı KUZGUN, Sookyoung KWAK, Veronika LAIPPALA, Lucia LAM, Lorenzo LAMBERTINO, Tatiana LANDO, Septina Dian LARASATI, Alexei LAVRENTIEV, John LEE, Phương LÊ HỒNG, Alessandro LENCI, Saran LERTPRADIT, Herman LEUNG, Maria LEVINA, Cheuk Ying LI, Josie LI, Keying LI, Yuan LI, KyungTae LIM, Bruna LIMA PADOVANI, Krister LINDÉN, Nikola LJUBEŠIĆ, Olga LOGINOVA, Stefano LUSITO, Andry LUTHFI, Mikko LUUKKO, Olga LYASHEVSKAYA, Teresa LYNN, Vivien MACKETANZ, Menel MAHAMDI, Jean MAILLARD, Aibek MAKAZHANOV, Michael MANDL, Christopher MANNING, Ruli MANURUNG, Büşra MARŞAN, Cătălina MĂRĂNDUC, David MAREČEK, Katrin MARHEINECKE, Héctor MARTÍNEZ ALONSO, Lorena MARTÍN-RODRÍGUEZ, André MARTINS, Jan MAŠEK, Hiroshi MATSUDA, Yuji MATSUMOTO, Alessandro MAZZEI, Ryan MCDONALD, Sarah MCGUINNESS, Gustavo MENDONÇA, Tatiana MERZHEVICH, Niko MIEKKA, Karina MISCHENKOVA, Margarita MISIRPASHAYEVA, Anna MISSILÄ, Cătălin MITITELU, Maria MITROFAN, Yusuke MIYAO, AmirHossein MOJIRI FOROUSHANI, Judit MOLNÁR, Amirsaeid MOLOODI, Simonetta MONTEMAGNI, Amir MORE, Laura MORENO ROMERO, Giovanni MORETTI, Keiko Sophie MORI, Shinsuke MORI, Tomohiko MORIOKA, Shigeki MORO, Bjartur MORTENSEN, Bohdan MOSKALEVSKYI, Kadri MUISCHNEK, Robert MUNRO, Yugo MURAWAKI, Kaili MÜÜRISEP, Pinkey

NAINWANI, Mariam NAKHLÉ, Juan Ignacio NAVARRO HORÑIACEK, Anna NEDOLUZHKO, Gunta NEŠPORE-BĒRZKALNE, Manuela NEVACI, Lương NGUYỄN THỊ, Huyền NGUYỄN THỊ MINH, Yoshihiro NIKAIDO, Vitaly NIKOLAEV, Rattima NITISAROJ, Alireza NOURIAN, Hanna NURMI, Stina OJALA, Atul Kr. OJHA, Adédayọ OLÚÒKUN, Mai OMURA, Emeka ONWUEGBUZIA, Petya OSENOVA, Robert ÖSTLING, Lilja ØVRELID, Şaziye Betül ÖZATEŞ, Merve ÖZÇELIK, Arzucan ÖZGÜR, Balkız ÖZTÜRK BAŞARAN, Hyunji Hayley PARK, Niko PARTANEN, Elena PASCUAL, Marco PASSAROTTI, Agnieszka PATEJUK, Guilherme PAULINO-PASSOS, Angelika PELJAK-ŁAPIŃSKA, Siyao PENG, Cenel-Augusto PEREZ, Natalia PERKOVA, Guy PERRIER, Slav PETROV, Daria PETROVA, Jason PHELAN, Jussi PIITULAINEN, Tommi A PIRINEN, Emily PITLER, Barbara PLANK, Thierry POIBEAU, Larisa PONOMAREVA, Martin POPEL, Lauma PRETKALNIŅA, Sophie PRÉVOST, Prokopis PROKOPIDIS, Adam PRZEPIÓRKOWSKI, Tiina PUOLAKAINEN, Sampo PYYSALO, Peng QI, Andriela RÄÄBIS, Alexandre RADEMAKER, Mizanur RAHOMAN, Taraka RAMA, Loganathan RAMASAMY, Carlos RAMISCH, Fam RASHEL, Mohammad Sadegh RASOOLI, Vinit RAVISHANKAR, Livy REAL, Petru REBEJA, Siva REDDY, Mathilde REGNAULT, Georg REHM, Ivan RIABOV, Michael RIESSLER, Erika RIMKUTĖ, Larissa RINALDI, Laura RITUMA, Putri RIZQIYAH, Luisa ROCHA, Eiríkur RÖGNVALDSSON, Mykhailo ROMANENKO, Rudolf ROSA, Valentin ROȘCA, Davide ROVATI, Olga RUDINA, Jack RUETER, Kristján RÚNARSSON, Shoval SADDE, Pegah SAFARI, Benoît SAGOT, Aleksi SAHALA, Shadi SALEH, Alessio SALOMONI, Tanja SAMARDŽIĆ, Stephanie SAMSON, Manuela SANGUINETTI, Ezgi SANIYAR, Dage SÄRG, Baiba SAULTE, Yanin SAWANAKUNANON, Shefali SAXENA, Kevin SCANNELL, Salvatore SCARLATA, Nathan SCHNEIDER, Sebastian SCHUSTER, Lane SCHWARTZ, Djamé SEDDAH, Wolfgang SEEKER, Mojgan SERAJI, Syeda SHAHZADI, Mo SHEN, Atsuko SHIMADA, Hiroyuki SHIRASU, Yana SHISHKINA, Muh SHOHIBUSSIRRI, Dmitry SICHINAVA, Janine SIEWERT, Einar Freyr SIGURÐSSON, Aline SILVEIRA, Natalia SILVEIRA, Maria SIMI, Radu SIMIONESCU, Katalin SIMKÓ, Mária ŠIMKOVÁ, Kiril SIMOV, Maria SKACHEDUBOVA, Aaron SMITH, Isabela SOARES-BASTOS, Shafi SOUROV, Carolyn SPADINE, Rachele SPRUGNOLI, Steinór STEINGRÍMSSON, Antonio STELLA, Milan STRAKA, Emmett STRICKLAND, Jana STRNADOVÁ, Alane SUHR, Yogi Lesmana SULESTIO, Umut SULUBACAK, Shingo SUZUKI, Zsolt SZÁNTÓ, Chihiro TAGUCHI, Dima TAJI, Yuta TAKAHASHI, Fabio TAMBURINI, Mary Ann C. TAN, Takaaki TANAKA, Dipta TANAYA, Samson TELLA, Isabelle TELLIER, Marinella TESTORI, Guillaume THOMAS, Liisi TORGA, Marsida TOSKA, Trond TROSTERUD, Anna TRUKHINA, Reut TSARFATY, Utku TÜRK, Francis TYERS, Sumire UEMATSU, Roman UNTILOV, Zdeňka UREŠOVÁ, Larraitz URIA, Hans USZKOREIT, Andrius UTKA, Sowmya VAJJALA, Rob VAN DER GOOT, Martine VANHOVE, Daniel VAN NIEKERK, Gertjan VAN NOORD, Viktor VARGA, Eric VILLEMONTE DE LA CLERGERIE, Veronika VINCZE, Natalia VLASOVA, Aya WAKASA, Joel C.

WALLENBERG, Lars WALLIN, Abigail WALSH, Jing Xian WANG, Jonathan North WASHINGTON, Maximilan WENDT, Paul WIDMER, Sri Hartati WIJONO, Seyi WILLIAMS, Mats WIRÉN, Christian WITTERN, Tsegay WOLDEMARIAM, Tak-sum WONG, Alina WRÓBLEWSKA, Mary YAKO, Kayo YAMASHITA, Naoki YAMAZAKI, Chunxiao YAN, Koichi YASUOKA, Marat M. YAVRUMYAN, Arife Betül YENICE, Olcay Taner YILDIZ, Zhuoran YU, Arlisa YULIAWATI, Zdeněk ŽABOKRTSKÝ, Shorouq ZAHRA, Amir ZELDES, He ZHOU, Hanzhi ZHU, Anna ZHURAVLEVA, and Rayan ZIANE (2023), Universal Dependencies 2.13, http://hdl.handle.net/11234/1-5287.

George Kingsley ZIPF (2013), *The psycho-biology of language*, Routledge, ISBN 9781136310461, doi:10.4324/9781315009421.

Micha Elsner
0000-0002-1432-2129
elsner.14@osu.edu

Ohio State University

Sacha Beniamine
0000-0003-2584-3576
s.beniamine@surrey.ac.uk

University of Surrey

Micha Elsner and Sacha Beniamine (2024), *Computational approaches to morphological typology*, Journal of Language Modelling, 12(2):271–286
https://dx.doi.org/10.15398/jlm.v12i2.431