

DEmA: the Pavia Diachronic Emergence of Alignment database

Sonia Cristofaro (Università degli Studi di Pavia) & Guglielmo Inglese (KU Leuven, FWO)

Digital atlases such as WALS (Dryer & Haspelmath 2013) have established themselves as important tools in typological research, by making available a significant amount of structured typological data. WALS can be queried to retrieve information regarding the distribution of typological variables across (a sample of) languages as well as possible correlations between variables. In WALS, there are three chapters dedicated to alignment patterns (Comrie 2013a; Comrie 2013b; Siewierska 2013). These offer a strictly synchronic overview of the alignment patterns attested in the world's languages (e.g. nominative, ergative, and active alignment, hierarchical alignment), various types of alignment splits, and their possible morphosyntactic realization and geographical distribution.

As is well-known, typological features arise through specific historical processes. Unfortunately, typological resources that may supply information on how specific typological traits develop over time are not numerous. An example is DiACL (Carling 2017), which features an inventory of typological features akin to that of WALS, but crucially differs from the latter in that the data is structured with the explicit purpose of facilitating the investigation of language change. In a way, DiACL enables users to track down typological shifts in the history of individual languages. In the case of alignment, this means that in addition to information on the distribution of individual patterns, one can also learn that a shift from nominative to (split-)ergative alignment has taken place in the history of Indo-Aryan, by comparing e.g. Sanskrit (nominative) to Pashto (ergative). However, data on typological shifts is only indirectly inferable from DiACL, and historical details on the direction and dynamics of such shifts are lacking altogether.

Progress in grammaticalization studies and historical linguistics have brought to light an increasing body of evidence regarding the possible origins of different alignment patterns. In particular, individual patterns have been shown to

originate from various types of pre-existing constructions. Such historical information is sparse across publications, not easily comparable nor accessible to non-specialists, and it is not linked to general resources such as WALS.

The *Pavia Diachronic Emergence of Alignment* (DEmA) project aims to build a comprehensive open access database on the emergence of alignment patterns cross-linguistically. Historical research has mostly focused on (i) shift from one overall alignment system to the other (e.g. nominative > split-ergative in Indo-Aryan) and (ii) etymological sources of case marking (e.g. demonstrative > ergative). Less attention has been paid to the features of source constructions or to the type of mechanisms involved. The goal of DEmA is to supply data on this less researched point. The data is systematized in such a way that one can distinguish the role of the different components at play, including features and distribution of source constructions, specific developmental mechanisms, features and distribution of the resulting construction, and the relationship of the resulting construction with the global system of the language.

Datapoints in DEmA are individual languages. For each language, the following information is provided:

- Original alignment pattern and resulting alignment pattern (e.g. neuter > accusative);
- Source construction and resulting construction (e.g. passive with oblique marked agent > active with overtly marked A argument and zero marked P argument)
- (A)symmetry in the encoding of the resulting alignment pattern;
- Source element of alignment markers (e.g. demonstrative) and grammatical domain of the resulting construction (e.g. case marking, verb indexing)
- Constraints on the distribution of the resulting alignment pattern (e.g. tense-based split ergativity)

- Mechanisms of language change (e.g. grammaticalization, reanalysis of argument structure)
- Type of data (e.g. historical data, internal reconstruction)

Metadata on the identification, genealogical classification and location of languages (e.g. Glottocodes) is also provided, so as to allow linkage to other typological databases.

DEmA is still under construction. Entries are manually filled by means of a web interface built for the purpose. The data is structured through SQL. Our main source of data is the existing literature on the history of alignment patterns. At present, 21 languages from 15 different language families have been inserted. Note that available data on the emergence of alignment patterns is much more limited than that on the synchronic distribution of these patterns. This means that DEmA will necessarily have a much smaller language coverage as compared to e.g. WALS.

We plan to release the first online and access free version of DEmA by January 2021. Once completed, the database will be fully searchable, allowing users to query the database for all parameters and combinations thereof. The database is also expandable, and we encourage scholars working on the diachrony of alignment to also share their data through DEmA.

From a theoretical point of view, DEmA will offer a solid empirical basis to tackle a number of questions regarding traditional assumptions about the motivations underlying alignment patterns (Harris & Campbell 1995; Gildea 1998; Mithun 2005; Creissels 2008; Cristofaro 2012, 2013, 2014, among others). What are the possible source constructions of different alignment patterns? Which mechanisms of language change may lead to the emergence of individual patterns? To what extent are particular patterns motivated by the properties of specific source constructions, rather than general semantic or pragmatic principles pertaining to those patterns in themselves? To what extent does the distribution of individual patterns across different contexts (alignment splits) follow from such principles, and to what extent is it related to the properties of the source construction? Can specific alignment patterns be related to a single underlying principle, or is each pattern an epiphenomenal result of several distinct diachronic processes, each motivated by a different principle?

At a more general methodological level, the architecture of DEmA is also unique in that it offers a theoretically well-grounded and explicit systematization of several parameters pertaining to language change (e.g. source constructions, developmental process, type of data), so that these can be effectively implemented into a searchable format. In this respect, we hope that DEmA will also provide a suitable model for future typological resources dealing with the diachrony of different grammatical domains.

References

- Carling, Gerd (ed.) 2017. *Diachronic Atlas of Comparative Linguistics Online*. Lund: Lund University. (<https://diacl.ht.lu.se/>)
- Comrie, Bernard. 2013. Alignment of Case Marking of Full Noun Phrases. In Dryer & Haspelmath (eds.) (<http://wals.info/chapter/98>)
- Comrie, Bernard. 2013. Alignment of Case Marking of Pronouns. In Dryer & Haspelmath (eds.) (<http://wals.info/chapter/99>)
- Creissels, Denis. 2008. Direct and indirect explanations of typological regularities: the case of alignment variations. *Folia Linguistica* 42, 1–38.
- Cristofaro, Sonia. 2012. Cognitive explanations, distributional evidence, and diachrony. *Studies in Language* 36, 645–70.
- Cristofaro, Sonia. 2013. The referential hierarchy: Reviewing the evidence in diachronic perspective. In Dik Bakker & Martin Haspelmath (eds), *Languages across Boundaries: Studies in the Memory of Anna Siewierska*, 69–93. Berlin: de Gruyter.
- Cristofaro, Sonia. 2014. Competing motivations and diachrony: what evidence for what motivations? In Brian MacWhinney, Andrej Malchukov & Edith Moravcsik (eds.), *Competing motivations in grammar and usage*, 282–98. Oxford: OUP.
- Dryer, Matthew S. & Haspelmath, Martin (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute (<http://wals.info/>)
- Gildea, Spike. 1998. *On reconstructing grammar: Comparative Cariban morphosyntax*. Oxford: OUP.
- Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: CUP.
- Mithun, Marianne. 2005. Ergativity and language contact on the Oregon Coast: Alsea, Siuslaw, and Coosan. *Proceedings of the Berkeley Linguistic Society*, 77–95.
- Siewierska, Anna. 2013. Alignment of Verbal Person Marking. In Dryer & Haspelmath (eds.). (<http://wals.info/chapter/100>)