Data Transparency and Citation in Gesture

Lauren Gawne,1 Chelsea Krajcik,2 Helene N. Andreassen,3 Andrea L. Berez-Kroeker,4 Barbara F. Kelly4
1 La Trobe University, 2 SQAS, University of London, 3 UiT The Arctic University of Norway, 4 University of Hawaii at Manoa, 5 University of Melbourne

Background
Gesture Studies has a strong history of research that spans multiple fields, but there is still not a robust culture of valuing reproducibility.

Reproducibility provides benefits including:
- Accountability in research by facilitating access to the underlying data and methods ensuring that other researchers may also reach the same conclusions (Gezelter 2014).
- Raised professional valuation of developing corpora that can be reused (Haspelmath & Michaelis 2014; Margetts et al. 2016; Berez-Kroeker et al. 2018).

Skubisz’s (2017) survey of data coding and terminological definitions in GESTURE demonstrated that these key features of research are often underspecified in articles published in the journal to date.

Data Transparency in linguistics
- Language Description 100 grammars (2003-2012) (Gawne et al. 2017a). Vast majority did not provide citations to underlying data.
- Linguistic Typology 50 articles from 5 years of Linguistic Typology (Gawne et al. 2017b). Low frequency of authors citing own data.
- 10 leading linguistics journals 270 articles (2003-2012) (Berez-Kroeker et al. 2017). Different subfields have different strengths in methods descriptions and data citation.

Survey of Data Citation in GESTURE
- 5 years of research articles in GESTURE (2012-2017 vol. 12.1-16.1)
- Total of 56 articles
- Discussion and introductory articles omitted
- The type of data, and what languages are the target of the analysis
- Total of 56 articles
- Based on methods from previous surveys (above)

Data Citation Conventions
Data citation directs the reader back to the specific source of the data. Sources could be datasets (publicly accessible or private), published texts (e.g. Bible translations), or other academic publications. (n=53)

**Most Common Data Types**
Perhaps unsurprising, given the diversity of work in the field, there is diversity in the types of data surveyed. (n=56)

**Discussion**
This survey demonstrates that we need a more robust culture of data accountability in gesture research. Researchers are mostly drawing on their own data, but are not stating the location of their data, and are not providing citation of individual examples.

GESTURE has recently adopted the standards of the Center for Open Science, which requires thorough description of methods and analyses, plus presentations of data in online data repositories.

As a research community we need to foster a culture of valuing research reproducibility.