John Benjamins Publishing Company

Data transparency and citation in the journal *Gesture*

Lauren Gawne,[1] Chelsea Krajcik,[2] Helene N. Andreassen,[3] Andrea L. Berez-Kroeker,[4] and Barbara F. Kelly[5]
[1] La Trobe University | [2] SOAS University of London | [3] UiT The Arctic University of Norway | [4] University of Hawai'i at Mānoa | [5] University of Melbourne

Data is central to scholarly research, but the nature and location of data used is often under-reported in research publications. Greater transparency and citation of data have positive effects for the culture of research. This article presents the results of a survey of data citation in six years of articles published in the journal *Gesture* (12.1–17.2). Gesture researchers draw on a broad range of data types, but the source and location of data are often not disclosed in publications. There is also still a strong research focus on only a small range of the world's languages and their linguistic diversity. Published papers rarely cite back to the primary data, unless it is already published. We discuss both the implications of these findings and the ways that scholars in the field of gesture studies can build a positive culture around open data.

Keywords: gesture studies, data citation, open data, data management

Introduction

Gesture studies is a field founded on an empirical research method; our understanding of gesture is based on evidence from data which is analysed and disseminated in research publications. Data is central to the formulation of analysis, but it is rarely presented in a way that is transparent to the reader. The transparency of data can refer to a number of features. These festures include how well the data is described in a research article, and whether the data is accessible in its entirety, or as a subset of specific examples, or has access restrictions. Transparency also includes citing the data to varying levels of granularity, directing the reader to a whole corpus, or to specific examples within that collection. There are many advantages to

having greater transparency of data in research practice – for authors, readers, and the field as a whole. These include heightened professional valuation of data collection and sharing (Haspelmath & Michaelis, 2014; Thieberger, Margetts, Morey, & Musgrave, 2016) and greater accountability in research by facilitating access to the underlying data and methods (Gezelter, 2009).

In order to best understand where the field of gesture studies is heading with regards to the use of data, we seek to understand the current state of practice. To do this, we conducted a six year survey of research publications in the journal *Gesture*, from 12.1 (in 2012) to 17.2 (in 2018). This survey examines how researchers describe the source and location of their data, and whether they cite examples back to the primary source. We also look at the types of data and the languages that researchers in gesture studies are working with, to better understand the support that will be needed to continue to develop a culture of research data transparency.

While researchers in this field draw on a broad range of data types, the nature of this data is rarely made clear in publications. This has implications for the future progress of research. We discuss the results of our survey in light of the broader 'open access' movement, as well as the specific ethical implications of working with gestural, and particularly video, data. We also discuss the results in light of the move by *Gesture* to require greater transparency in data reporting.

Background

In the field of gesture studies, perhaps more than any other field in human communication, the means by which data is collected and analysed becomes crucial to the development and interrogation of theories underpinning the frameworks of data analysis. Gesture research draws on a range of different methodologies for analysing multimodality, particularly manual gestures and gaze. In early studies gestures were characterised as relatively static visual signs rather than dynamic signs changing across space and time (e.g., de Jorio, 1832; Morris, Collett, Marsh, & O'Shaughnessay, 1979). Thanks to affordable video capture and computers for analysis, recent research tends towards more empirical studies presenting transcribed, coded, and analysed gestures and affiliated spoken language. These empirical methods and analytic approaches yield ideal data sets for the replicability and reproducibility of findings. Gesture studies has a strong history of qualitative and quantitative research that spans multiple research fields. One thing that links all research in this area is a clear acknowledgement of the role of primary data in shaping our understanding of the form of gesture and its role in communication. The discipline-spanning nature of gesture studies means that as a field we need to consider the multiple ways in which data transparency can lead to subsequent research.

Replicability and *reproducibility* have each received a good deal of attention in the social sciences lately, especially from those interested in the Open Access and Open Science initiatives (Buckheit & Donoho, 1995; de Leeuw, 2001; Donoho, 2010; Gawne & Styles, forthcoming, inter alia). While these terms may seem interchangeable, the differences between them are crucial to the future of the language sciences. *Replicability* is probably the more widely familiar of the two concepts and is one that has underpinned the scientific process for a long time. Replicable studies are those studies that are created, executed, and subsequently described in such a way that another researcher could recreate the study down to the smallest detail. The results of this replicated study would either confirm the previous results – lending them credence – or disconfirm them. The aim of replicability is to ensure some level of scientific rigor in the research process, as well as to provide a mechanism by which results can be "checked" by those with a healthy degree of skepticism. Granted, it may not be enjoyable having ones research disconfirmed, but that is part of doing good science, and it says something positive about our methods that they were replicable in the first place.

Replicability is the standard for scientific studies in which variables can be carefully controlled, such as in laboratory experiments. However, a great deal of science deals with data that is a little more "wild" (cf. the 2011 special issue of *Science* on reproducibility edited by Jasny, Chin, Chong, & Vignieri). This includes the behavioral data that is the basis of language-based research in many disciplines (Berez-Kroeker, Gawne, et al., 2018). It is nearly impossible to create language studies that are truly replicable in the original sense of the word because it is very difficult to control for every factor that leads to the use of a particular word or gesture in a given linguistic context (be it naturalistic, elicited, or experimental). Even in the most tightly controlled language experiments, it would be impossible to control for a subject's previous experience with a particular sound, word, phrase, or gesture. In such situations, the notion of *reproducibility* becomes valuable: reproducible research, therefore, is research that facilitates access to not just the methods used in the study, but also to the data collected in the study, and the tools (software, scripts, etc.) used to collect and analyse it. Another researcher could then examine or even reanalyse the data to reach similar or different conclusions. Thus, when replicability is impossible, reproducibility steps in to ensure a level of rigor and accountability in the scientific process.

Research that is reproducible or replicable requires a high degree of transparency on the part of scientists who must effectively communicate to their audiences about every aspect of their methodology, from collection to processing to analysis. Doing so would allow someone else to recreate the original study to test if the original hypothesis and analysis is supported. Replicability further requires clear description of the location of the underlying data set and how one would gain access.

The Open Data movement began gaining momentum around the same time that the field of gesture studies was formalised. The earliest initiatives in open access publishing in the 1990s coalesced in the Budapest Open Access Initiative,[1] which advocated for open access journals, in 2002, the same year the International Society for Gesture Studies was founded.[2] Researchers recognised the profound effect that the internet had in making it easier than ever for knowledge to be shared openly with a wide audience. In 2003 the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities took the Budapest statement a step further, focusing on the dissemination of all research knowledge, including primary data:

> Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society. New possibilities of knowledge dissemination not only through the classical form but also and increasingly through the open access paradigm via the Internet have to be supported. We define open access as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community.[3]

A culture of valuing data transparency in gesture studies is also beginning to coalesce. The flagship journal *Gesture* has recently adopted the standards of the Center for Open Science,[4] which requires thorough description of methods and analyses, plus lodgement of data in publicly accessible online data repositories. *Gesture* was founded in 2001. While there are other journals that publish research on gesture, as well as book series and monographs, the journal demonstrates how gesture studies has grown and diversified over the last two decades. In 2007 (volume 7), the journal began publishing three issues a year rather than two, and *Gesture* continues to include articles on new topics in the field.

The growth of gesture studies allows us to take stock of where we have come from and where we are going with regard to research methodology. Skubisz (2017) undertook a survey of data coding and terminology definitions in quantitative papers in *Gesture* from the foundation of the journal until 2016. In this survey, she demonstrated that key features of research design and methods are often underspecified. Our survey complements Skubisz (2017) as we look at how researchers manage data rather than methodology. We also include both quantitative and qualitative research. We focus on six recent years of publication as Skubisz did not notice any trends in research practice changing over the history of the journal.

---

1. http://budapestopenaccessinitiative.org/read visited Nov 8 2018.

2. http://gesturestudies.com/history.php visited Nov 8 2018.

3. http://openaccess.mpg.de/Berlin-Declaration visited Nov 8 2018.

4. http://benjamins.com/#catalog/journals/gest/guidelines visited Nov 8 2018.

Researchers in gesture studies are not alone in reconsidering the role of data in research. The related fields of social psychology and linguistics are also experiencing a raised awareness of the need to move towards more transparent research methods. In social psychology, a series of separate events regarding non-replicability of findings occurred across a number of high profile publications in major journals that called into question many long-standing research practices in the field (Ioannidis, 2012). While many of these practices are based around particular approaches to statistical methods and the way research questions are framed in the collection of data and their presentation in final publications, the overarching theme of this 'crisis of confidence' in social psychology has been that these events were enabled by a culture that did not value open science and replicability (Chambers, 2017; Nelson, Simmons, & Simonsohn, 2018). This lead to the founding of the Center for Open Science in 2012,[5] and the publication of the Open Science Collaboration (2015) which replicated a hundred key papers in social psychology, finding very low rates of result replication.

In linguistics, there have been a number of surveys conducted that look at the transparency and research methods in different subfields. This has included a survey of 270 articles across ten leading linguistics journals published between 2003 and 2012 (Berez-Kroeker, Gawne, Kelly, & Heston, 2017). This survey found that different subfields have different strengths in methods descriptions; for example articles in the journal Studies in Second Language Acquisition consistently provide some description of methodology, while articles in Journal of Sociolinguistics consistently give some metadata on research participants. In a parallel survey, Gawne, Kelly, Berez-Kroeker, & Heston (2017) examined one hundred descriptive grammars, finding that there was a great deal of variation in the methodological detail provided and that the vast majority did not provide citations to underlying data. These surveys fed into a position statement which called for stronger valuation of open data and reproducibility in linguistic research (Berez-Kroeker, Gawne, et al., 2018), and the distillation of this position statement into the Austin Principles of Data Citation in Linguistics (Berez-Kroeker, Andreassen, et al., 2018).

We wish to contribute to this positive shift in research practice by interrogating where we come from as a field with regard to data and how we can move forward. In this paper we present a survey of all research articles published in *Gesture* from 2012 to 2018. For each article we seek to understand how transparent each published article is in regard to the presence of both clear research methods and the citation of data to a source that would allow the reader to analyse the data for themselves.

---

5. http://cos.io/ visited Nov 8 2018.

Survey of data citation in *Gesture*

To gain an understanding of the state of data citation in the field of gesture studies, we conducted a survey of almost six years of research articles in the journal *GESTURE*. We took articles from volumes 12.1 to 17.2 (2012–2018). We focused specifically on research articles, omitting commentaries, book reviews or introductions to special issues that do not include extensive review and discussion.

There were 81 articles in total. Our survey is based on methods from previous surveys (Gawne et al., 2017; Berez-Kroeker et al., 2017). We collected information on the type of data in each article to understand the nature of how researchers in gesture studies approach research. This included the source of data, location of data, the type of data, and what languages the data is sourced from. We then looked at how transparent each article is in regard to citation of data to a source that would allow the readers to analyse the data for themselves.

In this section we discuss each of the features that we coded for and what categories we coded. While the discussion is mostly presented in aggregate, the survey data is presented as supplementary material in a spreadsheet hosted online with this publication. Examples of coded categories are given in the results section, where relevant.

Source of data

Researchers can draw upon data they collect themselves, or data from others. We coded for the source of the data used in each article, and allowed for multiple sources. Sources include:

> OWN: Own collected data
> PUBD: Published, either as a corpus or in an existing manuscript
> UNK: Source is unknown from this article

We also had categories UNPUBD for explicitly-stated unpublished data and OTHER to allow for other possibilities, but no articles were coded for either of these categories.

Location of data

We coded for where the data are currently located, if stated by the author. Options included:

ARCH: archived, location described even briefly
ONL: data online somewhere other than article, but not clear if it is archived
PUBD: in another publication (the author's or someone else's)
HERE: all data are included with the article (e.g., as appendix)
HERESUMMARY: all data summarised in the article (e.g., from an experiment)
UNK: unknown from this article

We distinguished between archives, i.e., physical or digital repositories with an institutional commitment to long-term preservation, and presentation online, where the long term stability of the data is not made clear. For something to explicitly count in either of these categories, the author needs to indicate the archive or online location in the paper. We also had the coding category HERE, for when the article contains the data, and is its own main source. This is a common way of presenting data in fields with experiment-based methods (see, for example, the journal *Second Language Learning and Teaching* in Berez-Kroeker et al., 2017), but no articles were categorised as this in our survey.

Data types

We coded for the types of data found in each article. Since it is not uncommon to draw upon multiple data types in the study of gesture, we included a category MULTI, rather than coding each article across multiple categories. Options included:

CONVO: conversation
NARR: narrative
TASK: task
EXPER: experiment
MULTI: a range of data types, or multiple genres (e.g., speeches, conversation and song)
REVIEW: review of existing literature
OTHER: other

For MULTI we made note of the various types of data. For OTHER we made note of what data was collected, and this is described in the results.

Languages included

Although the languages included in this study are not strictly a matter of data cita-
tion, there are a number of reasons to consider the languages that are targets of
research in gesture studies. The first is that the management of citation and trans-
parency in minority languages has particular challenges that may not be faced by
languages with larger populations where anonymity may be more easily provided.
The second is that a field dominated by larger languages may not be providing
the breadth of data to be able to approach anything like a typologically-driven
approach to gesture, nor providing the range of data necessary to make claims
about the extent to which differences in use are motivated by language, culture
and/or cognition.

    We collected information about what languages were included as part of the
analysis of each article. We did not start with a pre-determined list, but made a
note of the languages referred to in each article.

Data citation conventions

We coded for citation conventions used in examples. Citation conventions include:

> NONE: no citation
> STD: use of APA referencing to other publication
> CODEEX: a code that is explained in the text or in a footnote
> CODEUNX: a code that is not explained
> NUMBER: examples are numbered in the order they appear in the original
> recordings or are discussed
> URL: a URL link to the data
> NAME: name of performance, story, or speaker

Illustrative examples of the citation conventions used are given in the results sec-
tion.

Results

Source of data

Researchers draw on both their own data, and existing data, but still mostly col-
lect their own data (Table 1). Multiple sources were counted for eight papers leav-
ing 73 papers with a single source of data. Raw totals and broad percentages are
given for each category, here and in all further tables in the results section.

Table 1. Source of data

Code	Detail	Total	%
OWN	author's own data	51	63%
PUBD	Published	17	20%
UNK	unknown source	6	7%
OWN & PUBD	Uses both author's own & published data	8	10%

By far, the most common source of data in the journal is that collected by the researchers themselves. Whether this be conducting an experiment, or making a series of recordings of naturalistic conversation for analysis, most researchers collect their own data. There are good reasons that this is the case. For example, experimental methods require the formulation of a hypothesis and then conducting a well-crafted experiment to test the hypothesis. There is also still a paucity of publically available corpora that are of interest or use to gesture researchers, especially beyond a small set of languages. The reliance on researchers' own data is not a problem in-and-of-itself, however as we discuss in subsequent sections of these results it makes the need to be transparent about the location of the data, and data citation, all the more pressing.

There are two different types of published data. The first is the use of published data where some form of original data is available, and the researchers perform their own analysis of the data. This can include corpora that are available for research. For example, Kimura & Kazik (2017) use the Corpus of English for Academic and Professional Purposes (CEAPP) (2014) in their study of how speakers of English as a second language use gesture to assist in the learning of grammar. Other researchers use existing data from other sources, such as Lempert's (2017) use of publicly televised political debate, or Looney & Meier's (2014) analysis of pointing gestures drawing on all publically available footage of Genie, a child who had been raised with minimal linguistic input.

The second type of published data is when researchers draw upon the existing literature of gestural analysis. Of course, all academic research publications do this while setting up the motivation for their study, but some of the papers in the six year sample exclusively synthesised the published literature in a particular area to advance a theoretical position. For example, Corballis (2012) draws on a range of literature on primates and humans to argue that language evolved from manual gestures, while Bavelas & Healing (2013) undertook a review of the literature on mutual visibility and its effect on gesturing.

Eight papers draw upon multiple sources for primary data. For all eight this was a combination of the author's own data and published data. Agwuele (2014) draws upon both commercially released films as well as the author's own fieldwork

recordings in an analysis of the repertoire of Yoruba hand and face gestures. Cibulka (2013) draws on both the author's own recordings as well as the TalkBank corpus (MacWhinney, 2007) to look at the use of writing gestures in Japanese conversation. The use of multiple sources allows the authors to draw upon a broader range of data than they otherwise would have had access to.

The six papers with 'unknown' sources involve video recordings, which are most likely the authors' own, but there is no clear explanation in the methods as to how the recordings were made or obtained, so this cannot be confirmed. Regardless of whether researchers work with their own data, or publically available data, it is important to have transparent research methods that make the source of data clear, even in those cases where the authors cannot share the data itself.

Location of data

Stating data location increases opportunity for reproducibility and replicability, because others can return to the original data on which an analysis is built. The vast majority of articles represent the only known location of the data, or a summary of the data (Table 2). There are multiple locations noted for data in three papers.

Table 2. Location of data

Code	Detail	Total	%
UNK	Unknown	37	46%
HERESUMMARY	A summary of the data is given in the paper	26	32%
PUBD	In another publication (the author's or someone else's)	10	12%
ONL	website or other non-archive internet storage	3	4%
ARCH	Archived	2	2%
multiple	ARCH & UNK (1), PUBD & UNK (1), HERESUMMARY & ONL (1)	3	4%

As we discussed above, there are a number of articles where the data source was the existing research (PUBD). The reader is able to go back to the original research publication to see the original analysis (however limitations in terms of clarity of source and location of data may still hold).

There are many papers in which a summary of the data is given in the publication, such that the readers can get an overview of the major features of the data, or a synthesis of it, but cannot themselves access the original recordings, or the original coding of the data to confirm the statistical analysis. These papers were

exclusively constrained tasks, experiments and analysis of particular features of conversations. While data transparency can be facilitated by better access to some of the underlying data, at least a summary of the data allows for basic review and inclusion in meta-analysis.

Four articles indicated that the data were available online. One example is the use of the NCCU Corpus of Spoken Mandarin (Chui & Lai, 2008) by Chui (2012), who provided a link to the data in the article. Another example is Sutton-Spence & Napoli (2013), who drew upon online recordings of performed poetry in British Sign Language and American Sign Language. While these links to data are currently useful, unless the data are housed in an archive with a mandate for long term storage, it is possible that access to the data for the interested reader may not be maintained. We discuss the need to consider that data is both accessible and stable in the discussion section.

There are three articles we categorised as having the data archived. Cibulka (2013) uses TalkBank (MacWhinney, 2007), Gawne (2018) uses Syuba data archived open access with PARADISEC,[6] and finally Kamunen (2018) uses the Oulu Video Corpus of English and Finnish and the Oulu Corpus of US British Television Interviews, available to the local research community and others on request. Other articles referenced corpora, or collections of materials, but did not make it clear to the reader where these materials were archived, or if they were available and which parts of the corpora were analysed. We therefore labeled them as location UNK 'unknown'.

Table 3 gives a list of all of the published sources of data used in papers in Gesture in the period we surveyed. This table may help researchers find data to use in their own work, or provide a model for making their own data available. There are likely many other corpora used in publications in this survey that are open access, or at least available on request, but without making this clear in publications it is difficult to make use of them.

The category with the largest number of papers is that where the location of data is not made clear to the reader at all. This is not to say that the data may not be housed somewhere secure, nor that it is inaccessible to the reader, but that the authors have not made this clear. Close reading of a small number of papers suggests that the data are located in corpora or archives that may be researcher-accessible, however there is a lack of clear citation of the data.

---

6. http://catalog.paradisec.org.au/collections/SUY1 shorter clips with the specific examples are bundled together in a FigShare collection to accompany the article: https://figshare.com/articles /_/6462284

Table 3. Published data and the papers in which this data is used

Data source	Publication(s)	Language	Data type
NCCU Corpus of Spoken Mandarin (Chui, 2012)	Chui & Lai (2008)	Mandarin	Spontaneous face-to-face recordings of Mandarin, Hakka, and Southern Min
TalkBank (MacWhinney, 2007)	Cibulka (2013)	Various	An open access corpus of more than 34 languages
YouTube	Mihas (2018); Sutton-Spence & Napoli (2013)	British Sign Language; American Sign Language; Northern Kampa Arawaks	Video recordings of performances (does not constitute a long-term archive)
Kagate (Nepal) (Gawne, 2009)	Gawne (2018)	Syuba	Video recordings
Oulu Video Corpus of English and Finnish	Kamunen (2018)	English, Finish	Video recordings of naturally occurring everyday conversations
Oulu Corpus of US British Television Interviews	Kamunen (2018)	English	Video recordings broadcast English language interviews from 2001 to 2015

Data types

Perhaps unsurprising, given the diversity of work in the field, there is diversity in the types of data surveyed (Table 4).

Table 4. Data types

Code	Detail	Total	%
EXPER	experimental data	20	25%
CONVO	conversation data	15	19%
TASK	task-based data	14	17%
MULTI	multiple data types	13	16%
REVIEW	review of existing literature	9	11%
NARR	narrative	6	7%
OTHER	other data types	4	5%

There is a lot of research in *Gesture* that uses experimental data, task-based data or conversational data. We took a very broad approach to each of these genres. For example Wehling (2018) uses televised interviews, which we include 'conversation'

as the focus of the paper on the use of gestures to manage discourse in conversation. There are also review articles that synthesise existing data across genres.

Some authors draw upon multiple data types in their work. Cooperrider & Núñez (2012), Sandler (2012), and Mihas (2013) all draw upon ethnographic or anthropological methods that involve the collection of data across a range of genres in their analysis of particular gestural phenomena (nose-pointing, gesture grammaticalisation in signed language, and gesture-ideophone use respectively).

In the category of 'other' data types, Matoesian & Gilbert (2016) examined the use of gesture by attorneys in closing arguments of a case, Sutton-Spence & Napoli (2013) looked at signed poetry performances, Kettner & Carpendale (2013) examined parents journals of their children's acquisition of shaking and nodding gestures, and Lefebvre (2016) studied recordings of Aikido training sessions.

The variety in the data types used in gesture studies is one of the strengths of the field, but it means that a move toward stronger practices of data citation must take into account a range of approaches.

Languages included

There were 37 languages included as targets of research in 73 of the 81 papers. English is the dominant target of study, with a rapidly falling long tail of other languages where there are 4 or fewer articles. A list of all the languages, and the papers in which they feature, is given in the Appendix.

There were some papers which we coded as 'general' or 'no' language, these were predominantly review articles, or articles that focused on primate behaviour (Cissewski & Boesch, 2016). There was also Lefebvre's (2016) Aikido training session, where speech was not analysed and the language of the participants was not stated.

Table 5 is a summary of the languages include in articles. Percentages total to greater than 100 as some articles drew upon multiple languages. For languages with only 1 use, we group them by their modality (spoken or signed). Of the 37 languages, 21 are spoken languages and 16 are signed languages.

The 'Other – Spoken' languages are: Anyi, Arabic, Ashéninka Perené, Hebrew, Japanese, Malay, Maori, Northern Kampa Arawaks, Norwegian, Siwu, Spanish, Swedish, Syuba, Yoruba, Yupno.

The 'Other – Signed' languages are: Anmatyerr Sign Language, Armenian alternate sign language, Auslan, British Sign Language, Cape York Peninsula alternate sign language, Kuuk Thaayorre Sign Language, Nepali Sign Language, New Zealand Sign Language, Ngaanyatjarra/Ngaatjatjarra Sign Language, Norwegian Sign Language, Protactile American Sign Language, Yolŋu Sign Language.

Table 5. Languages in analysis (for papers see the Appendix)

Language	Number of articles	% of total
English	38	47%
American Sign Language	4	5%
Al-Sayyid Bedouin Sign Language (ABSL)	3	4%
German	3	4%
Israeli Sign Language (ISL)	3	4%
Mandarin	3	4%
Dutch	2	2%
French	2	2%
Homesign	2	2%
Italian	2	2%
Other – Signed	13	
Other – Spoken	15	
None/Unknown/General	8	10%

As a final observation on data transparency in academic publications with regards to the languages that feature in research papers, when the language was mentioned in only one or two papers, we noticed the author was more likely to make the target languages clear in the title of the paper, or at least the abstract and keywords. When the language of analysis was English, this was much less likely to be the case.

Different types of research have particular skews in language. Of the 20 experimental papers 13 were exclusively on English, and two were on English and other languages (e.g., English and French in Tutton, 2012, English, with 2 other spoken languages and 3 signed languages in Padden et al., 2013). Research that draws on multiple data types is a more heterogeneous set, with no language included in this category twice, and only one focused on English (Alibali et al., 2013), in a study of classroom interactions that were analysed and also repackaged for an experimental design.

Data citation conventions

Data citation directs the reader back to the specific source of the data (Table 6). Many papers included no citation, and very few cited data in a way that could lead the reader back to the underlying data.

Table 6. Data citation conventions

Code	Detail	Total	%
NONE	no citation convention	35	43%
NAME	name of speaker or text	15	18%
NUM	numbered in order of original recordings or discussion	15	18%
STD	Standard citation to published source	9	11%
UNEX	a citation code that is unexplained	2	3%
EXPL	an explained citation code that links back to materials	2	3%
URL	a weblink to the location of the data online	0	0%
multiple	NONE & STD (1), URL & NAME (1), EXPL & STD (1)	3	4%

Given the high number of publications in which the location of the data is unstated, it is perhaps unsurprising that we find a paucity of data citation. The relationship between citing the data back to a source, and having a source be accessible to the reader in some form, is the reason that data citation as an end-stage practice needs to be considered in the larger context of replication and reproducibility.

Other than having no citation back to data, the most common way to cite data was to give a name of the speaker of the text that was being discussed in a particular example. In Harrison (2014) each numbered example is given a name, which refers to the particular topic of that interaction, example 2 is titled 'not to be a politician'. The example consists of a string of dialogue with speaker turns marked by single initials (B. & J. in this example). This is coupled with a cropped screenshot of an ELAN tier, correlated with stills of the performance of the gesture, further annotated with arrows to show direction of movement. The initials and the images of the participants makes it clear if examples come from the same speakers, but it's not always clear if they're from the same narrative. In Hauser (2014), the recordings of Japanese student conversations are analysed with speakers referred to by pseudonym. For example, excerpt 5 is from a conversation between Yoshida and Nishi. This is a useful piece of metadata, as it narrows down which speaker or which conversation is being analysed, but it does not necessarily make it easier for the reader to find this particular interaction in the original data.

While almost all examples are numbered sequentially throughout an article, in some papers this is the only form of citation used.

We found three examples of data cited with a code that resolved back to the original data, and which was clearly explained by the author. Cibulka explains that examples taken from the TalkBank corpus are cited using a code, and explains the code. The example with the citation [Talkbank/CABank/Sakura04 17:52] is 17 minutes and 52 seconds into the Sakura04 recording from the TalkBank corpus,

while [Bq/1 54:00] is 54 minutes into the researcher's own recording 'Bq'. The Talk-Bank recordings are resolvable back to the original corpus for the interested reader, but it is not clear where the researcher's own recordings are archived, if they are at all. Gawne (2018) and Kamunen (2018) also used codes that resolved back to the specific point in the specific recording that is under discussion.

There are also a small number of papers that have citation codes that are not explained to the reader. Tutton (2012) gives examples citations such as [EngDesc8] and [FrenDesc11]. The reader can figure out that these refer to numbered recordings of descriptive tasks in English and French respectively, but the use of any citation code should be clearly explained in the publication itself.

Alongside citing the name of the poem and the poet in their analysis of signed poetry, Sutton-Spence & Napoli (2013:10) give a URL to one of the poems which is hosted publically on YouTube. This kind of direct linking can be convenient, particularly for readers accessing the journal digitally, but like all data citation requires that the data remains hosted stably at that URL.

We have counted eleven papers that use 'standard' citation of existing publications. These include review articles that drew exclusively on published data. The other use of 'standard' data citation was to cite back to a specific corpus, for example Kok, Bergmann, Cienki, & Kopp (2016) cited the Bielefeld Speech and Gesture Alignment corpus (Lücking, Bergman, Hahn, Kopp, & Rieser, 2013), which was the basis of their materials, Chui (2012) cited Chui & Lai, (2008) for the NCCU Corpus of Spoken Mandarin, and Cibulka (2013) cited (MacWhinney, 2007) for the TalkBank corpus. In each of these three cases, the reference was to a 'proxy publication' that provided the reader with information about the corpus, rather than the corpus itself.

It should also be noted that every single paper we looked at used appropriate standard citation to existing literature when referring to published data or claims that were not the authors' own. That all authors follow a citation practice when it is codified in a style sheet, and in a set of social expectations makes us optimistic that with the right support, data citation can also become a common practice.

Discussion

Gesture researchers are drawing on a wide variety of data types, and the research area includes data from a wide range of languages across both spoken and signed modalities. However, 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. In this discussion we begin by looking at some of

the challenges that scholars in gesture studies face with regard to the presentation of data, and how these can be navigated with a mindset that centres open access.

One of the most immediate concerns that many researchers in this field have is that their work includes the collection of video data, which is not easily de-identifiable or sharable (Green, Woods, & Foley, 2011). Current technological infrastructure facilitates access to primary data, and the linking of research publications to this data. However, this infrastructure has potential negative consequences in that sensitive information can be easily spread. Thus, in order to protect their research participants, researchers need to be aware of the risks and current regulations, and how to carry out research data management that is both ethically and legally sound. Researchers are already aware of their ethical and legal obligations within the institution and country they work in, as well as in regards to the communities they work with. There is growing concern regarding how the European Union's General Data Protection Regulation (GDPR)[7] will affect research data management, both within the EU but also abroad, as many universities and funders often set their agenda to the most conservative possible set of regulations.

The discussion about whether it is appropriate to publish the primary data demonstrates the difference between transparency and openness: If the data cannot be published openly due to ethical or legal reasons, citation of the data to a closed-access repository and link to its metadata, or the publication of de-identified secondary or aggregated data, at least make these restrictions transparent. At the very least, discussion of why it is ethically or legally inappropriate to share this data makes the research method and process transparent, even in cases where there are good grounds not to make it open.

Different repositories allow researchers to select different levels of access to data that may suit particular projects. Some repositories allow some elements of the research data to be made open, and for others the data remains closed or only accessible upon invitation. Some repositories allow data to be embargoed for a required period if necessary. 'Openness' with regard to research data is not a binary of fully open vs. fully closed, but a series of choices researchers needs to consider. While it can be tempting to always default to the most closed option, there are good reasons to build open access into a project early, and find a repository that best supports your data needs. For a list of repository evaluation criteria, see Whyte (2015).

Once the data management plan has been established (Jones, 2011; Kung, forthcoming), data has been collected and stored in an appropriate repository, there is still a need to link the data to subsequent research publications. While acad-

---

7. http://eugdpr.org/ and http://gdpr-info.eu/ visited Nov 8 2018.

emics routinely cite existing publications, we do not have the same training, history of practice or style guide resources for citing our own primary data. A research publication should include citation back to the body of data as a whole, be it a corpus or a small experimental data set, and where relevant also cite the individual examples to their location within the data set. There are a growing number of resources for this kind of citation practice (cf. Ball & Duke, 2015), and archives now routinely provide automatically formatted citations. The increased use of persistent identifiers such as digital object identifiers (DOIs) are providing useful infrastructure for this kind of data citation.

Although we have focused on the logistics of data management and citation so far, building this into research practice has many benefits. As scholars it can help us think critically about our motivation for collecting data before commencing a project (John, Loewenstein, & Prelec, 2012), and help minimise the loss of data through securing it in a repository (Vines et al., 2014). Citability of data also puts it on more equal footing with publications as a product of research, helping to make an argument for the development of research data (particularly that which can be reused) as an important output that should be recognised in grant, job and promotion applications (Haspelmath & Michaelis, 2014; Thieberger et al., 2016). This avoids the need to use proxy publications, where the author cites a publication about the data collection, as the data itself is acknowledged as a valid research output. As readers, transparent data citation allows us to more easily replicate or reproduce research, or use the data to ask different research questions all together. Funders, publishers and research institutions are also beginning to see the benefit of transparent and open research, particularly with regards to higher rates of dissemination, value for money through reuse, and the minimisation of questionable research practices (Harris, 2017).

Researchers in gesture studies are already beginning to move towards including more open research practices in their work. As mentioned above, *Gesture* has recently adopted submission guidelines that require researchers to clearly describe methods and materials and link to videos that any still figures in the article are taken from (where it is ethical to do so). As part of this move towards open data, *Gesture* is participating in the badge program of the Center for Open Science.[8] Researchers can now add badges to their publication if it has 'open data', 'open materials' or 'preregistered' methods. However, as long as these guidelines are optional rather than mandatory adoption will occur piecemeal. We recommend that *Gesture* and other journals adopt a timeframe in which the requirement to make a transparent statement about data becomes obligatory.

---

8. GESTURE publication guidelines: http://benjamins.com/catalog/gest/guidelines; COS badges: http://cos.io/our-services/open-science-badges/ visited Nov 8 2018.

Further to this, now is the ideal time for the gesture community to develop clearer guidelines around the citation of data in publications to the original source. It is relatively easy to direct the reader to a particular online repository, and many repositories now provide formatted citations, but there is still no good set of guidelines for how to resolve specific examples of gesture use to a particular place in a particular video in a corpus. We also need to formalise expectation that this should be done for all examples in a research publication. Moving forward on such a project is the current work of the Linguistics Data Interest Group of the Research Data Alliance.[9]

Beyond these specific concrete actions, we also need to build a positive and supportive culture to encourage our colleagues to build openness and transparency into their research practice. Transparency is a fundamental guiding principle in research. If transparency builds trust between peers, whether they are readers, peer reviewers or collaborators, it also helps researchers achieve confidence in their relationships with subjects, research institutions and funding bodies.

Conclusion

Gesture studies is a field that draws upon varied methods and varied data to better understand a broad range of multimodal phenomena. To ensure that the field is in the best possible position to build on the existing decades of research, we need to start thinking more critically as a discipline about the role that data plays in our research methods and publications. Greater transparency with regard to the description of data used in publications, and a more open approach to data sharing and citation, can have many positive benefits, for individual scholars and for the field as a whole.

Funding

9.　www.rd-alliance.org/groups/linguistics-data-ig visited Nov 8 2018.

Acknowledgements

References

Agwuele, Augustine (2014). A repertoire of Yoruba hand and face gestures. *Gesture*, *14* (1), 70–96. https://doi.org/10.1075/gest.14.1.04agw

Alibali, Martha W., Andrew G. Young, Noelle M. Crooks, Amelia Yeo, Matthew S. Wolfgram, Iasmine M. Ledesma, Mitchell J. Nathan, Ruth Breckinridge Church, & Eric J. Knuth (2013). Students learn more when their teacher has learned to gesture effectively. *Gesture*, *13* (2), 210–233. https://doi.org/10.1075/gest.13.2.05ali

Andreassen, Helene N., Andrea L. Berez-Kroeker, Lauren Collister, Philipp Conzett, Christopher Cox, Koenraad De Smedt, Bradley McDonnell, and the Research Data Alliance Linguistic Data Interest Group (2019). Tromsø recommendations for citation of research data in linguistics (Version 1). *Research Data Alliance*. https://doi.org/10.15497/RDA00040

Andrén, Mats (2014). Multimodal constructions in children: Is the headshake part of language? *Gesture*, *14* (2), 141–170. https://doi.org/10.1075/gest.14.2.02and

Ball, Alex & Monica Duke (2015). How to cite datasets and link to publications. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk /resources/how-guides

Bavelas, Janet & Sara Healing (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, *13* (1), 63–92. https://doi.org/10.1075/gest.13.1.03bav

Benazzo, Sandra & Aliyah Morgenstern (2014). A bilingual child's multimodal path into negation. *Gesture*, *14* (2), 171–202. https://doi.org/10.1075/gest.14.2.03ben

Berez-Kroeker, Andrea L., Lauren Gawne, Barbara F. Kelly, & Tyler Heston (2017). *A survey of current reproducibility practices in linguistics journals, 2003–2012.* https://sites.google.com /a/hawaii.edu/data-citation/survey

Berez-Kroeker, Andrea L., Helene N. Andreassen, Lauren Gawne, Gary Holton, Susan Smythe Kung, Peter Pulsifer, & Lauren B. Collister, The Data Citation and Attribution in Linguistics Group, & the Linguistics Data Interest Group (2018). *The Austin Principles of Data Citation in Linguistics.* Version 1.0. http://site.uit.no /linguisticsdatacitation/austinprinciples/ (accessed Nov 23, 2018).

Berez-Kroeker, Andrea L., Lauren Gawne, Susan Smythe Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David I. Beaver, Shobhana Chelliah, Stanley Dubinsky, Richard P. Meier, Nicholas Thieberger, Karen Rice, & Anthony C. Woodbury (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, *56* (1), 1–18. https://doi.org/10.1515/ling-2017-0032

Buckheit, Jonathan B. & David Donoho (1995). WaveLab and reproducible research. In A. Antoniadis & G. Oppenheim (Eds.), *Wavelets and statistics* (pp. 55–81). New York: Springer. https://doi.org/10.1007/978-1-4612-2544-7_5

Chambers, Chris (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton: Princeton University Press.

Chui, Kawai (2012). Cross-linguistic comparison of representations of motion in language and gesture. *Gesture*, *12* (1), 40–61. https://doi.org/10.1075/gest.12.1.03chu

Chui, Kawai (2018). Spatial conceptualization of sequence time in language and gesture. *Gesture*, *17* (1), 176–195. https://doi.org/10.1075/gest.00015.chu

Chui, Kawai & Huei-ling Lai (2008). The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics*, *6* (2), 119–144.

Cibulka, Paul (2013). The writing hand: Some interactional workings of writing gestures in Japanese conversation. *Gesture*, *13* (2), 166–192. https://doi.org/10.1075/gest.13.2.03cib

Cissewski, Julia & Christophe Boesch (2016). Communication without language: How great apes may cover crucial advantages of language without creating a system of symbolic communication. *Gesture*, *15* (2), 224–249. https://doi.org/10.1075/gest.15.2.04cis

Cooperrider, Kensy & Rafael Núñez (2012). Nose-pointing: Notes on a facial gesture of Papua New Guinea. *Gesture*, *12* (2), 103–129. https://doi.org/10.1075/gest.12.2.01coo

Corballis, Michael C. (2012). How language evolved from manual gestures. *Gesture*, *12* (2), 200–226. https://doi.org/10.1075/gest.12.2.04cor

Corina, David P. & Eva Gutierrez (2016). Embodiment and American Sign Language. *Gesture*, *15* (3), 291–305. https://doi.org/10.1075/gest.15.3.01cor

Corpus of English for academic and professional purposes. (2014). *Corpus of videos and accompanying transcripts from educational contexts*. Unpublished raw data. http://crellt.la.psu.edu/ceapp-1

de Jorio, Andrea (1832). *La mimica degli antichi investigata nel gestire napoletano*. Naples: Stamperia e cartiera del Fibreno.

de Leeuw, Jan (2001). Reproducible research: The bottom line. *UCLA Department of Statistics papers*. http://escholarship.org/uc/item/9050x4r4 (accessed October 9, 2018).

de Nooijer, Jacqueline A., Tamara van Gog, Fred Paas, & Rolf A. Zwaan (2014). Words in action: using gestures to improve verb learning in primary school children. *Gesture*, *14* (1), 46–69. https://doi.org/10.1075/gest.14.1.03noo

Dingemanse, Mark (2013). Ideophones and gesture in everyday speech. *Gesture*, *13* (2), 143–165. https://doi.org/10.1075/gest.13.2.02din

Donoho, David L. (2010). An invitation to reproducible computational research. *Biostatistics*, *11*, 385–388. https://doi.org/10.1093/biostatistics/kxq028

Edwards, Terra (2018). Sign-creation in the Seattle DeafBlind community: A triumphant story about the regeneration of obviousness. *Gesture*, *16* (2), 305–328. https://doi.org/10.1075/gest.16.2.06edw

Fasolo, Mirco & Laura D'Odorico (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture*, *12* (1), 1–15. https://doi.org/10.1075/gest.12.1.01fas

Ferrara, Lindsay & Rolf Piene Halvorsen (2018). Depicting and describing meanings with iconic signs in Norwegian Sign Language. *Gesture*, *16* (3), 371–395. https://doi.org/10.1075/gest.00001.fer

Fleming, Luke (2014). Negating speech: Medium and modality in the development of alternate sign languages. *Gesture*, *14* (3), 263–296. https://doi.org/10.1075/gest.14.3.01fle

Fuks, Orit (2016). Intensifier actions in Israeli Sign Language (ISL) discourse. *Gesture*, *15* (2), 192–223. https://doi.org/10.1075/gest.15.2.03fuk

Gawne, Lauren (collector) (2009). *Kagate (Nepal) (SUY1)*, Digital collection managed by PARADISEC. [Open Access]. https://doi.org/10.4225/72/56E976A071650

Gawne, Lauren (2018). Contexts of use of a rotated palms gesture among Syuba (Kagate) speakers in Nepal. *Gesture*, *17* (1), 37–64. https://doi.org/10.1075/gest.00010.gaw

Gawne, Lauren, Barbara F. Kelly, Andrea L. Berez-Kroeker, & Tyler Heston (2017). Putting practice into words: The state of data and methods transparency in grammatical descriptions. *Language Documentation & Conservation*, *11*, 157–189. 10125/24731

Gawne, Lauren & Suzy J. Styles. (still in press). Situating linguistics in the social science data movement. In Andrea L. Berez-Kroeker, Bradley McDonnell, & Eve Koller (Eds.), *The open handbook of linguistic data management*. Cambridge, M.A.: MIT Press Open.

Gezelter, Dan (2009). Being scientific: Falsifiability, verifiability, empirical tests, and reproducibility. *The OpenScience project*. Online: http://www.openscience.org/blog/?p:312 (accessed Nov 23, 2018).

Green, E. Mara (2018). Performing gesture: The pragmatic functions of pantomimic and lexical repertoires in a natural sign narrative. *Gesture*, *16* (2), 329–363. https://doi.org/10.1075/gest.16.2.07gre

Green, Jennifer, Anastasia Bauer, Alice Gaby, & Elizabeth Marrkilyi Ellis (2018). Pointing to the body: Kin signs in Australian Indigenous sign languages. *Gesture*, *17* (1), 1–36. https://doi.org/10.1075/gest.00009.gre

Green, Jennifer, Gail Woods & Ben Foley (2011). Looking at language: Appropriate design for sign language resources in remote Australian Indigenous communities. In Nicholas Thieberger, Linda Barwick, Rosey Billington, & Jill Vaughan (Eds.), *Sustainable data from digital research: Humanities perspectives on digital scholarship* (pp. 66–89). Melbourne: University of Melbourne.

Gruber, James, Jeanette King, Jen Hay, & Lucy Johnston (2016). The hands, head, and brow: A sociolinguistic study of Māori gesture. *Gesture*, *15* (1), 1–36. https://doi.org/10.1075/gest.15.1.01gru

Harris, Richard (2017). *Rigor mortis: how sloppy science creates worthless cures, crushes hope, and wastes billions*. New York: Basic Books.

Harrison, Simon (2014). The organisation of kinesic ensembles associated with negation. *Gesture*, *14* (2), 117–140. https://doi.org/10.1075/gest.14.2.01har

Haspelmath, Martin & Susanne Maria Michaelis (2014). Annotated corpora of small languages as refereed publications: A vision. *Diversity Linguistics Comment*. Online: http://dlc.hypotheses.org/691 (accessed Nov 23, 2018).

Hauser, Eric (2014). Solution strokes: Gestural component of speaking trouble solution. *Gesture*, *14* (3), 297–319. https://doi.org/10.1075/gest.14.3.02hau

Haviland, John B. (2013). The emerging grammar of nouns in a first generation sign language: specification, iconicity, and syntax. *Gesture*, *13* (3), 309–353. https://doi.org/10.1075/gest.13.3.04hav

Hunsicker, Dea & Susan Goldin-Meadow (2013). How handshape type can distinguish between nouns and verbs in homesign. *Gesture*, *13* (3), 354–376. https://doi.org/10.1075/gest.13.3.05hun

Ioannidis, John P.A. (2012). Why science isn't necessarily self-correcting. *Perspectives on Psychological Science*, *7* (6): 645–654. https://doi.org/10.1177/1745691612464056

Jasny, Barbara R., Gilbert Chin, Lisa Chong, & Sacha Vignieri (2011). Introduction to special issue: Again, and again, and again. *Science*, *334* (6060), 1225. https://doi.org/10.1126/science.334.6060.1225

John, Leslie K., George Loewenstein, & Drazen Prelec (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23* (5), 524–532. https://doi.org/10.1177/0956797611430953

Johnston, Trevor (2013). Towards a comparative semiotics of pointing actions in signed and spoken languages. *Gesture*, *13* (2), 109–142. https://doi.org/10.1075/gest.13.2.01joh

Jones, Sarah (2011). How to develop and data management and sharing plan. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk /resources/how-guides

Kamunen, Antti (2018). Open Hand Prone as a resource in multimodal claims to interruption. *Gesture*, *17* (2), 291–321. https://doi.org/10.1075/gest.17002.kam

Kettner, Viktoria A. & Jeremy I.M. Carpendale (2013). Developing gestures for no and yes: Head shaking and nodding in infancy. *Gesture*, *13* (2), 193–209. https://doi.org/10.1075/gest.13.2.04ket

Kimura, Daisuke & Natalia Kazik (2017). Learning in-progress. *Gesture*, *16* (1), 127–151. https://doi.org/10.1075/gest.16.1.05kim

Kok, Kasper, Kirsten Bergmann, Alan Cienki, & Stefan Kopp (2016). Mapping out the multifunctionality of speakers' gestures. *Gesture*, *15* (1), 37–59. https://doi.org/10.1075/gest.15.1.02kok

Kung, Susan Smythe. (still in press). Developing a data management plan. In Andrea L. Berez-Kroeker, Bradley McDonnell, & Eve Koller, *The open handbook of linguistic data management*. Cambridge, M.A.: MIT Press Open.

Lefebvre, Augustin (2016). The coordination of moves in Aikido interaction. *Gesture*, *15* (2), 123–155. https://doi.org/10.1075/gest.15.2.01lef

Lempert, Michael (2017). Uncommon resemblance. *Gesture*, *16* (1), 35–67. https://doi.org/10.1075/gest.16.1.02lem

Li, Heng (2018). Time on hands: Deliberate and spontaneous temporal gestures by speakers of Mandarin. *Gesture*, *16* (3), 396–415. https://doi.org/10.1075/gest.00002.li

Looney, Veronica & Richard P. Meier (2014). Genie's middle-finger points and signs: A case study. *Gesture*, *14* (1), 97–107. https://doi.org/10.1075/gest.14.1.05loo

Lücking, Andy, Kirsten Bergman, Florian Hahn, Stefan Kopp, & Hannes Rieser (2013). Data-based analysis of speech and gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications. *Journal on Multimodal User Interfaces*, *7* (1/2), 5–18. https://doi.org/10.1007/s12193-012-0106-8

MacWhinney, Brian (2007). The TalkBank project. In Joan C. Beal, Karen P. Corrigan, & Hermann L. Moisl (Eds.), *Creating and digitizing language corpora: Synchronic databases* (Vol. *1*, pp. 163–180). Houndmills: Palgrave-Macmillan. https://doi.org/10.1057/9780230223936_7

Matoesian, Gregory & Kristin Gilbert (2016). Multifunctionality of hand gestures and material conduct during closing argument. *Gesture*, *15* (1), 79–114. https://doi.org/10.1075/gest.15.1.04mat

Mechraoui, Amal & Faridah Noor Binti Mohd Noor (2017). The direction giving pointing gestures of the Malay Malaysian speech community. *Gesture*, *16* (1), 68–99. https://doi.org/10.1075/gest.16.1.03mec

Mihas, Elena (2013). Composite ideophone-gesture utterances in the Ashéninka Perené 'community of practice', an Amazonian Arawak society from Central-Eastern Peru. *Gesture*, *13* (1), 28–62. https://doi.org/10.1075/gest.13.1.02mih

Mihas, Elena (2018). Interactional functions of lip funneling gestures: A case study of Northern Kampa Arawaks of Peru. *Gesture*, *16* (3), 432–479. https://doi.org/10.1075/gest.00004.mih

Mittelberg, Irene (2018). Embodied frames and scenes: Body-based metonymy and pragmatic inferencing in gesture. *Gesture*, *16* (2), 203–244. https://doi.org/10.1075/gest.16.2.03mit

Morris, Desmond, Peter Collett, Peter Marsh, & Marie O'Shaughnessay (1979). *Gestures: Their origins and distribution*. London: Cape.

Müller, Cornelia (2018). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, *16* (2), 277–304. https://doi.org/10.1075/gest.16.2.05mul

Murillo, Eva & Mercedes Belinchón (2012). Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture*, *12* (1), 16–39. https://doi.org/10.1075/gest.12.1.02mur

Nelson, Leif D., Joseph Simmons, & Uri Simonsohn (2018). Psychology's renaissance. *Annual Review of Psychology*, *69*, 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nyst, Victoria (2016). The depiction of size and shape in gestures accompanying object descriptions in Anyi (Côte d'Ivoire) and in Dutch (The Netherlands). *Gesture*, *15* (2), 156–191. https://doi.org/10.1075/gest.15.2.02nys

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science*, *349* (6251), aac4716. https://doi.org/10.1126/science.aac4716

Padden, Carol A., Irit Meir, So-One Hwang, Ryan Lepic, Sharon Seegers, Tory Sampson (2013). Patterned iconicity in sign language lexicons. *Gesture*, *13* (3), 287–308. https://doi.org/10.1075/gest.13.3.03pad

Sandler, Wendy (2012). Dedicated gestures and the emergence of sign language. *Gesture*, *12* (3), 265–307. https://doi.org/10.1075/gest.12.3.01san

Sikveland, Rein O. & Richard A. Ogden (2012). Holding gestures across turns: Moments to generate shared understanding. *Gesture*, *12* (2), 166–199. https://doi.org/10.1075/gest.12.2.03sik

Skubisz, Joanna (2017). A systematic review of the methods reported in the journal GESTURE. *iGesto*. Porto: February 2–3.

Sutton-Spence, Rachel & Donna Jo Napoli (2013). How much can classifiers be analogous to their referents? *Gesture*, *13* (1), 1–27. https://doi.org/10.1075/gest.13.1.01sut

Tkachman, Oskana & Wendy Sandler (2013). The noun-verb distinction in two young sign languages. *Gesture*, *13* (3), 253–286. https://doi.org/10.1075/gest.13.3.02tka

Thieberger, Nicholas, Anna Margetts, Stephen Morey, & Simon Musgrave (2016). Assessing annotated corpora as research output. *Australian Journal of Linguistics*, *36* (1), 1–21. https://doi.org/10.1080/07268602.2016.1109428

Tutton, Mark (2012). When and why the lexical Ground is a gestural Figure. *Gesture*, *12* (3), 361–386. https://doi.org/10.1075/gest.12.3.04tut

Vines, Timothy H., Arianne Y. K. Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, & Diana J. Rennison (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24* (1), 94–97. https://doi.org/10.1016/j.cub.2013.11.014

Wehling, Elisabeth (2018). Discourse management gestures. *Gesture*, *16* (2), 245–276. https://doi.org/10.1075/gest.16.2.04weh

Whyte, Angus (2015). *Where to keep research data: DCC checklist for evaluating data repositories*. V.1.1. Edinburgh: Digital Curation Centre. Available online: www.dcc.ac.uk /resources/how-guides

Appendix

Below is a list of languages other than English included in the journal, and the references to the papers in which they appear.

Language	Reference(s)
Al-Sayyid Bedouin Sign Language	Sandler, 2012; Tkachman & Sandler, 2013; Padden et al., 2013
American Sign Language	Sutton-Spence & Napoli, 2013; Padden et al., 2013; Looney & Meier, 2014; Corina & Gutierrez, 2016
Anmatyerr Sign Language	Green et al., 2018
Anyi	Nyst, 2016
Arabic	Padden et al., 2013
Armenian alternate sign language	Fleming, 2014
Ashéninka Perené	Mihas, 2013
Auslan	Johnston, 2013
British Sign Language	Sutton-Spence & Napoli, 2013
Cape York Peninsula alternate sign language	Fleming, 2014
Dutch	de Nooijer et al., 2014; Nyst, 2016
French	Tutton, 2012; Benazzo & Morgenstern, 2014
German	Kok et al., 2016; Mittelberg, 2018; Müller, 2018
Hebrew	Padden et al., 2013
Homesign	Haviland, 2013; Hunsicker & Goldin-Meadow, 2013
Israeli Sign Language	Sandler, 2012; Tkachman & Sandler, 2013; Fuks, 2016
Italian	Fasolo & D'Odorico, 2012; Benazzo & Morgenstern, 2014

Language	Reference(s)
Japanese	Cibulka, 2013
Kuuk Thaayorre Sign Language	Green et al., 2018
Malay	Mechraoui & Noor, 2017
Mandarin	Chui, 2012; Li, 2018; Chui, 2018
Maori	Gruber, King, Hay, & Johnston, 2016
Nepali Sign Language	Green, 2018
New Zealand Sign Language	Padden et al., 2013
Ngaanyatjarra/Ngaatjatjarra Sign Language	Green et al., 2018
Northern Kampa Arawaks	Mihas, 2018
Norwegian	Sikveland & Ogden, 2012
Norwegian Sign Language	Ferrara & Halvorsen, 2018
Protactile American Sign Language	Edwards, 2018
Siwu	Dingemanse, 2013
Spanish	Murillo & Belinchón, 2012
Swedish	Andrén, 2014
Syuba	Gawne, 2018
Yolŋu Sign Language	Green et al., 2018
Yoruba	Agwuele, 2014
Yupno	Cooperrider & Núñez, 2012

Online appendixes

Online appendixes can be found here: https://doi.org/10.1075/gest.00034.gaw.additional

Address for correspondence

Lauren Gawne
La Trobe University
Victoria, 3086
Australia
l.gawne@latrobe.edu.au

Biographical notes

Lauren Gawne is a Lecturer in Linguistics at La Trobe University, Australia. Lauren's research primarily focuses on use of grammar and gesture in Tibeto-Burman languages of Nepal.

Chelsea Krajcik is a PhD candidate at SOAS, University of London, England. She works on small-scale multilingualism in the Lower Casamance region of Senegal. Her particular research interests includes co-speech gestures, semantics, and the effects of multilingualism multimodally.

Helene N. Andreassen has a PhD in French linguistics from the UiT The Arctic University of Norway. She holds the position as subject librarian for linguistics at the same institution, specialising in research data management and research ethics. Helene's research primarily focuses on French phonology, in particular L1 and L2 acquisition, and dialectal variation.

Andrea L. Berez-Kroeker is an associate professor in the Department of Linguistics at the University of Hawai'i at Mānoa. She is a documentary linguist specialising in endangered language preservation.

Barbara F. Kelly is a linguist at the University of Melbourne. She has investigated carer-child multimodal communication across communities in the Himalayas, remote Indigenous Australia, and urban industrial settings.