Natural action processing conversation analysis and big interactional data

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

Association for Computing Machinery (ACM)

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

LICENCE

REPOSITORY RECORD

Natural Action Processing

Conversation Analysis and Big Interactional Data

William Housley

School of Social Sciences, Cardiff University, Cardiff, UK, housleyw@cardiff.ac.uk

Saul Albert

School of Social Sciences, Loughborough University, Loughborough, UK, s.b.albert@lboro.ac.uk

Elizabeth Stokoe

School of Social Sciences, Loughborough University, Loughborough, UK, e.h.stokoe@lboro.ac.uk

ABSTRACT

This position paper identifies a crucial opportunity for the reciprocal exchange of methods, data and phenomena between conversation analysis (CA), ethnomethodology (EM) and computer science (CS). Conventional CS classification of sentiment, tone of voice, or personality do not address what people *do* with language or the paired sequences that organize actions into social interaction. We argue that CA and EM can innovate and substantially enhance the scope of the dominant CS approaches to big interactional data if artificial intelligence-based natural *language* processing systems are trained using CA annotated data to do what we call natural *action* processing.

CCS CONCEPTS

• Computing methodologies~Natural language processing • Computing methodologies~Discourse, dialogue and pragmatics • Applied computing~Sociology

KEYWORDS

Conversation Analysis, Ethnomethodology, Social Interaction, Artificial Agents

1 Background

Developments in computational methods have been applied to the data and phenomena of discourse analysis with promising results. Computer-assisted transcription systems [25,34], topic and category search functionality across data corpora [9], and the identification of turn-taking patterns across heterogeneous data sets [4] demonstrate the viability of this interdisciplinary engagement. However, despite a rich history of ethnomethodological (EM) studies of human-computer interaction and computer-supported collaborative work [8,43], there has been relatively little application of developments within conversation analysis (CA) and ethnomethodology to the core methods, data and phenomena of computer science [46]. This is particularly puzzling given that spoken and text-based interactions are becoming primary sites for networked human-computer interaction. They also form crucial modes of interaction that often serve as both the medium and milieu for real time 'big' and 'broad' social data extraction [18].

This separation is in part due to the continued conflation of computational rules and conversational norms [7]. This category mistake re-issues long-broken promises about the development of 'general AI' every few decades [12], although scholars are becoming less optimistic about when, if ever, this may occur [27]. While domain-specific speech recognition and synthesis systems are now used in basic service interactions, fundamental questions about users' and designers' conceptualization of these systems are still largely unresolved [30]. Despite vast increases in computing power, and EM's increasing impact in user

experience design [26], nothing has so far achieved the kind of 'speech understanding system' that early collaborations between conversation analysts and computer scientists anticipated arriving by the early 90s [14]. This may, on the one hand, be attributable to irreconcilable differences in the way both sides understand social action and human behavior [11:12–13] leaving little common ground for mutual interdisciplinary engagement. On the other hand, and more optimistically, these early explorations of the space between conversation analysis and computer science (CS) may simply have come too early to configure their meeting as a methodologically coherent interdisciplinary engagement.

2 Three potential collaborative framings

In the following sections of this article we consider three prospective avenues for collaboration between CS and CA. Our purpose is not to advocate one position over the other. Rather, our intention is to identify three epistemological and techno-methodological frames through which CA and CS might harness the deluge of big interactional data through a shared analytical focus on social action.

2.1 Automatic annotation of big interactional data?

In order to make interactional data (i.e., video and audio recordings of talk, gesture, and embodied and material conduct) computationally analyzable, researchers must find principled ways to operationalize CA's inductive observations [35], which focus on social actions, rather than clusters of words or other granular linguistic 'features'. Although CA has tended to defer quantification in favour of detailed procedural descriptions [37], recent developments of Sacks's [36] methods have included early-stage experiments using speech recognition for automated CA transcription [25], as well as interactionally grounded approaches to categorization [13,41] and coding [39], The methodological viability of these developments are still under debate within CA [5,38]. However, transcription tools and coding schema that are empirically configured for CA can capture routinely structured generic social practices in everyday talk such as question/answer sequences, specific forms of repair, or task-oriented activities produced within specific institutional contexts such as service call settings [1,42].

CA transcription [16] is not only a means for communicating analyses in research publication, it is a highly developed and well-theorized form of interaction analysis in itself [5,28]. Machine-readable CA transcripts provide interactional detail at orders of magnitude greater than the corpora that are commonly used for machine learning, and will provide the basis for unanticipated discoveries for both fields. What is required is the development of new annotation tools for collaborative transcription enable 'crowdsourcing' of large scale CA data for open access corpora such as in the CABNC [3].

Since CA coding schemes describe actions discovered in real-time interaction, the ecologically-grounded annotated data they produce should, we argue, underpin AI. A large, manually-coded data corpus could be used to train supervised learning classifiers to pick out interactional features in new data. Where algorithms misclassify cases or fail to identify them, collections of these 'edge cases' could inspire new detailed, qualitative analyses [38] and provide useful data for testing and refining new iterations of the system. For example, in a recent study of antagonistic language on social media, twitter threads and stand-alone examples were identified from a large corpus of tweets. These data sets were annotated via a form of inductive inspection that identified a series of interactional features based on CA findings, such as turn taking, recipient design and routine membership categorization practices [19,20]. These data can also be reused to discover more examples, new variations and deviant cases for further inductive inspection [38]. We argue that automatic annotation will create a virtuous research cycle where CA methods for identifying action and automated action classification systems can work together to identify patterns in interactive text and talk.

2.2 Augmenting model and hypothesis construction?

Hypothesis construction and modelling are key aspects of computational social science. However, the ways in which interactional and text-based data are understood, classified and operationalized are a crucial, but often unreported, determining factor in the research process [6]. The same problems of reproducibility that have been plaguing social psychology [29] pose similar threats to machine learning research [24]. The recent experimental turn in CA [21], and calls for theoretical interfaces between CA's inductive methods and the deductive logic of falsifiability [35] offer some practical responses to this unfolding reproducibility crisis [2]. CA and its sister field of Discursive Psychology (DP) have been critical of the widespread tendency to rely on models of interaction that are "largely stipulative or intuitive rather than based on detailed empirical work" [32:18] or that perform only cursory pre-hypothetical qualitative phases of research [31]. Despite long-held skepticism within CA and DP about the viability of hypothetico-deductive reasoning in interaction research, interaction analysts are finding practical solutions to long-standing problems with ecological validity in experimental design [22] where CA and video analysis provide the methodological basis for naturalistic experiments [15] and a 'natural laboratory' [40] for a principled approach to interaction research.

This approach also lends itself to studies of big interactional data on social media, gaming and 'born-digital' text-based mediated interaction. For example, current work on social media analytics has enabled an examination of clustering, homophily and differentiation at the interactional level in relation to Twitter-based activism and online campaigns [45]. This allows us to associate dimensions of a twitter thread with the likelihood of differentiation and 'off topic' conversation and a move towards more explicit forms of 'identity' work and claims management. Such observational findings may be transposed into hypotheses that relate to thread length and counter speech, for example, in modelling the spread of hateful content via social media networks. It remains to be seen whether related concerns can shed light onto the relatively recent emergence and proliferation of voice based big interactional data. This new 'data deluge' is being driven by 'digital assistants' in the home, the workplace and 'artificially intelligent' conversational technologies associated with contemporary marketing strategies. We argue that any movement towards a conversation analytic approach to voice-based big interactional data requires a radical reconsideration of the way we conceptualize features associated with 'voice data' and the framework in and through which such data is understood and classified.

2.3 Natural action processing: Automating CA?

The state of the art in consumer voice interfaces still only allows systems like Amazon's *Alexa* to respond to variations on a single request-formatted phrase [30]. Far more sophisticated 'natural language processing' (NLP) systems are needed than the kind of 'sentiment analysis' often used to process social media [33] or transcribed dialogues, which usually associate simple word and phrase frequencies with meanings via limited statistical models of lexical semantics [10]. For example, in Extract 1 Salesperson (S) has called a prospective Customer (C). One common measure of customer satisfaction may be coded by a customer's 'thank you'.

```
1    S:    >Yeah.< But what I want to find out is when that contract's
2          up for review=Then we can contact at [  t h e  ] ti:me.
3    C:                                          [Yeah w'll]
4    C:    We:ll we're happy with u:hm ↓the people that we're currently
5          using.
6             (0.3)
7    S:    <I'm sure you are, but I wanted to find out when the
8          contract's up for review: so then I can
9          c[all maybe nearer the time, ]
10   C:     [Yeah no we're happy with wh]o we're currently using.
11              (0.5)
12   S:    ↑You don't know when the [contract's up for re]view¿
13   C:                             [ O k a y. Thank you.]
14   C:         ((Hangs up))
```

Extract 1: Part of a call during which a salesperson (S) 'cold calls' a customer (C).

The salesperson re-issues the question of when the customer's contract is due on three occasions at lines 1, 7 and 12. Each time, the customer's response demonstrates that they are not treating what the salesperson says as a *question* to answer, but as an *offer* to refuse. Furthermore, the customer designs their refusals as *assessments* of their existing service, rather than make explicit refusals.

These mismatched turn-and-action formats highlight the arbitrariness of mapping between lexical semantics and social action in conversation. The customer's "Thank you" in the midst of the salesperson's last turn in line 13 highlights the dangers of operationalizing fixed words or phrases as communicative measures. How often the words "thank you" appear in a dialogue is as absurd a measure of gratitude as laughs-per-minute are a measure of amusement [37]. Out of place, these utterances can be rude since their meaning depends on precisely where and how they are used within a sequence of *action*. If CA and computer science researchers work together, they could explore new approaches to natural *action* processing (NAP) that take the centrality of social action into account.

In order to have an impact in machine learning and related approaches in CS, CA needs to work at a larger scale, with new software tools and protocols for data collection, transcription, and for sharing annotations for interactional features. CA has developed a bottom-up inductive research cycle [17] for refining analysis and identifying future cases by studying collections of individual cases in detail. However, if CA were equipped with computer-readable annotations, enabling work on large scale corpora, this ongoing research activity would allow CS access to detailed data for training machine learning systems and algorithm development.

We propose NAP as a critical and practical approach to automation using conversation analysis and related methods for interaction-based feature identification. Replacing the 'L' of NLP with 'A' underlines the importance of starting with action, not language (and interaction, not dialogue). This will involve building systems for scoping and annotating networked communications, as social media moves from an emphasis on text to incorporate speech and embodiment with the arrival of conversational and interactionally enhanced machines.

3 Conclusion

Greater interdisciplinary engagement between conversation analysis (CA) and computer science (CS) would enable CS to use CA data to build more sensitive systems for classifying patterns of behavior in real-time communications, the results of which could further inform understanding of key features and effects of large-scale social organization. The identification of common interactional features could then be re-deployed on data from spoken, multimodal and text-based interaction across a wide variety of social and institutional settings, online platforms and media environments. Successful uses of deep learning to deal with clearly defined problems such as automated translation [44] tend to arise from using new or better data. Using CA-based annotation and coding systems would not only provide new, more detailed data. CA's core focus on social action may help to define new research questions. Using CA at key points in the development of algorithms used for the automated scoping of big social data in real time would engender a move from a design ontology configured by hyper-individualized data points [23] to one that is nuanced and commensurate with social media as an interactional and relational environment.

A final consideration is the extent to which the three approaches outlined above can act as a means of assisting the manual inspection of data streams, augmenting model building and 'normal science' or end up merely supporting the 'dash for automation'. We suggest that all these workflows might be accommodated in a mutually reinforcing manner; albeit within the context of interdisciplinary, reflexive and collaborative work supported by new infrastructural requirements. However, it also represents engagement with 'boundary objects': processes and phenomena that are understood quite differently within computer science, conversation analysis and their cognate disciplines. This form of interdisciplinary engagement is fraught with difficulties and tensions; but it remains a creative and potentially rewarding process. To this extent the practice and process of interdisciplinarity remains key to exploring these avenues robustly, reflexively and with due care and attention.

REFERENCES

[1] Saul Albert, William Housley, Rein Ove Sikveland, and Elizabeth Stokoe. frth. The Conversation Analytic Turing Test: Using Conversation Analysis to Evaluate Google Duplex. (frth.).

[2] Saul Albert and J. P. De Ruiter. 2018. Improving Human Interaction Research through Ecological Grounding. *Collabra: Psychology* 4, 1 (July 2018), 24. DOI:https://doi.org/10.1525/collabra.132

[3] Saul Albert, J. P. De Ruiter, and Laura De Ruiter. 2015. The CABNC. Retrieved from https://saulalbert.github.io/CABNC/

[4] Daniel Angus, Sean Rintel, and Janet Wiles. 2013. Making sense of big text: a visual-first approach for analysing text data using Leximancer and Discursis. DOI:https://doi.org/10.1080/13645579.2013.774186

[5] Galina B. Bolden. 2015. Transcribing as Research: "Manual" Transcription and Conversation Analysis. *Research on Language and Social Interaction* 48, 3 (July 2015), 276–280. DOI:https://doi.org/10.1080/08351813.2015.1058603

[6] Phillip Brooker, William Dutton, and Christian Greiffenhagen. 2017. What would Wittgenstein say about social media? *Qualitative Research* (June 2017). DOI:https://doi.org/10.1177/1468794117713058

[7] Graham Button. 1990. Going Up a Blind Alley: Conflating Conversation Analysis and Computational Modelling. In *Computers and Conversation*, Paul Luff, Nigel Gilbert and David Frolich (eds.). Academic Press, London, 67–90. DOI:https://doi.org/10.1016/B978-0-08-050264-9.50009-9

[8] Graham Button and Paul Dourish. 1996. Technomethodology: paradoxes and possibilities. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. Retrieved from http://dl.acm.org/citation.cfm?id=238394

[9] Susan Conrad. 2002. Corpus Linguistic Approaches for Discourse Analysis. *Annual Review of Applied Linguistics* 22, (March 2002), 75–95. DOI:https://doi.org/10.1017/S0267190502000041

[10] Robin Cooper. 2012. Type theory and semantics in flux. *Handbook of the Philosophy of Science* 14, (2012), 271–323.

[11] Jeff Coulter. 1989. *Mind in action*. Polity Press, Cambridge.

[12] Hubert L Dreyfus. 2012. A history of first step fallacies. *Minds and Machines* 22, 2 (2012), 87–99.

[13] Richard Fitzgerald and William Housley. 2015. *Advances in Membership Categorisation Analysis*. Sage, London.

[14] Nigel Gilbert, Robin Wooffitt, and Norman Fraser. 1990. Chapter 11 - Organising Computer Talk. In *Computers and Conversation*, Paul Luff, Nigel Gilbert and David Frolich (eds.). Academic Press, London, 235–257. DOI:https://doi.org/10.1016/B978-0-08-050264-9.50016-6

[15] Christian Heath and Paul Luff. 2017. The Naturalistic Experiment. *Organizational Research Methods* (December 2017), 1–23. DOI:https://doi.org/10.1177/1094428117747688

[16] Alexa Hepburn and Galina B Bolden. 2017. *Transcribing for social research*. Sage, London.

[17] E. M. Hoey and K. H. Kendrick. 2017. Conversation Analysis. In *Research Methods in Psycholinguistics: A Practical Guide*, A. M. B. de Groot and P.Hagoort (eds.). WileyBlackwell, Hoboken, NJ, 151–173.

[18] W. Housley, R. Procter, A. Edwards, P. Burnap, M. Williams, L. Sloan, O. Rana, J. Morgan, A. Voss, and A. Greenhill. 2014. Big and broad social data and the sociological imagination: A collaborative response. *Big Data & Society* 1, 2 (July 2014). DOI:https://doi.org/10.1177/2053951714545135

[19] William Housley, Helena Webb, Adam Edwards, Rob Procter, and Marina Jirotka. 2017. Digitizing Sacks? Approaching social media as data. *Qualitative Research* (2017), 1468794117715063.

[20] William Housley, Helena Webb, Adam Edwards, Rob Procter, and Marina Jirotka. 2017. Membership categorisation and antagonistic Twitter formulations. *Discourse & Communication* 11, 6 (September 2017), 567–590. DOI:https://doi.org/10.1177/1750481317726932

[21] Kobin H. Kendrick. 2017. Using Conversation Analysis in the Lab. *Research on Language and Social Interaction* (January 2017), 1–11. DOI:https://doi.org/10.1080/08351813.2017.1267911

[22] Alan Kingstone, Daniel Smilek, and John D. Eastwood. 2008. Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology* 99, 3 (August 2008), 317–340. DOI:https://doi.org/10.1348/000712607x251243

[23] Robert W Lake. 2017. Big Data, urban governance, and the ontological politics of hyperindividualism. *Big Data & Society* 4, 1 (June 2017). DOI:https://doi.org/10.1177/2053951716682537

[24] Zachary C. Lipton and Jacob Steinhardt. 2019. Troubling Trends in Machine Learning Scholarship. *Queue* 17, 1 (February 2019), 80:45–80:77. DOI:https://doi.org/10.1145/3317287.3328534

[25] Robert J. Moore. 2015. Automated Transcription and Conversation Analysis. *Research on Language and Social Interaction* 48, 3 (July 2015), 253–270. DOI:https://doi.org/10.1080/08351813.2015.1058600

[26] Robert J. Moore. 2018. A Natural Conversation Framework for Conversational UX Design. In *Human–Computer Interaction Series*. Springer International Publishing, 181–204. DOI:https://doi.org/10.1007/978-3-319-95579-7_9

[27] Vincent C. Müller and Nick Bostrom. 2016. Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In *Fundamental Issues of Artificial Intelligence*. Springer International Publishing, 555–572. DOI:https://doi.org/10.1007/978-3-319-26485-1_33

[28] E Ochs. 1979. Transcription as theory. In *Developmental pragmatics*, Elinor Ochs and B. B. Schieffelin (eds.). Academic Press, New York, 43–72.

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

[30] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 ACM Conference on Human Factors in Computing Systems (CHI'18)*.

[31] Jonathan Potter. 2012. Re-reading Discourse and Social Psychology: Transforming social psychology. *British Journal of Social Psychology* 51, 3 (December 2012), 436–455. DOI:https://doi.org/10.1111/j.2044-8309.2011.02085.x

[32] Jonathan Potter and H te Molder. 2005. Talking cognition: Mapping and making the terrain. In *Conversation and cognition*, Jonathan Potter and Derek Edwards (eds.). 1–54.

[33] Matthew Purver and Stuart Battersby. 2012. Experimenting with Distant Supervision for Emotion Classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (EACL '12), 482–491. Retrieved from http://dl.acm.org/citation.cfm?id=2380816.2380875

[34] Luis Rodríguez, Francisco Casacuberta, and Enrique Vidal. 2007. Computer Assisted Transcription of Speech. In *Pattern Recognition and Image Analysis* (Lecture Notes in Computer Science), 241–248.

[35] J. P. de Ruiter and Saul Albert. 2017. An Appeal for a Methodological Fusion of Conversation Analysis and Experimental Psychology. *Research on Language and Social Interaction* 50, 1 (January 2017), 90–107. DOI:https://doi.org/10.1080/08351813.2017.1262050

[36] Harvey Sacks. 1995. *Lectures on conversation*. Basil Blackwell, Oxford.

[37] Emanuel A Schegloff. 1993. Reflections on Quantification in the Study of Conversation. *Research on Language & Social Interaction* 26, 1 (1993), 99–128. DOI:https://doi.org/10.1207/s15327973rlsi2601_5

[38] Jakob Steensig and Trine Heinemann. 2015. Opening Up Codings? *Research on Language and Social Interaction* 48, 1 (January 2015), 20–25. DOI:https://doi.org/10.1080/08351813.2015.993838

[39] Tanya Stivers. 2015. Coding Social Interaction: A Heretical Approach in Conversation Analysis? *Research on Language and Social Interaction* 48, 1 (January 2015), 1–19. DOI:https://doi.org/10.1080/08351813.2015.993837

[40] E. Stokoe, R. O. Sikveland, and J. Symonds. 2016. Calling the GP surgery: patient burden, patient satisfaction, and implications for training. *British Journal of General Practice* (August 2016). DOI:https://doi.org/10.3399/bjgp16x686653

[41] Elizabeth Stokoe. 2012. Moving forward with membership categorization analysis: Methods for systematic analysis. *Discourse Studies* 14, 3 (June 2012), 277–303. DOI:https://doi.org/10.1177/1461445612441534

[42] Elizabeth Stokoe, Rein Ove Sikveland, Magnus Hamann, Saul Albert, and William Housley. frth. Can humans simulate talking like other humans? Comparing simulated clients to real customers in service inquiriesh. *Discourse Studies* (frth.).

[43] Lucy A Suchman. 1987. *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press, Cambridge.

[44] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (NIPS'14), 3104–3112. Retrieved from http://dl.acm.org/citation.cfm?id=2969033.2969173

[45] Helena Webb, Pete Burnap, Rob Procter, Omer Rana, Bernd Carsten Stahl, Matthew Williams, William Housley, Adam Edwards, and Marina Jirotka. 2016. Digital Wildfires: Propagation, Verification, Regulation, and Responsible Innovation. *ACM Trans. Inf. Syst.* 34, 3 (April 2016), 15:1–15:23. DOI:https://doi.org/10.1145/2893478

[46] Malte Ziewitz. 2017. A not quite random walk: Experimenting with the ethnomethods of the algorithm. *Big Data & Society* 4, 2 (2017). DOI:https://doi.org/10.1177/2053951716682537