Time-constrained Multi-layer Corpus Creation

Katarzyna Budzynska[1,2], *Martin Pereira-Fariña*[2,3], *Dominic De Franco*[2], *Rory Duthie*[2], *Núria Franco-Guillén*[4], *Annette Hautli-Janisz*[5], *Mathilde Janier*, *Marcin Koszowy*[1], *Luana Marinho*[6], *Elena Musi*[7], *Alison Pease*[2], *Brian Plüss*[2], *Chris Reed*[2], *Jacky Visser*[2]

[1]Centre for Argument Technology, Polish Academy of Sciences, **Poland**

[2]Centre for Argument Technology, University of Dundee, **UK**

[3]Universidade de Santiago de Compostela, **Spain**, [4]Griffith University, **Australia**

[5]University of Konstanz, **Germany**, [6]University of Brasília, **Brasil**

[7]Columbia University, **US**

Goals of the annotation process. The paper proposes a new complex method of corpus creation under the constraint of bounded, short period of time available for the annotation process. One important consequence of such a constraint is that it does not leave time for the traditional techniques of corpus evaluation of Inter-Annotator Agreement, IAA. Therefore, we designed, tested and improved a multi-layer annotation process with each subsequent layer aiming to replace IAA with an alternative method allowing for the creation of high-quality corpus.

We built our method on two approaches to corpus creation: *iterative enhancement* (IE) which aims to improve the annotation in several iterations using automatic techniques to look for inconsistencies in the manual annotation [5], and *agile corpus creation* (ACC) which replaces the traditional, linear-phase approach with a cyclic and iterative small-step process [12, 1]. The layers in our approach can be viewed as such iterative cycles which aim to improve the result of the annotation, however, our process is also adapted to handle time-constraint and the annotation of complex linguistic phenomena (dialogical argumentation) where (semi-)automatic methods such as IE cannot be successfully applied. Moreover, the full multi-layer annotation process was iterated three times which allowed us to not only improve the corpus as in ACC, but also to improve the annotation process itself.

Figure 1: Infographics in *Argument Analytics*: like-mindedness (on the left) and divisive issues (on the right).

The annotation process was designed for a project run in 2017 in partnership with the BBC, which aimed to develop *Argument Analytics* [7, 10], a set of analytics (metrics [8]), for the BBC Radio 4 programme, *Moral Maze*. This sense-making argument technology provides infographics (i.e. an intuitive overview of the debate using graphic visualisations) presented to a large, non-expert audience in a real environment on the BBC webpages (see Figure 1 and `bbc.arg.tech`).

We worked with two radio programmes (from 2012 & 2017) and one TV programme (2017) on the morality of abortion. In order to release Argument Analytics in real-time, i.e. at the same time as the 2017 programmes were broadcast, they were pre-recorded allowing us to run the whole process of preparing Argument Analytics in the 48 hours before broadcast.

In each run for these three 45 minute programmes, the *Argument Analysis Team* (AAT) was allocated 8 hour time window to analyse a programme using OVA+ tool [6] and an annotation scheme [2] built upon Inference Anchoring Theory, IAT [3] (`arg.tech/iatguidelines`).[1] The time required for the annotation was significantly longer than 8h (estimated as 45 hours for the basic layer of annotation), thus we ran three rounds of training in IAT annotation for 60 candidates from whom 10 passed the final test and were recruited to AAT. As a result, the design of the annotation process had to address several challenges, including time-constraint, complexity of annotation scheme and newly recruited, less experienced AAT members.

[1] Argument data was in this work generated manually, but the process could in principle be automated by applying argument mining techniques (*cf.* [9, 4]). Still the current state-of-the-art does not guarantee high quality of such an automatic annotation.

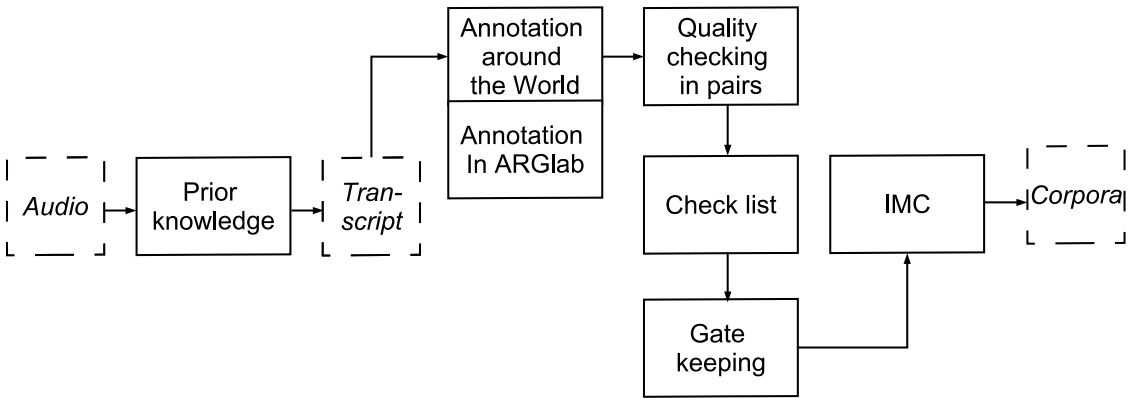

Figure 2: Six layers of time-constrained process of corpus creation.

Design of the annotation process. We developed a process of corpus creation with six layers of iterative cycles of manual annotation (see Figure 2). Argument Analysis Team consists of 18 members, including 8 AATs who were experienced in IAT analysis (6 on-site in the lab and 2 joining us remotely through an Internet communication platform), and 10 inexperienced AATs (6 on-site and 4 joining remotely). The annotators from different parts of the world helped us to slightly speed up the process by making use of different time zones, where the annotation could have started during the night in the UK where most of the team was based. The whole process was coordinated by one member of the team.

The preparatory phase, the *Prior Knowledge Layer*, aimed to capitalise on having the audio earlier than the transcript to familiarise ourselves with the content of the programmes while waiting for the transcription to be prepared. All AATs available on the late evening before the day of annotation process (i.e. at the beginning of the full 48 hours) met to listen together and discuss a programme. This helped us to create a general overview of the content which was then particularly useful when annotating smaller, isolated excerpts into which a transcript was split for annotation.

Next, in the *Basic Annotation Layer* excerpts were allocated to all AATs to be analysed around the world and then together in the lab to allow for discussing how to annotate difficult parts of the programme. The excerpts were annotated here using a standard corpus linguistic procedure applying IAT annotation scheme.

Once we had an initial set of argument maps, it was passed to the *Quality Assessing Layer*. At this point, the team was divided into smaller task groups with only some of AATs continuing basic annotation. During this phase, an author of each map was explaining to another AAT the decisions behind the annotation in order to reach an agreement on the annotation created in the previous step. Inexperienced AATs were always paired with an experienced AATs.

Then in the *Check-List Layer*, another group of annotators was going through each map to compare it against a check-list with the basic rules from IAT guidelines (`arg-tech.org/IATchecklist`) to avoid the most common mistakes in annotation. In order to create final versions of argument maps for each excerpt, in the *Gate Keeping Layer* the two most experienced analysts were running the last check-up on each map and apply final corrections. Finally, in the *IMC Layer* all maps where connected together in one large argument network by using our technique of Inter-Map Correspondence [11]. This network was then submitted to a corpus.

The annotation process was additionally changed and improved after each run of annotation. Between the first and the second run we focused on reducing the time of annotation and the number of errors by introducing, e.g.: the allocation of the longest excerpts to the best and fastest AATs; and the addition of the layers of Check-list and Gate-keeping. Between the second and the third run, the improvements focused on making the process more structured and controlled by: reducing the length of the excerpts to make them easier to manage; and allocating AATs to specific layers which helped them to concentrate on single task at a time.

The proposed process allows for creating high-quality corpora under the constraint of a short time available for the demanding task of argument annotation. As the standard methods of corpus evaluation were unfeasible, we improved the quality *internally* by iterative cycles of annotation at six layers, and *externally* by three cycles of full annotation process. As far as we know, this is the first time such a complex process has been tried, successfully, on real-world data from the media, and with a hard deadline for making the results available to the public.

Acknowledgments

We would like to acknowledge that the work reported in this paper has been supported in part by EPSRC in the UK under grants EP/N014871/1, EP/M506497/1, in part by the Volkswagen Foundation (VolkswagenStiftung) in Germany under grant 92 182, in part by "Ministerio de Economía y Competitividad", "Consellería de Cultura, Educación e Ordenación Universitaria" in Spain under grants TIN2014-56633-C3-1-R and TIN2014-56633-C3-3-R (accreditation 2016-2019, ED431G/08) and Postdoctoral Fellowship (accreditation 2016/048-0, ED481B) and the European Regional Development Fund (ERDF). We would also like to thank our colleagues from Argument Analysis Team – Rory Magowan, Nicole Orr, Javier Ruiz and Teodora Udrea.

References

[1] Bea Alex, Claire Grover, Rongzhou Shen, and Mijail Kabadjov. Agile corpus annotation in practice: An overview of manual and automatic annotation of CVs. In *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010*, pages 29 – 37, Uppsala, Sweden, 15-16 July, 2010.

[2] Katarzyna Budzynska, Mathilde Janier, Chris Reed, and Patrick Saint-Dizier. Theoretical foundations for illocutionary structure parsing. *Argument and Computation*, 2016.

[3] Katarzyna Budzynska and Chris Reed. Whence inference? Technical report, University of Dundee, 2011.

[4] Katarzyna Budzynska and Serena Villata. Processing natural language argumentation. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of Formal Argumentation*, pages 576–625, 2017.

[5] M. Dickinson and D. Tufiş. Iterative enhancement. In N. Ide and J. Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 257 – 276, 2017.

[6] Mathilde Janier, John Lawrence, and Chris Reed. OVA+: An argument analysis interface. In S. Parsons and et al., editors, *Proceedings of the Fifth International COMMA*, pages 463–464. IOS Press, 2014.

[7] John Lawrence, Rory Duthie, Katarzyna Budzynska, and Chris Reed. Argument analytics. *Proceedings of the Sixth International COMMA. Frontiers in Artificial Intelligence and Applications*, pages 371–378, 2016.

[8] John Lawrence, Mark Snaith, Barbara Konat, Katarzyna Budzynska, and Chris Reed. Debating technology for dialogical argument: Sensemaking, engagement, and analytics. *ACM Transactions on Internet Technology*, 17(3):1–23, 2017.

[9] Marco Lippi and Paolo Torroni. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, 16(2):10:1–10:25, 2016.

[10] Chris Reed, Katarzyna Budzynska, Martin Pereira-Fariña, Dominic De Franco, Rory Duthie, Marcin Koszowy, John Lawrence, Alison Pease, Brian Pluss, Mark Snaith, Debela Tesfaye, and Jacky Visser. Large-scale deployment of argument analytics. In *Argumentation and Society – the workshop at the 7th International Conference on Computational Models of Argument (COMMA 2018)*, 20xx, under review.

[11] Jacky Visser, Rory Duthie, John Lawrence, and Chris Reed. Intertextual correspondence for integrating corpora. In *11th edition of Language Resources and Evaluation Conference (LREC)*, pages 1–7, 2018.

[12] Holger Voormann and Ulrike Gut. Agile corpus creation. *Corpus Linguistics and Linguistic Theory*, 4(2):235 – 251, 2008.