

MMAX2 for coreference annotation

Mateusz Kopec

Institute of Computer Science, Polish Academy of Sciences,
Jana Kazimierza 5, 01-248 Warsaw, Poland
m.kopec@ipipan.waw.pl

Abstract

This article presents major modifications in the MMAX2 manual annotation tool, which were implemented for the coreference annotation of Polish texts. Among other, a new feature of adjudication is described, as well as some general insight into the manual annotation tool selection process for the natural language processing tasks.

1 Introduction

Recently published Polish Coreference Corpus (PCC) (Ogrodniczuk et al., 2013) contains a large number of Polish texts annotated manually with coreference. During the initial stage of this project in 2011, a tool had to be selected for the manual text annotation with coreference.

First issue considered during the selection was the alternative of desktop versus online annotation tool. Recently, online annotation tools are becoming increasingly popular (see for example BRAT (Stenetorp et al., 2012)), with their advantages such as the possibility to monitor the current state of annotation and make changes to the annotation tool easily, without the need to communicate with the annotators. However, in our opinion the choice should be made mainly based on the annotators' preferences, as their work efficiency is crucial for the cost of the task.

10 linguists (which were annotators in previous projects conducted by the authors of this paper) were asked in anonymous survey to choose one of following three options: 1) that they prefer an online annotation tool (not requiring any installation), 2) a desktop tool with the possibility to work without constant internet access, 3) that they do not have preference. Only one person chose online tool and one person chose the third option, leaving no choice for the annota-

tion task organizers other than to prepare a desktop application. The drawback of this approach was the need to manage text distribution among the annotators, as all the work was done on local computers. Distribution was controlled by the DistSys application, available at the webpage zil.ipipan.waw.pl/DistSys.

After some analysis of the features required by the project's scope, the choice was narrowed to only two tools: MMAX2 (Müller and Strube, 2006) and Palinka (Orăsan, 2003). The problem with the latter was that it was not error-prone, and lack of publicly available sources did not allow to make project-specific corrections. Therefore MMAX2 environment was chosen for the manual coreference annotation. It is a general purpose cross-platform desktop annotation tool (written in Java), created to be configurable for many natural language annotation efforts. Its source is publicly available at the webpage mmax2.net. MMAX2 interface consists of the main window with the text being annotated (see for example figure 5) and several smaller windows facilitating various tasks (all the other figures).

2 Annotation scope

The annotation scope of the Polish Coreference Corpus project consisted of several subtasks for an annotator to perform for each text. Because the automatic preannotation was used, annotator's job consisted not only of addition of new markables, but also removal and correction of existing ones. Subtasks to perform were to:

- mark all mentions,
- mark each mention's semantic head (choose one word the mention consists of),
- cluster coreferent mentions,
- mark dominant expression of each cluster,

- mark quasi-identity links.

MMAX2 was easy to configure for most of the subtasks, except the dominant expressions (there is no possibility to define attributes of clusters, only markables) and the choice of semantic head (available types of attributes did not include a possibility to define a mention-dependent attribute).

Because an estimate of inter-annotator agreement had to be calculated for the corpus, some texts were annotated independently by two annotators. Agreement measures were then calculated, but as single gold standard annotation was needed for the corpus, they also had to be merged into single version by the adjudicator. This feature was also not present in MMAX2.

3 New features

Even with it's great number of features, there was no possibility to use MMAX2 without any changes in it's source code. Some changes were the result of project's scope requirements, some were added in response to annotator requests. New implemented features include:

1. Internationalization – the interface of the tool is available in English and Polish and can be easily translated to other languages. Polish version was used by the annotators, but for international articles about the tool (such as this one) the English interface was used.
2. Semantic head selection – a dropdown list allows to choose one of the tokens mention consists of as it's semantic head. This head is also underlined in markable browser.
3. Storing user setting – which windows are opened, where are they located, what is the font and it's size – these and other user settings are saved and automatically restored when the application is reopened.
4. Dominant expressions – clusters can have their attribute: a dominant expression, which can be selected as one of the mentions from the cluster or any other expression entered by the user.
5. Undo button, reverting last action – very useful feature to revert the last change, regardless of it's nature.

Mention	A_mentions.xml	B_mentions.xml
[gorzką czekoladę], której polykam od 2 - 10 kostek dziennie]	-	+
[gorzką czekoladę]	-	+
[pianki i][żelki anyżkowe]	-	+

Figure 1: Adjudication of mentions

Mention	A_mentions.xml	B_mentions.xml
[gorzką czekoladę]	cluster 1 (2 occ.)	singleton
[tą 99 %]	singleton	cluster 19 (2 occ.)
[jakieś super słodkie ciasteczka , pianki i]	cluster 4 (4 occ.)	cluster 20 (2 occ.)
[żelki anyżkowe]	cluster 4 (4 occ.)	cluster 21 (2 occ.)
[nadają]	cluster 1 (2 occ.)	singleton
[gorzką czekoladę]	cluster 1 (2 occ.)	singleton
[tą 99 %]	singleton	cluster 19 (2 occ.)
[jakieś super słodkie ciasteczka , pianki i]	cluster 4 (4 occ.)	cluster 20 (2 occ.)
[żelki anyżkowe]	cluster 4 (4 occ.)	cluster 21 (2 occ.)
[nadają]	cluster 4 (4 occ.)	singleton
[nadają się]	singleton	cluster 21 (2 occ.)

Figure 2: Adjudication of clustering

6. Merge two mentions – user can merge two mentions into one with a single click, summing their words and merging their clusters. Very useful feature when for example one is correcting automatic annotation, which failed to recognize a long named entity name and instead created two entities, each in it's separate cluster.
7. Improved browser operability – browsers allow to operate on mentions, links and clusters, not only to view them.
8. Adjudication feature – it will be covered in detail in the next section.

4 Adjudication feature

Adjudication feature of the new *Superannotation plugin* allows to compare two versions of annotation of the same text and merge them into one, adjudicated version. The design is based on the original MMAX2 Diff plugin, which allowed to see the differences between two annotations, yet it was not possible to merge them into one. The readability of the differences was also limited and it was improved in our tool.

The adjudication process starts with opening one annotation in standard way and then the other via the menu in *Superannotation plugin* and consist of several steps, each merging particular layer:

1. Mentions – first we need to merge mention annotations. Differences between the two annotations are shown in the figure 1. First column shows the mention content (and this is constant in all steps of adjudication), second shows if that mention is in the first annotation, third column shows if it is in the second annotation ("+" if yes, "-" if not). Single click at "+" or the first column highlights given span in the main window. Double click at one of the last two columns selects the clicked version as the proper one and changes the annotation in the other file to match the clicked version. After such double click, the difference disappears and that row vanishes. After all rows from that step are gone, mention annotations in both versions are the same and we can proceed to the next step.
2. Comments – this time first column again shows each mention, for which there is a difference in comments in both annotations. Double clicking at 2nd or 3rd column resolves the difference in given row.
3. Heads – similar to comments, by double-clicking we can adjudicate differences in head annotation.
4. Links – analogously as with heads, we merge near-identity links annotations.
5. Clusters – this is the most complex adjudication task. At this point we surely have the same set of mentions in both annotations, but they may be clustered differently. Figure 2 presents how differences in clustering are visualized. Mentions with the same color in one column are in the same cluster (they have also the same cluster number). For example, two occurrences of mention *gorzką czekoladę* are in the same cluster according to the first annotation, and are singletons according to the second annotation. Single click on any of these options will show it in the main application window, while double click will choose the clicked version as the gold one and update the other to match it.
6. Dominating expressions – as the clusters are now the same, the only annotation left considers cluster attributes: dominating expressions.

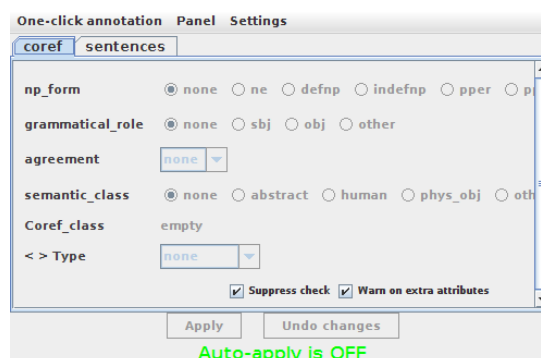


Figure 3: Mention attributes – original MMAX2

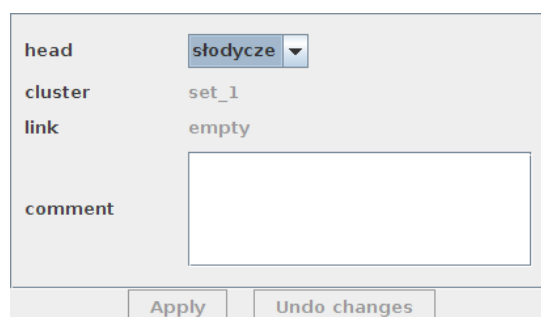


Figure 4: Mention attributes – simplified

Key point of the adjudication procedure is to merge all differences at a given level before proceeding to the next one. This way, after we resolve all differences in the dominating expressions, we are certain that our annotations are fully merged and in fact the same.

5 Removed features

Because MMAX2 is a generic tool, the first impression is that it is very complicated. Numerous options, many windows and menus are overwhelming and could increase the time of creating a manual for the annotators (often writing a lot of text only to inform which options should not be changed). Therefore we removed many options and simplified the interface to leave only the features required by our annotation task. Compare for example the mention attribute window from the original MMAX2 in figure 4 and in our version in figure 3. Removed features included:

- Distinction of multiple annotation levels – scope of the project considers only one level, and the need to explicitly select it in many places (for example in markable browser) is unnecessary.
- Possibility to edit the base text – as we per-

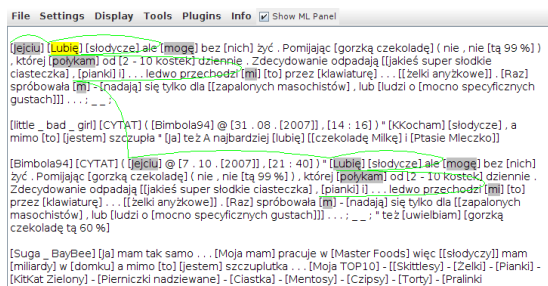


Figure 5: Unnecessary arcs

formed the inter-annotator agreement analysis, the base text could not be changed.

- Arcs between coreferent mentions in cluster (see figure 5 for original visualization) – from our experience, they decrease the readability of the cluster annotation. As the mentions in cluster are already highlighted, there is no need to show the arcs connecting them (the arcs are not clickable as in BRAT).
- MMAX Query Language – MMAX2 facilitates a query language to search for the annotations fulfilling given properties. In our opinion this feature seems more appropriate for an analyst, not an annotator. Moreover, results of such querying would be more informative for a collection of texts, not a single document.
- Kappa statistic and coincidence matrix calculation for multiple annotations of a single text – again, this feature seems more appropriate for an analyst and for the whole corpus, not a single text.

6 Conclusion

Every unnecessary action, which has to be repeated numerous times by a human annotator, has a significant cost in terms of time and money. We claim that annotation efforts are more efficient when there is a step of tool customization (or even design and implementation from scratch) beforehand and also during the process, based on the feedback from the annotators. Using general-purpose tools has a clear benefit of cheap and fast initialization of the project, but also there are major drawbacks: a compromise between the project needs and the tool capabilities. As we have seen, even a tool with great customization options such as MMAX2 doesn't have all the features one would need.

Experience of the PCC project shows, that instead of trying to provide a general, configurable annotation tool (which is very complex due to its wide application possibilities), another way to proceed is to create simple, well designed tool focused on specific task. Such tool can be then customized or extended by qualified programmers without much effort and then provide great efficiency of the annotation process.

Presented version of MMAX2 with its source code is available at <http://zil.ipipan.waw.pl/MMAX4CORE> webpage. We encourage it's use and modification for other coreference annotation projects.

Acknowledgments

The work reported here was cofounded by the Computer-based methods for coreference resolution in Polish texts project financed by the Polish National Science Centre (contract number 6505/B/T02/2011/40) and by the European Union from resources of the European Social Fund. Project PO KL „Information technologies: Research and their interdisciplinary applications”.

References

- Christoph Müller and Michael Strube. 2006. Multi-level annotation of linguistic data with mmax2. In Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, pages 197–214. Peter Lang, Frankfurt a.M., Germany.
- Maciej Ogrodniczuk, Katarzyna Głowińska, Mateusz Kopeć, Agata Savary, and Magdalena Zawisławska. 2013. Polish coreference corpus. In Zygmunt Vetulani, editor, *Proceedings of the 6th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 494–498, Poznań, Poland. Wydawnictwo Poznańskie, Fundacja Uniwersytetu im. Adama Mickiewicza.
- Constantin Orăsan. 2003. PALinkA: a highly customizable tool for discourse annotation. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialog*, pages 39 – 43, Sapporo, Japan, July, 5 -6.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France, April. Association for Computational Linguistics.