

Play the Language: Play Coreference

Barbora Hladká and Jiří Mírovský and Pavel Schlesinger

Charles University in Prague

Institute of Formal and Applied Linguistics

e-mail: {hladka, mirovsky, schlesinger@ufal.mff.cuni.cz}

Abstract

We propose the PlayCoref game, whose purpose is to obtain substantial amount of text data with the coreference annotation. We provide a description of the game design that covers the strategy, the instructions for the players, the input texts selection and preparation, and the score evaluation.

1 Introduction

A collection of high quality data is resource-demanding regardless of the area of research and type of the data. This fact has encouraged a formulation of an alternative way of data collection, "Games With a Purpose" methodology (GWAP), (van Ahn and Dabbish, 2008). The GWAP methodology exploits the capacity of Internet users who like to play on-line games. The on-line games are being designed to generate data for applications that either have not been implemented yet, or have already been implemented with a performance lower than human. Moreover, the players work simply by playing the game - the data are generated as a by-product of the game. If the game is enjoyable, it brings human resources and saves financial resources. The game popularity brings more game sessions and thus more annotated data.

The GWAP methodology was formulated in parallel with design and implementation of the on-line games with images (van Ahn and Dabbish, 2004) and subsequently with tunes (Law et al., 2007),¹ in which the players try to agree on a caption of the image/tune. The popularity of the games is enormous so the authors have succeeded in the basic requirement that the annotation is generated in a substantial amount. Then the *Onto games* appeared (Siorpaes and Hepp,

2008), bringing a new type of input data to GWAP, namely video and text.²

The situation with text seems to be slightly different. One has to read a text in order to identify its topics, which takes more time than observing images, and the longer text, the worse. Since the game must be of a dynamic character, it is unimaginable that the players will spend minutes reading an input text. Therefore, the text must be opened to the players 'part' by 'part'.

So far, besides the Onto games, two more games with texts have been designed: *What did Shannon say?*³, the goal of which is to help the speech recognizer with difficult-to-recognize words, and *Phrase Detectives*⁴ (Kruschwitz, Chamberlain, Poesio, 2009), the goal of which is to identify relationships between words and phrases in a text.

Motivated by the GWAP portal, the LGame portal⁵ has been established. Seven key properties that any game on the LGame portal will satisfy were formulated – see Table 1.

The LGame portal has been opened with the *Shannon game*, a game of intentionally hidden words in the sentence, where players guess them, and the *Place the Space* game, a game of word segmentation.

Within a systematic framework established at the LGame portal, the games *PlayCoref*, *PlayNE*, *PlayDoc* devoted to the linguistic phenomena dealing with the contents of documents, namely coreference, named-entities, and document labels, respectively, are being designed in parallel but implemented subsequently since the GWAPs are open-ended stories the success of which is hard to estimate in advance. These games are designed for Czech and English by default. However, the game rules are language independent.

¹www.gwap.org

²www.ontogame.org

³lingo.clsp.jhu.edushannongame.html

⁴www.phrasedetectives.org

⁵www.lgame.cz

1. During the game, the data are collected for the natural language processing tasks that computers cannot solve at all or not well enough.
2. Playing the game only requires a basic knowledge of the grammar of the language of the game. No extra linguistic knowledge is required.
3. The game rules are designed independently of the language of the game.
4. The game is designed for Czech and English by default.
5. During the game, the players have at least a general idea of what their opponent(s) do.
6. The game is designed for at least two players (also a computer can be an opponent).
7. The game offers several levels of difficulty (to fit a vast range of players).

Table 1: Key properties of the games on the LGame portal.

We have decided to implement the PlayCoref first. Coreference crosses the sentence boundaries and playing coreference offers a great opportunity to test players' willingness to read a text part by part, e.g. sentence by sentence. In this paper, we discuss various aspects of the PlayCoref design.

2 Coreference

Coreference occurs when several referring expressions in a text refer to the same entity (e.g. person, thing, reality). A *coreferential pair* is marked between subsequent pairs of the referring expressions. A sequence of coreferential pairs referring to the same entity in a text forms a *coreference chain*.

Various projects on the coreference annotation by linguists are running. We mention two of them – the Prague Dependency Treebank 2.0 and the coreference task for the sixth Message Understanding Conference.

Prague Dependency Treebank 2.0 (PDT 2.0)⁶ is the only corpus establishing the coreference annotation on a layer of meaning, so-called tectogrammatical layer (t-layer). The annotation includes grammatical and textual coreference. Extended textual coreference (covering additional categories) is being annotated in PDT 2.0 in an ongoing project (Nedoluzhko, 2007).

Sixth Message Understanding Conference – the coreference task (MUC-6)⁷ operates on a surface layer. The coreferential pairs are marked between pairs of the categories nouns, noun phrases, and pronouns.

⁶ufal.mff.cuni.cz/pdt2.0

⁷cs.nyu.edu/faculty/grishman/muc6.html

3 The PlayCoref Game

Motivation The PDT 2.0 coreference annotation (including the annotation scheme design, training of the annotators, technical and linguistic support, and annotation corrections) spanned the period from summer 2002 till autumn 2004. Each of two annotators annotated one half out of 3,165 documents. We are aware that coreferential pairs marked in the PlayCoref sessions may differ from the PDT 2.0 coreference annotation. However, the following estimates reinforce our motivation to use the GWAP technology on texts: assuming that (1) the PlayCoref is designed as a two-player game, (2) at least one document is being present in each session, (3) the session lasts up to 5 minutes and (4) the players play half an hour a day, then at least 6 documents will be processed a day by two players. This means that 3,165 documents will be annotated by two players in 528 days, by eight players in 132 days, by 32 players in 33 days etc., and by 128 players in 9 days.

Strategy The game is designed for two players. The game starts with several first sentences of the document displayed in the players' sentence window. According to the restrictions put on the members of the coreferential pairs, parts of the text are unlocked while the other parts are locked. Only unlocked parts of the text are allowed to become a member of the coreferential pair. In our case, only nouns and selected pronouns are unlocked.⁸ In Table 2, we provide a list of the locked pronoun's sub-part-of-speech classes (as designed in the Czech positional tag system). Pronouns of the other sub-part-of-speech classes are unlocked. The selection of the locked pronoun's sub-part-of-speech classes is based on the fact that some types of pronouns usually corefer with parts of the text larger than one word. This type of coreference cannot be annotated without a linguistic knowledge and without training. Therefore it must be omitted for the purposes of the PlayCoref game.

The players mark coreferential pairs between the unlocked words in the text (no phrases are allowed). They mark the coreferential pairs as undirected links.⁹ After the session, the coreference

⁸A tagging procedure is used to get the part-of-speech classes of the words.

⁹This strategy differs from the general conception of coreference being understood as either the anaphoric or cataphoric relation depending on "direction" of the link in the text. We believe that the players will benefit from this sim-

Locked pronouns: subPOS and its description	
D	Demonstrative ("ten", "onen", ..., lit. "this", "that", "that", ... "over there", ...)
E	Relative "což" (corresponding to English which in subordinate clauses referring to a part of the preceding text)
L	Indefinite "všechn", "sám" (lit. "all", "alone")
O	"svůj", "nesvůj", "tentam" alone (lit. "own self", "not-in-mood", "gone")
Q	Relative/interrogative "co", "copak", "cožpak" (lit. "what", "isn't-it-true-that")
W	Negative ("nic", "nikdo", "nijaký", "žádný", ..., lit. "nothing", "nobody", "not-worth-mentioning", "no"/"none")
Y	Relative/interrogative "co" as an enclitic (after a preposition) ("oč", "nač", "zač", lit. "about what", "on"/"onto" "what", "after"/"for what")
Z	Indefinite ("nějaký", "některý", "čikoli", "cosi", ..., lit. "some", "some", "anybody's", "something")

Table 2: List of the pronoun's sub-part-of-speech classes in the Czech positional tag system locked for the PlayCoref.

chains are automatically reconstructed from the coreferential pairs marked.

During the session, the number of words the opponent has linked into the coreferential pairs is displayed to the player. The number of sentences with at least one coreferential pair marked by the opponent is displayed to the player as well. Revealing more information about the opponent's actions would affect the independency of the players' decisions.

If the player finishes pairing all the related words in a visible part of the document (visible to him), he asks for the next sentence of the document. It appears at the bottom of his sentence window. The player can remove pairs created before at any time and can make new pairs in the sentences read so far. The session goes on this way until the end of the session time.

Instructions for the Players Instructions for the players must be as comprehensible and concise as possible. To mark a coreferential pair, no linguistic knowledge is required. It is all about the text comprehension ability.

Input Texts In the first stage of the project, documents from PDT 2.0 and MUC-6 will be used in the sessions, so that the quality of the game data can be evaluated against the manual coreference annotation.

Since the PDT 2.0 coreference annotation operates on the tectogrammatical layer and PlayCoref on the surface layer, the coreferential pairs of the t-layer must be projected to the surface first. The basic steps of the projection are depicted in Figure 1. Going from the t-layer, some of the coreferential

plication and that the quality of the game data will not be decreased.

pairs get lost because their members do not have their counterparts on surface.¹⁰ From the remaining coreferential pairs, those between nouns and unlocked pronouns are selected. In the final game documents, the difference between the grammatical, textual and extended textual coreference is omitted, because the players will not be asked to distinguish them. Table 3 shows the number of coreferential pairs in various stages of the projection.

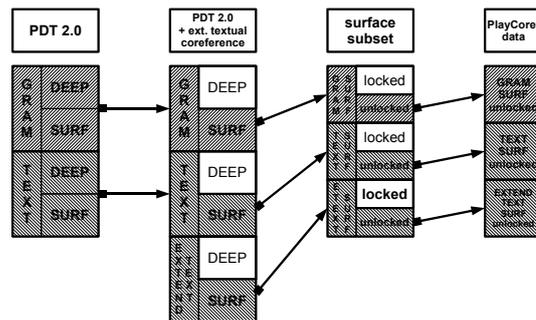


Figure 1: Projection of the PDT coreference annotation to the surface layer. The first step depicts the annotation of the extended textual coreference. Pairs that have no surface counterparts are marked *DEEP*, pairs with surface counterparts are marked *SURF*. Pairs suitable for the game are marked *unlocked*.

Data from the coreference task on the sixth Message Understanding Conference can be used in a much more straightforward way. Coreference is annotated on the surface and no projection is needed. The links with noun phrases are disregarded.

	PDT 2.0	PDT 2.0 + ext.	surface subset	PlayCoref
# coref. pairs	45	96	70	33

Table 3: Number of coreferential pairs (in thousands) in various stages of projection. Counts in the second, third and fourth columns are extrapolated on the basis of data annotated so far, which is about 200 thousand word tokens in 12 thousand sentences (out of 833 thousand tokens in 49 thousand sentences in PDT 2.0). Type of the coreferential pairs, either grammatical or textual one, is not distinguished.

Scoring The players get points for their coreferential pairs according to the equation $pts_A = w_1 * ICA(A, acr) + w_2 * ICA(A, B)$ where A and B are the players, acr is an automatic coreference resolution procedure, weights $0 \leq w_1, w_2 \leq 1$, $w_1, w_2 \in R$ are set empirically, and ICA stands for the inter-coder agreement that we can simultaneously express either by the F-measure or Krippen-

¹⁰Czech is a 'pro-drop' language, in which the subject pronoun on 'he' has a zero form (also in feminine, plural, etc.).

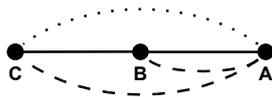


Figure 2: Player '1' pairs (A,C) – the dotted curve; player '2' pairs (A,B) and (B,C) – the solid lines; player '3' pairs (A,B) and (A,C) – the dashed curves. Although players '1' and '2' do not agree on the coreferential pairs at all, '1' and '3' agree only on (A,C) and '2' and '3' agree only on (A,B), for the purposes of the coreference chains reconstruction, the players' agreement is higher: players '1' and '2' agree on two members of the coreferential chain: A and C, players '1' and '3' agree on A and C as well, and players '2' and '3' achieved agreement even on all three members: A, B, and C.

dorff's α (Artstein and Poesio, 2008). The score is calculated at the end of the session and no running score is being presented during the session. Otherwise, the players might adjust their decisions according to the changes in the score. Obviously, it is undesirable.

Assigning a score to the players deals with the coreferential pairs. However, motivated by (Pasonneau, 2004) and others, the evaluation handles the coreferential pairs in a way demonstrated in Figure 2.

PlayCoref vs. PhraseDetectives At least to our knowledge, there are no other GWAPs dealing with the relationship among words in a text like PhraseDetectives and PlayCoref. Nevertheless, there are many differences between these two games – the main ones are enumerated in Table 4.

<i>PlayCoref</i>	<i>PhraseDetectives</i>
detection of coreference chains	anaphora resolution
two-player game	one-player game
a document presented sentence by sentence	a paragraph presented at once
–	checking the pairs marked in the previous sessions
pairing not restricted to the position in the text	the closest antecedent
simple instructions	players training
scoring with respect to the automatic coreference resolution and to the opponent's pairs	scoring with respect to the players that play with the same document before
coreferential pairs correction	no corrections allowed

Table 4: PlayCoref vs. PhraseDetectives.

4 Conclusion

We propose the PlayCoref game, a concept of a GWAP with texts that aims at getting the documents with the coreference annotation in substan-

tially larger volume than can be obtained from experts. In the proposed game, we introduce coreference to the players in a way that no linguistic knowledge is required from them. We present the game rules design, the preparation of the game documents and the evaluation of the players' score. A short comparison with a similar project is also provided.

Acknowledgments

We gratefully acknowledge the support of the Czech Ministry of Education (grants MSM-0021620838 and LC536), the Czech Grant Agency (grant 405/09/0729), and the Grant Agency of Charles University in Prague (project GAUK 138309).

References

- Ron Artstein, Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, December 2008, vol. 34, no. 4, pp. 555–596.
- Udo Kruschwitz, Jon Chamberlain, Massimo Poesio. 2009. (Linguistic) Science Through Web Collaboration in the ANAWIKI project. In *Proceedings of the WebSci'09: Society On-Line*, Athens, Greece, in press.
- Lucie Kučová, Eva Hajičová. 2005. Coreferential Relations in the Prague Dependency Treebank. In *Proceedings of the 5th International Conference on Discourse Anaphora and Anaphor Resolution*, San Miguel, Azores, pp. 97–102.
- Edith. L. M. Law et al. 2007. Tagatune: A game for music and sound annotation. In *Proceedings of the Music Information Retrieval Conference*, Austrian Computer Soc., pp. 361–364.
- Anna Nedoluzhko. 2007. Zpráva k anotování rozšířené textové koreference a bridging vztahů v Pražském závoslostním korpusu (Annotating extended coreference and bridging relations in PDT). Technical Report, UFAL, MFF UK, Prague, Czech Republic.
- Rebecca J. Pasonneau. 2004. Computing Reliability for Coreference. *Proceedings of LREC*, vol. 4, pp. 1503–1506, Lisbon.
- Katharina Siorpaes and Martin Hepp. 2008. Games with a purpose for the Semantic Web. *IEEE Intelligent Systems Vol. 23, number 3*, pp. 50–60.
- Luis van Ahn and Laura Dabbish. 2004. Labelling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, pp. 319–326.
- Luis van Ahn and Laura Dabbish. 2008. Designing Games with a Purpose. *Communications of the ACM*, vol. 51, No. 8, pp. 58–67.