Appeared in *Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium, Lisbon, September* 18-20, 2002.

Handling complex anaphora resolution cases

Dan Cristea*[♦], Gabriela E. Dima[■], Oana-Diana Postolache*, Ruslan Mitkov[□]

*University "Al.I.Cuza" of Iaşi, Faculty of Computer Science, 16 Berthelot St., 6600 Iaşi, Romania
{dcristea, oanap}@infoiasi.ro
[♦]Romanian Academy, Institute for Theoretical Computer Science, Iasi branch, 11, Bd. Carol I, 6600 Iaşi, Romania
[■]University "Al.I.Cuza" of Iaşi, Faculty of Letters, 11, Carol I Bvd., 6600 Iaşi, Romania
g.dima@mail.uaic.ro
[□]University of Wolverhampton, School of Humanities, Languages and Social Sciences,
Stafford St., WV1 1SB, Wolverhampton, United Kingdom
R.Mitkov@wlv.ac.uk

Abstract

The paper investigates difficult problems that could arise in anaphora resolution and proposes some solutions within the framework of a general anaphora resolution solver. Departing from the current research settings that deal with anaphora resolution on contiguous corpora, our investigation uses instead a collection of carefully hand chosen examples. The research is motivated by the belief that interpretation of free language in modern applications, especially those related to semantic web, requires more and more sophisticated tools.

1. Introduction

It is already well known that anaphora studies reveal a psychological threshold around the value of 80% precision and recall that seemed to resist to any tentative to overpass by the present day systems [Mitkov, 2002]. Drawn mainly on strict matching criteria, in which the morphological features are of great value, these systems are unable to base their resolution strategies on more subtle strategies that would allow for: number mismatch, gender variation, semantic type variation, subset/superset references, element-of/has-as-element references, etc.

We have previously defined (Cristea&Dima, 2001), (Cristea et al., 2002) a framework (called **AR-engine**, AR – from Anaphora Resolution) able to accommodate different anaphora resolution models in terms of four components: a set of attributes, a set of external sources that fetch values of these attributes from the text, a set of rules to evoke antecedents and a set of heuristics defining the domain of referential accessibility.

This paper discusses, in terms of the framework, the feasibility of implementing models capable to tackle with complex cases of anaphora. Unlike usual approaches that mainly validate anaphora resolution systems on contiguous corpora, our research aimed at collecting hand-chosen examples from real texts, not solved by existent anaphora resolution approaches. We have classified these examples and empirically evaluated the computational feasibility of a solution within our framework. The necessity to upgrade the AR-engine in one important aspect was also revealed by this study.

The investigation has used a collection of two pages of Romanian examples as well as a collection of English examples, also reported in (Barbu *et al.*, 2002).

The research is motivated by the belief that interpretation of free language in modern applications, especially those related to semantic web, requires more and more sophisticated tools. We think that our investigation is a step forward towards dealing with really hard anaphora resolution problems.

2. The framework

In (Cristea&Dima 2001), (Cristea *et al.*, 2002) a framework incorporating a general AR engine and able to accommodate different AR models is proposed. This approach involves three layers: the **text layer** – populated with referential expressions (REs) –, the **projected layer** – where feature structures are filled-in with information from the text layer (in the following, projected structures – PSs) and the deep **semantic layer** – where discourse entities (DEs) are placed. We say that a PS is **projected** from an RE and that a DE is **proposed** or **evoked** by a PS (Figure 1).

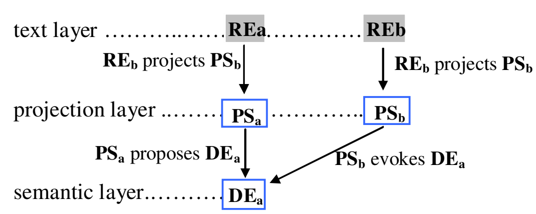

Figure 1. The tree-layer representation of two co-referring referential expressions

Any AR model is defined in terms of four components: **a set of primary attributes** that fill the PSs of the projection layer and are then reported to the semantic layer, a corresponding **set of knowledge sources** that fetch values from the text to these attributes, **a set of matching rules and heuristics** responsible to co-operate in order to decide whether an RE introduces a new DE and, if not, which of the existing DEs it refers and **a set of heuristics that configure the domain of referential accessibility** therefore establishing the order in which DEs have to be checked. The type of analysis supported by the framework is incremental, anaphoric references being processed from left to right (in left to right reading languages) miming the way humans read texts. Like in normal reading, anaphors are mostly resolved at the time of reading, but sometimes decisions are postponed until the acquisition of complementary information adds enough data to allow a disambiguation process. It is like when backwards eye movements reveal indecisions (Vonk, 1985). Processing presupposes four phases (times) at each resolution cycle which is triggered by an RE on the surface stream: a **projection phase**, a **proposing/evoking phase**, a **completion phase** and an optional **re-evaluation phase** (see section 4 for a detailed description).

The framework is language independent, in the sense that the adjustment to one language or another consists in defining a specific set of attributes, establishing the knowledge sources capable to fill them and devising evoking heuristics/rules specific to each language. We consider that the set of attributes may include morphological (lexical number, lexical gender and person), syntactical (the head of the RE, the syntactic role of the RE in the sentence, the immediate parent and daughters in a dependency description, the quality of being in a subject, or an apposition, or a predicative noun position, etc.), lexico-semantic (lemma, name – for proper nouns, natural gender, either the sense of the RE's head lemma as a synset index conforming to Wordnet (Miller *et al.*, 1993) if a sense disambiguation source is available or just all the senses, semantic roles, etc.), positional (the inclusion in a discourse unit), surface realisation (zero-pronoun, full-flagged pronouns, indefinite NP, definite NP, proper nouns, etc.). Identification of Wordnet senses opens the possibility for investigations over conceptual hierarchies. Also, features as animacy, natural gender and concreteness could be considered simplified semantic tags derived from a conceptual hierarchy. Out of semantic roles, selectional restrictions, inferential links, pragmatic limitations, semantic parallelism and object preference can be verified.

The set of rules include three type of rules:

- certifying rules (applied first), which, if evaluated to 'true' on a pair (PS, DE), certify without ambiguity the DE as an antecedent of the PS;
- demolishing rules (applied after the certifying rules), which rule out a possible DE as candidate of an PS (and its corresponding RE). These rules lead to a filtering phase that eliminates from among the candidates those discourse entities that cannot possibly be referred to by the RE under investigation;
- promoting/demoting rules (applied after the demolishing rules), which increase/decrease a resolution score associated with a pair (PS, DE). These rules allow for a subsequent selection phase, in which either the best candidate of a PS is chosen from the ones remaining after the demolishing rules have been applied, or a new entity is introduced.

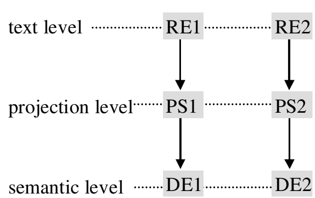

Figure 2: Notation conventions for objects assigned to the three layers

To facilitate the discussion to follow we will suppose the notation convention sketched in Figure 2, where RE1, RE2 are referential expressions, PS1, PS2 are projection structures and DE1, DE2 are discourse entities. In the text we will also mark REs by italic letters (as *a car*) while their corresponding DEs by a paraphrasing text in bold fonts and within square brackets (as [**the car**]).

3. Types of difficult anaphora

A set of difficult cases of anaphora resolution are discussed with the framework of the AR-engine. The intent is to find out if the knowledge sources and rules/ heuristics are computationally tractable, how much processing they require and of what kind. The discussion does not take into consideration the universal panacea for all the failures in AR, world knowledge (WK).

3.1. Relations triggered by positional (syntactic) constrains

Nested REs – usually genitival constructions

(1) *The University building*

(2) *Amenomfis the IVth's wife*

(3) *the face of the beautiful queen*

In constructions of these types, two DE's are involved and a certain relation holds between them. In (1) the two DEs are [**the University building**] and [**the University**], and [**the University building**] `belongs-to` [**the University**]. In (2) between [**Amenomfis the IVth's wife**] and [**Amenomfis the IVth**] a different type of `belongs-to` relation holds, perhaps a `commitment`, and the same is true in (3) where between [**the face of the beautiful queen**] and [**the beautiful queen**] a still different type of `belongs-to` relation holds, perhaps a `part-of` relation. In all cases, the possessed object (or the part) corresponds to the outer RE while the belonging entity (or the whole) corresponds to the inner RE on the surface string. In order to include such a relation on the semantic level it is likely that the belonging entity (corresponding to the inner RE) be processed before the possessed entity (the outer RE). If RE1 is nested on RE2 on the text level, a knowledge source should fetch the value `nesting(RE1)` to a `position` slot of the PS corresponding to RE2. This slot will later on be transformed on DE2 of the semantic level into a `belongs-to` attribute indicating the already existing DE1. Moreover, a demolishing rule should always prevent a co-reference relation between the two DE's.

Apposition

(4) *Nefertiti, Amenomfis the IVth's wife*

An apposition usually brings supplementary knowledge regarding a discourse entity. We will consider the two elements of the apposition as different REs, that should be processed into a co-reference anaphora.[1] Suppose RE2 is an apposition of RE1 on the text level. Relatively easy to detect by syntactic means, we will suppose a knowledge source capable to fetch a `position=apposition-of(RE1)` slot attached to

[1] Note that this is a simplifying assumption, because there are frequent cases when an apposition adds a metaphoric view to the antecedent DE.

PS2. As PS1 should have been processed into a DE1 the moment PS2 is being processed, a certifying rule should unify PS2 with DE1. As a result, DE1 will accumulate all the attributes of PS2.

The subject-predicative noun relation

(5) _Maria_ *is* _the best student of the whole class_.

(6) _Your rival_ *is* _a photo_.

(7) _The young lady_ *became* _a wife_.

Supposing RE1 is the subject and RE2 is the predicative noun, a knowledge source of a syntactic nature should be able to fetch a `position=predicative-noun(RE1)` attribute into the PS2 corresponding to the predicative noun RE, considered, in a simplifying assumption, co-referential with the subject. In a more elaborated approach, one could try to distinguish different types of relations in examples like (5) to (7). In (5), for instance, *the best student of the whole class* should be considered co-referential with the DE [**Maria**] and the resolution should aim at injecting into the DE [**Maria**] the information brought by the predicative noun RE. Suppose the DE [**Maria**] is something of the kind: [`name="Maria", sem=person1, Ngen=fem, num=sg`], where `person1` is the first sense of the word *person* according to Wordnet. Then, the fact that she is seen now also as a student must not affect any of the attributes `name`, `Ngen` (natural gender) or `num` (lexical number) but instead add into the description an attribute `lemma=student` (if only the head of the RE is considered in the representation, or a more sophisticated description if the constituents are also kept) and replace the `person1` value of the `sense` attribute with a more specific one: `student1`[2]. To correctly identify in example (6) that the rival (a human being) is not a photo (an inanimate) but the person faced in the photo necessitates deep WK. Example (7) reveals the `same-as` type of link that should chain discourse entities subject to time changes, as described in (Cristea&Dima, 2001). At the evoking time a certifying rule will unify the PS of the predicative noun with the DE corresponding to the subject. A complication arises when the resolution of RE1 (the subject) was postponed the moment RE2 (the predicative noun) is processed[3]. In this case, the unification should make PS2 to be co-referential with whatever will result as the evoked DE of the (for the moment) postponed PS1.

3.2. Lemma disagreement of common nouns

Common NPs displaying identical lexical number but different lemmas

(8) *Nefertiti, Amenomfis the IV*th*'s wife … the beautiful queen*

Unification in this case should mainly be similarity-based. In principle, a queen should be found more similar to a wife then to a pharaoh, supposing Amenomfis is known to be as such. If, instead, this elaborate knowledge is not known and all that is known about Amenomfis is his quality of being a man, the moment *the beautiful queen* is processed a queen should again be found more similar to a wife than to a man. Many approaches to measure similarity in NLP were proposed and many use Wordnet, for instance (Resnik, 1999). When a sense disambiguation procedure is lacking, than a Wordnet-driven similarity that counts the common hypernyms of all senses of the two lemmas could be a useful substitute in some cases[4].

Still, criteria to decide similarity are not elementary and a simple intersection of the hypernymic paths of the anaphor lemma and the candidate antecedent lemma often does not work. The following is an example of a chain of erroneous unifications based on this simplistic criteria: `the centre of the hall opposite the big telescreen | his place | some post so important and remote | the back of one's neck | a chair | places away | the end of the room | the protection of his foreign paymasters`[5].

Another useful criterion for the identification of co-referential common noun REs with different lemmas is the natural gender (queen and wife are both feminine in natural gender), but unfortunately this knowledge is often not available.

In other cases the antecedent could be recuperated by looking at the constituents of the head nouns. Consider example (9):

(9) *the most beautiful women… those beauties*

A scored rule should be able to confront the lemma *beauty* with constituents of the head *women* in the DE for [**the most beautiful women**].

Common NPs with different lexical number and different lemmas

(10) *A patrol … the soldiers*

(11) *The government… the ministers*

In two out of three senses a patrol is a group and in one sense out of 4 government is also a group. This suggests to look for this value of the `sense` feature in the hypernymic paths of the lemmas of the candidate antecedents of plural NPs as the ones in the examples (10) and (11). However this criterion is too weak because a lot of words have senses that correspond to groups (*a garden*, for instance, has a sense that means a group of flowers and in a text like *A patrol stopped by the garden. The soldiers…* there is high chance to find *the soldiers* co-referring [**the garden**] instead of [**the patrol**]. Different criteria should be combined to maximize the degree of confidence, among which a similarity criteria (for instance based on Wordnet definitions), or even the simple identification of the antecedents within a fixed collection

[2] The implicit assumption here was that word sense disambiguation capabilities were used as a knowledge source.

[3] The same is true for the apposition.

[4] There is good reason to believe that such an approach is successful when lexical ontologies, fine graded in word senses as Wordnet, are used, because most of the word senses that are very close to each other must have common hypernyms from a certain level up. Senses displaying common ancestors must be more similar than the ones whose hierarchical paths do not intersect.

[5] From G. Orwell, "1984".

of collective nouns, as suggested in (Barbu *et al.*, 2002). In principle, this case is similar to the preceding one if a constraint of being a group is added to the antecedent.

Seen from a different angle, this type of co-reference is imperfect, because one could argue that *the soldiers* are not co-referential with [**the patrol**], as they are only members of it, the same as ministers are members of [**the government**]. This is still a nuance difference which makes little sense on the semantic layer, unless it is desirable to see [**the patrol**] and [**the government**] as organizations and [**the soldiers**] and [**the ministers**] as their members. Regarding the computation within the framework, the affair must be tackled at the level of the scoring rules who should be fine-tuned to give a low matching score if the RE *the soldiers* is intended to introduce a new DE, different from the [**the patrol**] DE, and a high score if a co-reference relation is hunted for. In the first case a special method should be made responsible for filling-in an attribute `members-of` pointing to [**the patrol**] in the newly created DE [**the soldiers**] (as described in section 4).

Common nouns referring proper nouns

(12) *Bucharest… the capital*

There are no other means to solve this reference other than knowing that Bucharest is a capital and this WK should be acquired the moment Bucharest is processed and should be fixed in its DE, for instance under the `sense` attribute. If such a solution is available, then the framework processes the reference the same way it does with common nouns with different lemmas.

3.3. Number disagreement

A plural pronoun identifying a conjunction/disjunction of singular/plural NPs or a split antecedent

(13) *John waited for Maria. They went for a pizza.*

There is no reason to consider that during the interpretation of (13) above a discourse entity must have been proposed for the group [**John,Maria**] as soon as the referential expression *Maria* was parsed. Or else, we would have to face a very uncomfortable indecision regarding what groups to consider and when. This group is called for as a DE only because at a certain moment, as the text unfolds, a referent appeared: *they*. In (14) below there is no need for such a group, in the sense that the reader is perhaps not even conscious that such a group exists:

(14) *John waited for Maria. He invited her for a pizza.*

Neither vicinity in the location space of the story, nor textual vicinity are sufficient criteria to propose groups on the semantic layer, as examples (15) and (16) can prove:

(15) *John was in New York when Maria wrote him that she finally made up her mind. They got married the next month.*

(16) *John finished his classes. He went to a football match. As it was a rainy day, no more than 10 people were in the stadium. Maria happened to be there too. They went for a pizza and one month later got married.*

To make life even harder, note that in (16) 12 people are candidates for different groups of persons ([**John,10pers**], [**10pers,Maria**], [**John,10pers,Maria**], [**John,Maria**]). Nevertheless, the reader has no difficulty to identify *they* with only the group [**John,Maria**]. The second question is why not to attach to the group also [**John's classes**], [**the football match**], [**the rainy day**] or [**the stadium**]? Of course because none of the others can go for a pizza. And also because getting married is an occupation for exactly two people. But this is WK – therefore difficult to prove.

From the discussion above two ideas emerge: that groups should have a property of similarity of their elements and that group formation is triggered by a first referent to it. A group, unless it is verbalised as such in the text, does not exist until it is referred to. Still, two questions remain: how much we can do in the absence of WK for the group content identification and what the criteria to trigger the creation of group DEs are, therefore by what means a referent to a group can be identified. The answer to the first question is that we should again use some similarity measures to identify members of groups in the preceding text. As for the second question, the framework polity is to propose new DEs when no match between the current PS and the preceding DEs raises above a $threshold_{min}$. This policy is good enough for our purpose as long as no plural DEs are in the recent proximity, toward which the plural anaphor could match. If an ambiguity arises, then the second framework policy to postpone resolution until sufficient discrimination criteria lets a unique candidate within a $threshold_{diff}$ range is well suited again. The combination of these two policies in example (17) below, for instance, would maintain the indecision whether *they* should co-refer [**John,Maria**] or [**the classes**] as long as no WK is available to state that only people can go for a pizza, and this is a correct behaviour.

(17) *John waited for Maria when the classes were over. They went for a pizza.*

A plural noun identifying a conjunction/disjunction of singular/plural NPs or a split antecedent

Supplementary to the problems identified above, when the anaphor is a noun, the similarity criteria found to characterize the group should extend to the anaphor as well. Consider example (18):

(18) *Horia, Closca and Crisan … the heroes*

The similarity criteria sketched above yields `person, individual, someone, somebody, mortal, human, soul -- (a human being)` as the concept characteristic to the discovered group, while in two out of four senses the word *hero* means also a `person`. As such, there is enough evidence to conclude that the DE [**the heroes**] should point to the DEs [**Horia**], [**Closca**] and [**Crisan**] as members. As it will be seen in section 4, the need to tackle with cases similar to this one led to an upgrade of the framework, as it revealed the necessity to have methods responsible for decorating existing DEs with new attributes, in our case to complete the [**the heroes**] DE with an attribute `group-of = <x, y, z>`, with `x`, `y`, `z` being identifiers of the DEs [**Horia**], [**Closca**] and [**Crisan**]. More generally, such rules could do other actions related to the completion of a DE as a result of a successful PS-DE match (as, for instance, the way values of similar attributes should be combined).

3.4. Bridge anaphora

Elements to set references

(19) *all the weapons for the underwater hunting… the masque… the rifle… the ribbon paws*

In this example, to each of the REs *the masque*, *the rifle*, and *the ribbon paws* must correspond a proper DE. Moreover, each of them must have an attribute `element-of` pointing to the DE [**the weapons for the underwater hunting**]. *The rifle* against [**the weapons…**] is the only `element-of-set` relation that can be easily inferred based on a similarity computation. Also looking at the constituents of the candidate DE *the rifle paws* could be found somehow similar with underwater hunting. A masque is not in itself a weapon, although the context obliges to this inclusion. Only reasoning on deep WK would allow for such an assignment. But even if such WK would be available it still remains the problem of assigning the `element-of` links from all these DEs towards the [**the weapons…**] DE. The situation is similar to all cases when the values of a new attribute were set by a method during the completion phase (see the next section).

Hidden discourse entities

(20) *When I got into the room I saw a strange screen saver on the big monitor. The other computer was off.*

Interesting debates could arise around this example. Any human person reading this text is aware of the existence of two computers in the mentioned room: one with a big screen, on which a strange screen saver was running, and another one which was off. One question is whether both computers should be represented on the semantic layer or only [**the other computer**]. Since the mentioning of *the other computer* doesn't make sense but if [**some (first) computer**] exists, this can be taken as an implicit mentioning of the first computer. However there is no RE in the text explicitly referring this DE, excepting from the big screen which is part of this computer. But we see no reason why to consider only the `part-of` relation and to neglect others like `made-of`, spatial relations, etc. There is no end to describe all objects to which a certain mentioned object could consciously interact. For instance in the reader's mind at least the image of one table on which one or both computers lay is present. We think a saver solution (at this level of automatic reasoning which is insinuated by our framework) is to consider as candidates for being represented on the semantic layer strictly those objects which are explicitly mentioned in the text. If a more elaborated resolution model is to be attached on top of the work performed by the AR-engine, then those hidden DEs should be put into evidence through an inference mechanism which is beyond our interest at this level of processing.

What the engine would have to do in the case of example (20) is to build a DE corresponding to the RE *the big monitor* and another DE for the RE *the other computer*. No relations will have to link these representations. On the contrary, in a sequence like this:

(21) *When I got into the room I saw a strange screen saver on the big monitor. The computer was left open by my colleague.*

The DE [**the computer**] should display a `has-as-part` relation towards the DE [**the big monitor**]. In this example a method attached to an attribute `has-as-part` must be responsible for filling-in the value. The difference between examples (20) and (21) is that in (20) the method should prevent from retaining as the value of the attribute `has-as-part` of the DE [**the other computer**] the identifier of the DE [**the big monitor**], while in (21) is should mainly go for it.

4. Upgrading the framework

The main stream of processing in AR-engine is driven by the left-to-right scanning of the raw text in search for REs to be processed. At any moment during processing just one RE is under investigation, the current RE, and one processing cycle, as already mentioned, follows three compulsory phases and an optional one. The first (mandatory) phase, the **projection phase**, builds a PS on the projection layer out of the information, centered on the current RE, obtained from the text layer with the contribution of the knowledge sources available. The second (mandatory) phase, the **proposing/evoking phase**, is responsible of matching the PS towards one DE, either by proposing a new one or by deciding on the best candidate from the existent ones. This process involves first the running of certifying rules (if available), followed by the demolishing rules (if available), finally followed by the scoring rules. In the end either an existent DE was firmly identified by a certifying rule or matching scores between the current PS and a class of antecedent DEs were computed. Based on these scores, three possibilities can be judged: 1. all candidate DEs scored under $threshold_{min}$: a new DE is build; 2. the best rated scores are above $threshold_{min}$ but in the $threshold_{diff}$ range more than one candidate is placed: there is not enough evidence for just one antecedent among the best placed candidates, which makes that a postponement of the decision is decided in the hope that successive matches could give supplementary clues for the resolution of the current RE; 3. just one score is placed in the $threshold_{diff}$ range: the best candidate strongly identifies itself from the remaining candidates and is therefore chosen as the antecedent. In the third compulsory phase, the **completion phase**, a resolved PS triggers the combination of its data with the data contained in the antecedent DE if identified, or, simply, the projection of its data onto the new build DE, followed by the deleting of the resolved PS from the projection layer. Only bi-directional links between the RE and the identified DE are kept. Finally, the optional **re-evaluation phase**, is triggered if postponed PSs remained on the projection layer, with the intent to apply matching rules again on all of them.

When referential relations, different than strict co-referentiality, are intended to be revealed, attributes in DEs that are not triggered from the corresponding PSs appear as necessary. The framework must abide a transformation which, actually, also naturally generalizes the engine's behavior in two known cases: the filling-in of the attributes in the newly created DE during the proposing phase and the combination of attributes following an evoking phase. The default behavior of the engine in the completion phase was a copying action from the PS into the new DE, in the case of a proposing phase and an appending action of the values of corresponding attributes from the PS into the evoked DE, in the case of an evoking phase. None of these actions are, nevertheless, useful if a `belongs-to` (as in section 3.1 – nested REs) or a `members-of` (as in section 3.2 – common NPs with different lexical numbers and different lemmas) attribute is to be filled-in. A section dedicated to actions to be performed for the filling-in of specific attributes during the completion phase should be opened in the third component of the framework – that dedicated to rules and heuristics. Each such method should specify: the name of the attribute in focus, the signature of its value, conditions of application of the method, and the procedure of completion of its value.

5. Conclusions

Modern applications, especially those related to semantic web, compel for combined and complex methods in NLP. On these application environments more and more sophisticated tools will be put to work and AR methods should be prepared to tackle also hard problems raised by the interpretation of free language.

The paper investigates cases of difficult AR problems and proposes a set of solutions within the framework of a general AR solver, called AR-engine, previously introduced elsewhere (Cristea&Dima, 2001), (Cristea *et al.*, 2002) and shortly presented in section 2. The investigation oriented towards examples that failed to be solved in a previous evaluation tentative (Cristea *et al.*, 2002), on examples chosen by hand from free texts, as well as on examples reported by other authors to be difficult to tackle (Mitkov, 2001), (Barbu *et al.* 2002). In section 3 we have classified the extracted examples into four categories, that we believe display an ascending degree of difficulty: co-reference relations whose resolution could be triggered by positional (syntactic) constrains, co-reference relations in which the anaphor and the antecedent are common nouns with disagreement in lemma, noun and pronoun anaphors displaying number disagreement with the antecedents, and bridge anaphorae. For the first time, anaphoric references other than net co-references were experimented with the AR-engine. Crossing the barrier of one to one co-references revealed also the necessity to upgrade the framework in a number of aspects, described in section 4.

The languages under investigation were English and Romanian. In this stage of the research the interest focused, on one hand, in enhancing the engine and, on the other, in devising rules and heuristics that integrated into a model to give the expected solutions to the problems at hand. Due to the difficulty to annotate large fragments of texts we do not have yet a value of the precision and recall of the proposed solution. The implementation succeeded to solve most of the presented examples that did not necessitate world knowledge, in the sense that the expected antecedents were recuperated in short texts with rather low density REs. Still, it is difficult to appreciate at this moment what would be the behavior of the model when more referential expressions would conquer for solutions.

The truth that "when everything is done AR comes for free" seems to be proven again.

Acknowledgments. The authors are grateful to Cătălina Barbu for providing her set of plural disagreement examples.

6. References

Barbu, C., Evans, R., Mitkov, R. (2002). A corpus based investigation of morphological disagreement in anaphoric relations. In *Proceedings of Language Resources and Evaluation Conference - LREC 2002*, (pp. 1995—1999), vol. VI, Las Palmas, Spain.

Cristea, D. and Dima, G.E. (2001). An Integrating Framework for Anaphora Resolution. In *Information Science and Technology*, Romanian Academy Publishing House, Bucharest, vol. 4, no. 3.

Cristea, D., Postolache, O., Dima, E.G., Barbu, C. (2002). AR-Engine – a framework for unrestricted co-reference resolution. In *Proceedings of Language Resources and Evaluation Conference - LREC 2002*, vol. VI, Las Palmas, Spain.

Miller, G., Beckwith, R., Felbaum, C., Gross, D., and Miller, K. (1993). Introduction to wordnet: An on-line lexical database. Tech. rep., University of Princeton.

Mitkov, R. (2002) Anaphor resolution. Longman. Studies in Language and Linguistics.

Resnik, P. (1999) "Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language", vol. 11.

Vonk, W. (1985). The immediacy of inferences in the understanding of pronouns. In G. Rickheit and H. Stronher (eds.), *Inferences in text processing*.