Resolution of Anaphoric Reference

Information Status, Salience/Activation-based Reference Resolution

Vašek Němčík

January 15, 2003

The Main Idea

- building a hierarchy of objects already mentioned in the discourse
- this hierarchy should reflect the activation (salience) of each object in the hearer's/reader's memory
- the structure should be built inductively, step by step, based on its previous state, so that the activation of individual objects can be studied at any point of the processing
- the rank (more generally said, position) of each object in the structure corresponds to its likelihood of being referred to by an anaphor

Centering – the starting point

Text is a sequence of utterances with:

- **list of forward-looking centers** – individual objects mentioned in the utterance (ordered according to their position)
- **unique backward-looking center** – one of the previous forward-looking centers – the "center" (topic) of the utterance
- **center transition type** – the types ordered according to their coherence smoothness

Anaphors are resolved according to the ranking of forward-looking centers and/or transition smoothness.

Strube and his centering-inspired approach

- **S-list** – list of discourse objects structured according to their information status (reflecting their salience) in the text
- S-List element (information status) types:
  - OLD (hearer-old)
    * E (evoked) – already mentioned object
    * U (unused) – proper names and titles
  - MED (mediated)
    * I – inferrable
    * I^C – containing inferrable
    * BN^A – anchored brand-new
  - NEW (hearer-new)
    * BN (brand-new) – object mentioned for the first time

Information status ranking

Having utterances U_{i-1} and U_i, the discourse entities on the S-list can be ordered according to preference as follows:

$$OLD_i > OLD_{i-1} > MED_i > MED_{i-1} > NEW_i > NEW_{i-1}$$

(within individual groups only position preference applies)

The Resolution Algorithm

- The elements of the (ordered) S-list are tested until the test succeeds (agreement and binding aspects are considered).
- The S-list is updated according to the information status ranking (if it is identical for two elements, the linear position is the decisive factor).
- If the analysis of an utterance is finished, all entites from the S-list, which are not realized in it, are removed.
- Any occurrences of pleonastic "it" and quoted elements are ignored.

Example

<table>
<tr><td colspan="2">Brennan drives an Alpha Romeo.
S: [Brennan$_U$, Alpha Romeo$_{BN}$]</td></tr>
<tr><td colspan="2">She drives too fast.
S: [Brennan$_E$: she]</td></tr>
<tr><td>Friedman
S: [Friedman$_U$, Brennan$_E$: she]
races her on weekends.
S: [Friedman$_U$, Brennan$_E$: she]</td><td>A professional driver
S: [Brennan$_E$, Driver$_{BN}$]
races her on weekends.
S: [Brennan$_E$, Driver$_{BN}$]</td></tr>
<tr><td>She drives to Laguna Seca.
S: [Friedman$_E$: she, Lag. Seca$_U$]</td><td>She drives to Laguna Seca.
S: [Brennan$_E$: she, Lag. Seca$_U$]</td></tr>
<tr><td>She
S: [Friedman$_E$: she, Brennan$_E$: her]
often beats her.
S: [Friedman$_E$: she, Brennan$_E$: her]</td><td>She
S: [Brennan$_E$: she, Driver$_{BN}$]
often beats her.
S: [Brennan$_E$: she, Driver$_E$: her]</td></tr>
</table>

Remarks

- backward-looking centers and transition ranking are made redundant (all the information is already included in the S-list)
- only agreement and binding criteria used (no real-world knowledge)

Results

The approach was compared to the original centering approach (Brennan et al.) and its modification by Kameyama.

Presented approach showed better performance. Despite Kameyama's clause-level-based approach is more fine-grained, it's fine-grainedness is also a source of errors spoiling its performance.

The backbone of the algorithm is the preference for hearer-old entities. Whenever it is misleading, the algorithm gives wrong results. On the other hand, in this case, the other two approaches give wrong results aswell.

Similar approach – RAP (Lappin & Leass)

- designed to work with McCord's Slot Grammar parser
- accounts also for intrasentential and intersentential anaphora
- like the preceeding approaches, it doesn't use any semantic or real-world knowledge

RAP's components

- intrasentential syntactic filter
- morphological filter (agreement issues)
- pleonastic pronouns identification procedure
- salience parameters assignment procedure
  (assigns each NP its salience weight)
- procedure for identifying anaphorical identity equivalence classes
- decision procedure (for selecting the preferred element)

Some definitions:

P is in the *argument domain of* N $\Leftrightarrow$ P and N are arguments of the same head

P is in the *adjunct domain of* N $\Leftrightarrow$ N is an argument of head H, P is the object of preposition PREP and PREP is an adjunct of H.

P is in the *NP domain of* N $\Leftrightarrow$ N is a determiner of a noun Q and either P is an argument of Q or P is the object of a preposition PREP and PREP is an adjunct of Q.

P is *contained in* a phrase Q $\Leftrightarrow$ P is either an adjunct or an argument of Q

The syntactic filter

Pronoun P is non-coreferential with a noun phrase N if any of these conditions is met:

- P and N have incompatible agreement features.
  $*$*The woman*$_i$ *said that he*$_i$ *is funny.*
- P is in the argument domain of N.
  $*$*She*$_i$ *likes her*$_i$.
  John$_i$ *seems to want to see him*$_i$.
- P is in the adjunct domain of N.
  $*$*She*$_i$ *sat near her.*

The syntactic filter (cont.)

- P is an argument of a head H, N is not a pronoun, and N is contained in H.
  **He$_i$ believes that the man$_i$ is amusing.*
  **This is the man$_i$ he$_i$ said John$_i$ wrote about.*

- P is in the NP domain of N.
  **John$_i$'s portrait of him$_i$ is interesting.*

- P is a determiner of a noun Q, and N is contained in Q.
  **His$_i$ portrait of John$_i$ is interesting.*
  **His$_i$ description of the portrait by John$_i$ is interesting.*

Pleonastic pronouns test

The class of ModAdj modal adjectives:

necessary, good, economical, possible, useful, easy, certain, advisable, desirable, likely, convenient, difficult, important, sufficient, legal

The class CogV of cognitive verbs:

recommend, think, believe, know, anticipate, assume, expect

Pleonastic uses of it:

> It is ModAdj that **S**.
> It is ModAdj (for **NP**) to **VP**.
> It is CogV-ed that **S**.
> It seems/appears/means/follows (that) **S**.
> **NP** makes/finds if ModAdj (for **NP**) to **VP**
> It is time to **VP**.
> It is thanks to **NP** that **S**.

The Anaphor Binding Algorithm

The higher argument slot relation:

$$subj > agent > obj > (iobj \,|\, pobj)$$

An NP N is a possible antecedent for a lexical anaphor A

$$\Longleftrightarrow$$

N and A do not have incompatible agreement features and one of the following conditions holds:

- A is in the argument domain of N and N fills a higher argument slot than A.
  They$_i$ *wanted to see themselves*$_i$.
  Mary knows the people$_i$ *who John introduced to each other*$_i$.

The Anaphor Binding Algorithm (cont.)

- A is in the adjunct domain of N.
  He$_i$ *worked by himself*$_i$*.*
  Which friends$_i$ *plan to travel with each other*$_i$*?*

- A is in the NP domain of N.
  John liks Bill$_i$*'s portrait of himself*$_i$*.*

- N is an argument of a verb V, there is an NP Q in the argument or adjunct domain of N such that Q has no noun determiner and either A is an argument of Q, or A is an argument of PREP and PREP is an adjunct of Q.
  They$_i$ *told stories about themselves*$_i$*.*

The Anaphor Binding Algorithm (cont.)

- A is a determiner of a noun Q and either Q is in the argument domain of N and N fills a higher argument slot than Q, or Q is in the adjunct domain of N.

Salience weighting

- Salience of individual elements is determined using *salience factors*.
- Each salience factor has a weight corresponding to its contribution to the total salience of the considered element.
- The salience of each element is the sum of the weights of all salience factors, which have this element in their scope.
- The salience factors get degraded each time processing of a new sentence has begun (when it reaches 0, it is removed).

Some salience factor types and their initial weights

Factor type	Initial weight
Sentence recency	100
Subject emphasis	80
Existential emphasis	70
Accusative emphasis	50
Indirect object and oblique complement emphasis	40
Head noun emphasis	80
Non-adverbial emphasis	50

Description of selected salience factors

- *Sentence recency* – one of the most important factors. Gives preference to elements in the current sentence.
- *Existential emphasis* – predicate nominal in existential constructions
- *Head noun emphasis* – increases salience of NPs that are not embedded in another NP
- *Non-adverbial emphasis* – increases salience of NPs that are not contained in an adverbial PP

Equivalence classes

Let us view anaphorical reference as a mathematical relation between the referring expression (`y`) and the expression being referred to (`x`), and denote it `antecedes(x,y)` .

The relation `coref` of co-reference between elements is then a reflexive, symetric and transitive closure of the `antecedes` relation.

The relation `coref` defines classes of discourse referents. For each discourse element `u` let us define

$$\texttt{equiv(u)} = \{ \texttt{y} \mid \texttt{coref(u,y)} \}$$

as the class of terms referentially equivalent with `u`.

Each such class has a salience weight, which is the sum of salience factors in scope of which at least one element of the equivalence class lies.

The Resolution Procedure (pronouns)

- Create list of all NPs in the sentence and identify their type.
- With the found NPs:
  - Categorization – non-referential / evoking new discourse referents / evoking already mentioned referents
  - Apply salience factors to appropriate discourse referents
  - Apply syntactic filter (for ruling out syntactically inappropriate antecedents for personal/possessive pronouns & reflexive binding algorithm (pairing lexical anaphors with their possible antecedents).
- If any non-pleonastic pronouns remain in the sentence, proceed with their resolution as follows (lexical anaphors are handled by the anaphor binding algorithm):

Resolution procedure for third person pronouns

- A list of possible antecedent candidates (the last elements of each equivalence class) with their salience weights is created. Slight local (individual to each pronoun) changes in the weighting are applied (cataphora penalty, parallelism reward, etc.).
- A salience threshold is applied.
- Agreement features are determined and separately for sg and pl candidates these steps are carried out:
  - if there are no candidates, skip this step
  - apply the morphological filter
  - apply the syntactic filter (excluding referential pairs made inappropriate during the preceding stages).
  - the remaining candidate (if more of them, then the most salient one) is considered as the best one.

Resolution procedure (cont.)

- Having the best sg and pl candidates:
  - If one of the candidates is missing, the only one left is the antecedent.
  - If both candidates present, choose the one with higher salience weight.
- The mentioned candidate is presented as the antecedent of the anaphor being processed.

Remarks:

- Morphological and syntactic filters applied to reduce the number of candidates.
- NP salience measures (syntax-based)
- No real-world knowledge or semantic constraints are used.
- Proximity of the NP to the pronoun plays the decisive role in case of equal salience weights.
- Strong preference of anaphora to cataphora
- Good possibility too fine-tune the algorithm by re-weighting or introducing new salience factors.

RAP's weaknesses

- Various cases of intersentential anaphora:
  This green indicator is lit when the controller is on.
  It shows that the DC power supply voltages are at the correct levels.
  'indicator' and 'controller' are ranked equally, being both subjects – the proximity preference applied. However, not always does it lead to correct results.

- Instances with too low-weighted antecedents in structurally non-prominent positions

- Cases where semantic or pragmatic information is needed to resolve the antecedent.

RAPSTAT – a statistical extension of RAP

Main aim of the statistical unit in STATRAP is to overturn salience-based decision in cases where salience difference between to candidates is small while one of them has a considerable statistical preference.

According to performance tests, RAPSTAT demonstrates a slightly better performance than RAP.

The Prague Group's Theory

- The notion of activation and salience is almost the same as in the previous cases, but this theory is based on slightly different formal grounds.
- The Prague studies of salience and discourse structure is a part of a more complex system – Functional Generative Description (FGD) of Language.
- The heart of the approach is the study of discourse – as of a playing field for the speaker and hearer. The speaker is trying to change the knowledge of the hearer, based on the knowledge they share.

Basic concepts

The very central concept of the whole theory is the concept of TFA – Topic-Focus Articulation.

According to this theory, each sentence is divided into two parts:

- **The Topic (T)** – what the sentence is about
- **The Focus (F)** – what is being said about the topic

To give yet a different definition of the same thing, one can regard the topic as *contextually bound (CB)* (already mentioned/known) elements and the focus as *contextually unbound (CN)* (new, unmentioned) elements.

Basic concepts

This articulation can also be trivial – there needn't be a topic in a sentence. However, there has to be a focus (sentence without it would have no information value anyway).

Yet a similar view of the same thing is the hierarchy of *communicative dynamism* which describes the level of activation each elements gets based on its (syntactical and linear) position in the sentence. Usually, there is a whole range of various levels of dynamism specified (depending on the language etc.), but often articulation into only these three basic levels suffice:

- **Theme** – the least dynamic part of the sentence
- **Rheme** – the most dynamic part of the sentence
- **Transition**

The topic and focus & activation

- The highest degree of activation is assigned to the elements referred to in the previous sentence – in particular in its focus.
- The second highest degree of activation is possessed by the topic of the previous center. Although the topic is not as activated as the focus, it tends to stay activated a bit longer than the focus.
- The next lower degree of activation is assigned to items mentioned in any of the preceding sentences or items that are somehow associated with them.
- A pronominal reference seems to "strenghten" the activation of the item a bit less than reference with a full NP.
- Items not referred to for a time lose their activation and after a time become "unreachable" for anaphorical reference.

Ways of expressing TFA in a sentence

- a highly language specific issue
- In languages with the so-called "free word order" is the main tool for expressing TFA. Thus, such a word order can't actually be called free (since it has semantic consequences).
- In languages with fixed word order is the main instrument passivization and topicalization.
- A common way in both of these groups of languages is the sentential stress. The constituents being a part of the topic get stressed (without a need to change the word order).

TFA and its relevance

At this point, one can think TFA is more important in such fields like stylistics rather than in theoretical linguistics.

Surprisingly enough, this is far from the truth.

The fact is that two sentences differing only in their TFA can have different semantics:

Many people read FEW BOOKS.
Few books are read by MANY PEOPLE.
MANY PEOPLE read few books.

The first speaks about many people and that each of them reads few books. In contrast to this, the second (and third) sentence speaks about a distinct small number of books which are read by many people.

Sentence representation

As an element, which is semantically relevant, TFA should be a rightful part of the semantic representation of a sentence.

In the FGD, it belongs to the representation of the tectogramatical (semantic) level of a sentence. This representation is a dependency tree (with the verb as a root), whose nodes carry various information about indvidiual lexical elements they represent.

Following from that, it is incorrect to speak about two word orders (TFAs) of a sentence – to be correct in terminology, one has to speak about two distinct sentences differing in their TFA.

Word order

As Vilém Mathesius, the most important researcher for Czech in this field, stated as early as in the 30's, there are two main types of word order:

- `unmarked/objective/neutral` – theme precedes the rheme
- `marked/subjective` – rheme (or its part) comes before the topic (such positioned rheme should be stressed)

Distinction between these two types (which is needed to determine the TFA), especially in written language, is a somewhat grey area.

$\Longrightarrow$ we need reliable criteria to identify the TFA

Criteria for determining the TFA

- **The Question Test** – we are trying to make up a question, answer of which can be the examined sentence

  If we find such a question, it can be generally said that the items mentioned in the question belong to the topic and the items "asked for" constitute the focus.

  The capital of Italy is ROME.
  WHAT is the capital of Italy?

- **Systemic ordering (SO)** – ordering of complement/dependency types for which holds:

  If A precedes B in the systemic ordering, A precedes B in a sentence with unmarked word-order.

Systemic ordering for Czech

Systemic ordering is highly language-dependent (though several precedence claims seem to be valid in most of the languages).

It allows us to distinguish, whether the current sentence has a marked or unmarked word order, and in case its word order is marked, it helps us to find the focussed element.

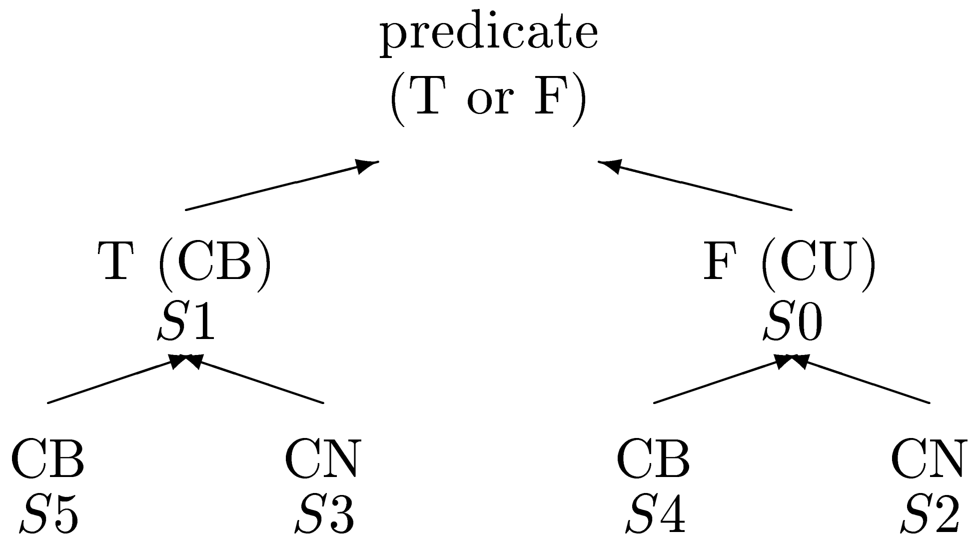
It has been suggested as a theory based on empirical studies – regularities in corpus text.

For Czech, SO has the following form:

1	*general dependency*	13	*through which place*
2	*agent*	14	*addressee (indirect object)*
3	*time*	15	*origin (or originator)*
4	*starting time*	16	*"from"-direction*
5	*finish time*	17	*patient*
6	*frequency*	18	*direction*
7	*duration*	19	*result*
8	*location*	20	*condition*
9	*manner*	21	*purpose*
10	*quantity*	22	*cause*
11	*benchmark*	23	*partitive*
12	*instrument*	24	*possessve*

Model of communicative dynamism hierarchy

Having a tectogramatical representation of a sentence, let's divide the individual nodes into groups S0 through S5 according to this schema:



Model of CD hierarchy (cont.)

- $S0$ and $S1$ are nodes directly dependent on the predicate (nodes of depth 1).
- Other groups consist of nodes dependent (directly or indirectly) on elements of the former groups.
- The groups with an odd number (higher than 1) are daughters of nodes in $S1$, the ones with an even number, are daughters of $S0$.
- The groups $S4$ and $S5$ are CB, $S3$ and $S4$ are CN.

Model of CD hierarchy (cont.)

Let's define these sentence positions (this definition holds for Czech and may be language-dependent):

$$P(0) = S0$$

$$P(1) = S1 \cup S4 \cup S5$$

$$P(2) = S2$$

$$P(3) = S3$$

Members of the same position set are equally activated in the SKK (Stock of Shared Knowledge), because they occur in equally prominent positions in the sentence.

Model of CD hierarchy (cont.)

Let T be a text consisting of sequence of n sentences $(V_1, \ldots, V_n)$.

Let O be the set of all discourse objects.

Then let's define a total binary function

$$B: \{1, \ldots, n\} \times O \longrightarrow \mathbb{N}_0 \cup \{\perp\},$$

which assigns each i and object t a number (or "undef" vlaue $\perp$), which represents:

- the index of the sentence the object t was last mentioned
- the position t's last occurrence had in a sentence

Model of CD hierarchy (cont.)

The value of $B(i,t)$ for each $i \in \{1,\ldots,n\}$ and $t \in O$ can be calculated using these recursive formulas: (with the use of values $B(i_0,t)$ for $i_0 < i$):

- If t is not mentioned in V_i and $B(i-1,t) = \perp$, then $B(i,t) = \perp$
- If t is not mentioned in V_i and $B(i-1,t) \neq \perp$, then $B(i,t) = B(i-1,t) + 4$
- If t is mentioned in V_i at position $P(j)$, then $B(i,t) = j$

Model of CD hierarchy (cont.)

The value of $B(i, t)$ can be used to:

- get the index of the sentence and the position, where the object t has been mentioned for the last time.
  This can be done using the remainder theorem – there exist p, j such that $B(i, t) = 4p + j$ and $p \in \mathbb{N}_0$ a $0 \leq j < 4$. Then, clearly, t has been last referred to in the sentence V_{i-p} at position $P(j)$.
- get the dynamism (salience, activation) of the object t after uttering the sentence V_i

Impact of the obtained information

With the knowledge of $B(i,t)$ for individual values of t, we can make the following conclusions for extrasentential anaphora occurring in the sentence V_{i+1}:

- if $B(i,t) \leq 2$, $B(i,t) = 4p$ or $B(i,t) = 4p+1$ pro $p > 0$, it IS possible to refer to t using weak anaphorical means (null, or unstressed forms).
- if $B(i,t) = 3$ or $B(i,t) = 4p+2$ for $p > 0$, t can NOT be an antecedent of an anaphor in V_{i+1}
- if $B(i,t) = 0$, stronger referential means are preferred (stressed personal, demonstrative or relative pronouns etc.).
- ellipsis requires an object, for which holds $B(i,j) \in \{0,1\}$

Evaluation & weaknesses

- the performance was relatively high – roughly 80%; but the approach would probably have to be extended considerably to improve
- the main weakness, as with other approaches, remain to be semantic influences
- the other remarkable weakness are intersentential anaphora – primarily processing of embedded clauses

Open issues and problems

- The discourse structure. It hasn't been formalized appropriately yet. In particular, the questions of topic locality stays a bit unclear. Should a topic be linked to a discourse segment or to the whole text instead of a single sentence?
- Semantic issues and associations still remain to be an open question and pose a giant challenge for psycholingustics and other empirical research.

References

[1] Eva Hajičová. *Focussing – a meeting point of linguistics and artificial intelligence.* B.V. North-Holland, the Netherlands, 1987.

[2] Eva Hajičová. *Issues of sentence structure and discourse patterns, volume 2 of Theoretical and computational linguistics.* Charles University, Prague, 1993.

[3] Eva Hajičová, Tomáš Hoskovec, and Petr Sgall. *Discourse modelling based on hierarchy of salience.* The Prague Bulletin of Mathematical Linguistics 64, Praha, 1995.

[4] Shalom Lappin and Herbert J. Leass. *An algorithm for pronominal anaphora resolution.* Computatinal Linguistics, 20(4), 1994.

[5] Petr Sgall, Eva Buráňová, and Eva Hajičová. *Aktuální členění věty v češtině.* Academia, Praha, 1980.

[6] Michael Strube. *Never look back: An alternative to centering.* In Proc. of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics, Montreal, Quebec, Canada, 1998.