

Resolution of Referents Groupings in Practical Dialogues

Alexandre Denis, Guillaume Pitel, Matthieu Quignard

LORIA

BP239 F-54206 Vandoeuvre-lès-nancy, France

denis@loria.fr, pitel@loria.fr, quignard@loria.fr

Abstract

This paper presents an extension to the Reference Domain Theory (Salmon-Alt, 2001) in order to solve plural references. While this theory doesn't take plural reference into account in its original form, this paper shows how several entities can be grouped together by building a new domain and how they can be accessed later on. We introduce the notion of super-domain, representing the access structure to all the plural referents of a given type.

1 Introduction

In the course of a discourse or a dialogue, referents introduced separately could be referenced with a single plural expression (pronoun, demonstratives, etc.). The grouping of these referents may depend on many factors: it may be explicit if they were syntactically coordinated or juxtaposed or implicit if they just share common semantic features (Eschenbach *et al.*, 1989). Time is also an important factor while it may be difficult to group old mentioned referents with new ones. Because of this multiplicity of factors, choosing the right discursive grouping for a referential plural expression is ambiguous, and this ambiguity needs to be explicitly described.

We present a model of grouping based on reference domains theory (Salmon-Alt, 2001) that considers that a reference operation consists of extracting a referent in a domain. However the original theory barely takes into account plural reference. This paper shows how several entities can be grouped together by building a new domain and how they can be accessed later on. It introduces also the notion of super-domain D^+

that represents the access structure to all the plural referents of type D . This work is currently being implemented and evaluated in the MEDIA project of the EVALDA framework, a national french understanding evaluation campaign (Devillers, 2004).

2 Groupings of Referents

Several kinds of clues can specify that referents should be grouped together, or at least could be grouped together. These clues may occur at several language levels, from the noun phrase level to the rhetorical structure level. We have not explored in detail the different ways of groupings entities together in a discourse or dialogue. What is described here are just some of the phenomenon we got confronted with while developing a reference resolution module for a dialogue understanding system.

- **Explicit Coordination** - The most basic way to explicitly express the grouping of two or more referents is using a connector such as *and, or, as well as*, etc.

*“Good afternoon, I would like to book a single room **and** a double room”*

- **Implicit Sentential Coordination** - An implicit coordination occurs when two or more referents of the same kind are present in one sentence, without explicit connector between them. *“Does the hotel de la gare have a restaurant, like the Holiday Inn?”*

- **Implicit Discursive Coordination** - Such a coordination occurs when several reference are evoked in separate sentences. The grouping must be done based on rhetorical structuring. Here we consider short pieces of dialogue, admitting only one level of implicit discursive coordination. *“I would like an hotel close to the sea... I also need an hotel downtown... And the hotels have to accept dogs.”*

- **Repetitions/Specifications** – In some particular cases, groupings make explicit a previous expression. For instance “*Two rooms. A single room, a double room*”.

3 Reference Domain Theory

We are willing to try a pragmatic approach to reference resolution in practical multimodal dialogues (Gieselmann, 2004). For example we need to process frequent phenomena like ordinals for choosing in a list (discursive, or visual) or otherness when re-evoking old referents. Hence keeping the track of the way the context is modified when introducing a referent or referring, is mandatory. The Reference Domains Theory (Salmon-Alt, 2001) supposes that every act of reference is related to a certain domain of interpretation. It endorses the cognitive grammar concept of domain, defined as a cognitive structure presupposed by the semantics of the expression (Kumar et al., 2003). In other words, a referring expression has to be interpreted in a given domain, highlighting and specifying a particular referent in this domain. A reference domain is composed of a group of entities in the hearer’s memory which can be discursive referents, visual objects, or concepts. It describes how each entity could be addressed through a referential expression.

This theory views the referring process as a dynamic extraction of a referent in a domain instead of a binding between two entities (Salmon-Alt, 2000). Hence doing a reference act consists in isolating a particular entity from other rejected candidates, amongst all the accessible entities composing the domain (Olson, 1970). This dynamic discrimination relies on projecting an access structure focusing the referent in the domain. The domain then becomes salient for further interpretations. The preferences for choosing a suitable domain are inspired from the Relevance theory (Sperber & Wilson, 1986) taking into account such focalization and salience.

Landragin & Romary (2003) have also studied the usage of reference domains in order to model a visual scene. The grouping factors for visual objects are those given by the Gestalt theory, proximity, similarity, and good continuation. Each perceptual groups or groups designated by a gesture could be the base domain for an extraction. Referential expressions work the same way either the domains are discursive, perceptual or gestural, they extract and highlight

referents in these domains. See (Landragin et al., 2001) for a review of perceptual groupings.

4 Basic Type

A referential domain is defined by:

- a set of entities accessible through this domain (ground of domain),
- a description subsuming the description of all these entities (type of domain),
- a set of **access structures** to these entities.

For instance: “*the Ibis hotel* (h_1) and *the hotel Lafayette* (h_2)” forms a referential domain, whose type would be Hotel, and whose accessible entities would be h_1 and h_2 , themselves defined as domains of type Hotel. These two hotels could be accessed later on by their names.

4.1 Access structures

We suppose that the distinction between the referents from the excluded alternatives requires highlighting a discrimination criterion opposing them. This criterion behaves like a **partition** of the accessible entities, grouping them together according to their similarities and their differences. A partition may have one of its parts **focused**. There are, at least, three kinds of discrimination criteria:

- **discrimination on description.** Entities can be discriminated by their type, their properties, or by the relations they have with other entities. For example the name of the hotels is a discrimination criterion in “*the Ibis hotel and the hotel Lafayette*”.
- **discrimination on focus.** Entities can also be discriminated by the focus they have when they are mentioned in the discourse or designed by a gesture. For example, “*this room*” would select a focused referent in a domain, whereas “*the other room*” would select a non-focused one.
- **discrimination on time of occurrence.** Entities can finally be discriminated by their occurrence in the discourse. For example “*the second hotel*” would discriminate this hotel by its rank in the domain.

4.2 Classical resolution algorithm

Each activated domain belongs to list of domains ordered along their recentness (the referential

space). The resolution algorithm consists of two phases:

1. Searching a suitable, preferred domain in the referential space when interpreting a referring expression. The suitability is defined by the minimal conditions the domain has to conform to in order to be the base of an interpretation (particular description, or presence of a particular access structure with focus or not). The main preference factor is the minimization of the access cost (recentness or salience), however other criteria like thematic structure could be taken into account and will be future work. Each domain is tested according to the constraints given by the referential expression. We allow several layers of constraints for each type of expression : if the stronger constraints are not met, then weaker constraints are tried.
2. Extracting a referent and restructuring the referential space, taking into account this extraction. It not only focuses the referent in its domain, but also moves the domain itself to a more recent place. When one referent acquires the focus, the alternative members of the same partition loose it.

This generic scheme is instantiated for each type of access modes (a modality plus an expression). For example a definite “*the N*” will search for a domain in which a particular entity of type “N” can be discriminated, and the restructuring consists in focalizing in this domain the referent found. See (Landragin & Romary, 2003) for a description of the different access modes.

The algorithm highlights the two types of ambiguities, domain or referent ambiguities, which occur when there is no preference available to make a choice between multiples entities in the first or the second phase. We guess that natural ambiguities should eventually be solved through the dialogue between the agents of the communication.

5 Super-Domains

In order to take groupings into account in the Reference Domains Theory, we introduce two constructs in our formal toolbox. Indeed, having only one kind of domain construct doesn’t allow for a correct distinction between different referent statuses.

First we distinguish plural and simple domains. The simple domains D serve as bases for profiling, or highlighting, a **subpart**, or **related part** of a simple referent. For instance, if $D =$

Room, then one can profile a *Price* from D . The plural domains D^* serve as either as a **generic base** or as a **plural representative** for profiling a simple domain D . A generic base is mandatory in our model to support the insertion of new extra-linguistic referents evoked with an indefinite construct (for instance “*I saw a black bird on the roof*”), while plural representatives are used for explicit groupings. A domain D^*_1 can also be profiled from a D^*_0 , provided D^*_1 profiles a subset of the elements of D^*_0 .

Second, we introduce the notion of **super-domain** D^+ , from which a D^* can be profiled. The relations allowed between domains are represented on figure 1. A super-domain D^+ is the domain of all groupings D^* , including a special D^*_{all} grouping which is the representative of all evoked instances of a given category. This configuration is not intended to deal with long dialogues where several, trans-sentential groupings occur, and where older groupings may become out of access. Doing this would require a rhetorically driven structuring of the D^*_{all} .

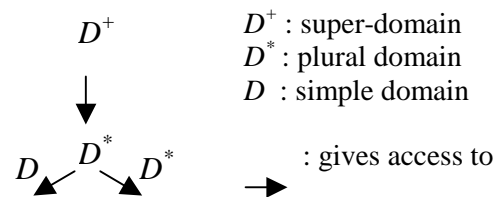


Figure 1: Access structure of Reference Domains

As Reference Domain Theory is primarily targeted toward extra-linguistic referents occurring in practical dialogue, the construction of the domain trees, representing the supposed structuring of referents accessibility, is based on ontology. As a consequence, for each “natural” type and each subtype (for instance $Room \wedge Single$), a domain tree is potentially created (actually, one can easily imagine how this creation may be driven ‘on-demand’).

Another evolution from the initial Reference Domain Theory is the possibility to focalize several items of a partition. Indeed, since the resolution algorithm can focalize a whole plural domain, all elements of this domain must be focalized in all the plural domains they occur in. In order to refer to plural entities the idea is to build plural domains dynamically : when some sentence-level grouping, either implicit or explicit occurs or when a plural extra-linguistic referent is evoked, a D^* is created and focussed

in D^+ , with each of its components as children, when possible (that is, when each component is described). When new extra-linguistic referents (singular or plural) are evoked, they are individually profiled under the D^*_{all} corresponding to their types (that is, their “natural” type, and all the subtypes they are eligible to).

In short, for all referents of type D :

- they become subdomains of D^*_{all}
- if they are plural referents, they also build up a focalized subdomain of D^+
- all the referents of a given type are then grouped together under a new focalized subdomain of D^+ .

Figure 2 illustrates the state of the $Hotel^+$ domain tree after a scenario with three dialogue acts, the first one introducing $Hotel_1$, the second one inserting a grouping of $Hotel_2$ and $Hotel_3$, and the third one referring to it.

U₁: The Ibis Hotel ($Hotel_1$) is too expensive

S₁: Maybe the Hotel Lafayette ($Hotel_2$) or the Hotel de la cloche ($Hotel_3$)

U₂: Those hotels are too far from the airport.

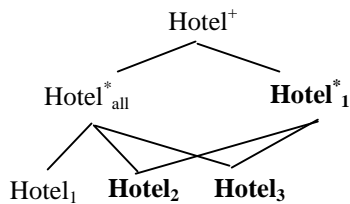


Figure 2: A domain tree built from a scenario above (focus in **bold**)

The operations are the following :

U₁ : $Hotel_1$ becomes a subdomain of $Hotel^*_{all}$ which gains focus in $Hotel^+$.

S₁ : $Hotel_2$ and $Hotel_3$ become subdomains of $Hotel^*_{all}$. In addition $Hotel_2$ and $Hotel_3$ are grouped in $Hotel^*_1$ which gains the focus in $Hotel^+$ while $Hotel^*_{all}$ loses it.

U₂ : The pronoun is solved in $Hotel^+$, and $Hotel^*_1$ is retrieved.

One can see that $Hotel^*_{all}$ is inaccessible by a generic expression like a demonstrative without modifiers but only by a special expression like “all the hotels”. In our point of view, the reason is that the grouping $Hotel^*_1$ lowers the salience of $Hotel^*_{all}$.

6 Implementation

We used description logics for modelling domains and domain-reasoning. One has to deal with plural entities and can follow (Franconi, 93) by using collection theory, representing collections as individuals and membership by a role (plus plural quantifiers). But we should use another way considering that the inference engine we use, Racer (Haarslev and Möller, 03), does not take into account ALCS. Hence we tried representing the domains by concepts, given their semantic are set of individuals. The domain D^+ corresponds to the concept D , and the domain-subdomain relation is a subsumption. All basic manipulation with domains could be done using Tbox assertions. Additionally, a partition structure is simply a sequence of subdomains which are different from each other (disjoint concepts) and whose elements could be focussed. The algorithm goes through the referential space and tests each domain in the recency order against the constraints given by the referential expression. Conceptual tests on the description and partition tests on the focus or possible discriminations are made to retrieve the domain and the referent. If none are found, they may be created by accommodation. Groupings are created only for explicit coordinations, implicit sentential coordinations (two referents could be grouped if they have the same basic type) and some kind of specifications.

Domains and groupings creation entails the creation of new concepts in the Tbox. Each concept insertion requires a costly reclassification, therefore we preferred an approximation considering only that new groupings assert primitive concepts. Other domains are concept terms *i.e.* descriptions which do not have to be asserted in the Tbox automatically.

Implicit discursive groupings are not implemented considering the need of a rhetorical structure (like in SDRT, Asher 93) or a mental space model. The following example shows the needs :

U₁ : I would like an hotel (h_1)

S₁ : I propose you the hotel Ibis (h_2) and the Lafayette hotel (h_3).

Hotel h_1 could very hardly be grouped with h_2 and h_3 , even by “all these hotels” (or maybe by a third speaker). We guess among other factors that they belong to different levels of interpretation, h_1 in the domain of the desires of

the user, and the others in the domain of existing hotels. The link between the two domains is possible if one knows that S_1 is an answer of to U's request. Such discrimination criterion and high level domains are not yet implemented. Instead we concentrated on extra-linguistic referents which are assumed to be interpreted in the real/system world (like hotels, rooms). We are currently testing the approach to see if it could be extended to any type of entities provided accurate discrimination criteria (like the predication).

7 Example

A sample dialogue (table 1) is analyzed through the preceding algorithm. This example shows how the referents introduced in an explicit coordination could be referenced as a whole “*the two hotels*”, or extracted discriminately by an ordinal “*the second one*” or by an otherness expression “*the other one*”. All the subdomains of H^+ (i.e. the plural domains of hotels) are indicated *after* each interpretation using a simplified notation. Only the ordered list of accessible entities and their focalization (bold) are noted for each subdomain. For instance $H_{all}^* = (h_1, h_2, \mathbf{h}_3)$ means that the domain H_{all}^* is focalized in H^+ , and that h_3 is focalized in H_{all}^* .

Dialogue	H^+
U: <i>Is there a bathroom at the Ibis hotel (h_1) and the hotel Lafayette (h_2)?</i>	$H_0^* = (\mathbf{h_1}, \mathbf{h_2})$ $H_{all}^* = (\mathbf{h_1}, \mathbf{h_2})$
S: <i>No they don't have bathrooms</i>	$H_0^* = (\mathbf{h_1}, \mathbf{h_2})$ $H_{all}^* = (\mathbf{h_1}, \mathbf{h_2})$
S: <i>But I propose you the Campanile hotel (h_3)</i>	$H_0^* = (h_1, h_2)$ $H_{all}^* = (h_1, h_2, \mathbf{h_3})$
U: <i>Hmm no, how much were the two hotels?</i>	$H_0^* = (\mathbf{h_1}, \mathbf{h_2})$ $H_{all}^* = (\mathbf{h_1}, \mathbf{h_2}, h_3)$
S: <i>The hotel Lafayette is 100 euros, the Ibis hotel is 75 euros</i>	$H_1^* = (\mathbf{h_2}, \mathbf{h_1})$ $H_0^* = (\mathbf{h_1}, \mathbf{h_2})$ $H_{all}^* = (\mathbf{h_1}, \mathbf{h_2}, h_3)$
U ₁ : <i>Ok, I take the second one</i>	$H_1^* = (h_2, \mathbf{h_1})$ $H_0^* = (\mathbf{h_1}, h_2)$ $H_{all}^* = (\mathbf{h_1}, h_2, h_3)$
U ₂ : <i>Ok, I take the third one</i> U ₃ : <i>and the other one ?</i>	$H_1^* = (h_2, \mathbf{h_1})$ $H_0^* = (h_1, h_2)$ $H_{all}^* = (h_1, h_2, \mathbf{h_3})$

Table 1: Example of dialogue (focus in bold)

In order to interpret U_1 , U_2 or U_3 one needs to rely on the previous structuring of H^+ . In U_1 , the previously focalized domain H_1^* is preferred to be the base for interpreting “*the second one*” because of the order discrimination. This leads to extracting h_1 hence focalizing it in H_1^* but also in H_0^* and in H_{all}^* . In U_2 , H_1^* cannot be the base for interpreting “*the third one*” because no entity could be discriminate this way. Therefore the only suitable domain is H_{all}^* . It is also impossible to interpret U_3 : “*the other one*” in H_1^* because of the lack of a focus discrimination between h_1 and h_2 .

It is however possible to choose H_{all}^* for the domain of interpretation: the excluded referents h_1 and h_2 are unfocused while h_3 gains focus.

8 Evaluation in progress

This work is currently being evaluated in the MEDIA/EVALDA framework, a national understanding evaluation campaign. (Devillers et al., 04). It aims to evaluate the semantic and referential abilities of systems with various approaches of natural language processing. The results of each system are compared to manually annotated utterances transcribed from a Woz corpus in a hotel reservation task. For the referential facet, referential expressions (excluding indefinites, and proper names) are annotated by a semantic description of their referents.

Our system which relies on a symbolic approach using deep parsing and description logics for semantic currently scores 64% (f-measure) for identifying and describing accurately the referents. We guess that such evaluation will be an occasion for us to test different hypothesis on reference resolution using domains (for exemple different criteria for grouping). However we do not have yet more precise results on plurals and ordinals specifically.

9 Conclusion

The extension we made to the Reference Domains Theory is still limited because it considers only extra-linguistic referents, i.e. those also having an existence outside discourse. In addition the trans-sentential groupings are not fully studied yet. We guess that such groupings should need a rhetorical description of the discourse or dialogue. In spite of its limits, the extension can render dynamic effects allowing ordinals and otherness in plural contexts. An

implementation in description logics is currently being evaluated in the MEDIA/EVALDA framework.

References

- Nicholas Asher. 1993. Reference to Abstract Objects in English: A Philosophical Semantics for Natural Language Metaphysics. In *Studies in Linguistics and Philosophy*, Kluwer, Dordrecht.
- Laurence Devillers, Hélène Maynard, Stéphanie Rosset, Patrice Paroubek, Kevin McTait, Djamel Mostefa, Khalid Choukri, Caroline Bousquet, Laurent Charnay, Nadine Vigouroux, Frédéric Béchet, Laurent Romary, Jean-Yves Antoine, Jeanne Villaneau, Myriam Vergnes, and Jérôme Goullian. 2004. The French MEDIA/EVALDA Project : the Evaluation of the Understanding Capability of Spoken Language Dialog System. In *Proceedings of LREC 2004*, Lisbon, Portugal.
- Carola Eschenbach, Christopher Habel, Michael Herweg, Klaus Rehkämper. 1989. Remarks on plural anaphora. In *Proc. Fourth Conference of the European Chapter of the Association for Computational Linguistics*.
- Enrico Franconi. 1993. A treatment of plurals and plural quantifications based on a theory of collections. *Minds and Machines* (3)4:453-474, Kluwer Academic Publishers, November 1993
- Petra Gieselmann: 2004. Reference Resolution Mechanisms in Dialogue Management. In: *Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue* (CATALOG), Barcelona, 2004.
- Volker Haarslev, and Ralf Möller. 2003. Racer: A Core Inference Engine for the Semantic Web. In *Proceedings of the 2nd International Workshop on Evaluation of Ontology-based Tools* (EON2003), located at the 2nd International Semantic Web Conference ISWC 2003, Sanibel Island, Florida, USA, October 20, 2003, pp. 27-36.
- Ashwani Kumar, Susanne Salmon-Alt, and Laurent Romary. 2003. Reference resolution as a facilitating process towards robust multimodal dialogue management: A cognitive grammar approach. In *International Symposium on Reference Resolution and Its Application to Question Answering and Summarization*.
- Frédéric Landragin, and Laurent Romary. 2003. Referring to Objects Through Sub-Contexts in Multimodal Human-Computer Interaction. In *Proc. Seventh Workshop on the Semantics and Pragmatics of Dialogue* (DiaBruck'03), Saarbrücken, Germany, 2003, pp. 67-74.
- Frédéric Landragin, Nadia Bellalem and Laurent Romary. 2001. Visual Saliency and Perceptual Grouping in Multimodal Interactivity. In: *First International Workshop on Information Presentation and Natural Multimodal Dialogue*, Verona, Italy, 2001
- David R. Olson. 1970. Language and Thought: Aspects of a Cognitive Theory of Semantics. *Psychological Review*, 77/4, 257-273.
- Susanne Salmon-alt. 2000. Interpreting referring expressions by restructuring context. *Proc. ESSLLI 2000, Student Session*, Birmingham, UK, August 2000.
- Susanne Salmon-Alt. 2001. Reference Resolution within the Framework of Cognitive Grammar. *Proc. International Colloquium on Cognitive Science*, San Sebastian, Spain
- Dan Sperber and Deirdre Wilson. 1986. *Relevance, Communication and Cognition*. Basil Blackwell, Oxford.