

# Towards a Functional Approach of the Extensional Resolution of Intrinsic Reference in Practical Dialogue

Guillaume Pitel, Jean-Paul Sansonnet

LIMSI – CNRS  
Bât. 508 BP 133 F-91403 Orsay CEDEX  
[Guillaume.Pitel@limsi.fr](mailto:Guillaume.Pitel@limsi.fr)

## Abstract

In this paper, we propose a functional approach to overcome the lack of expressiveness of predicative approaches for extensional reference resolution. This later approach indeed forces designers of dialogue systems to use specific heuristics when solving some reference cases, which they can't integrate in a unified representation framework. As we focus on extensional reference resolution in a generic framework for human computer dialogue system, we point the fact that the process of resolving referential expressions needs to take context into account even for expressions involving intrinsic characteristics of elements. Thus we propose a representation model for the semantics of **intrinsic referential extractors**. This model is built on three functions, the first one serves to calculate the similarity ratio between two elements, the two others serve to partition a reference domain which is previously sorted by the first function

## Keywords

Reference Resolution, Intrinsic Reference, Functional Approach, Human Computer Dialogue

# 1 Introduction

In the domain of human machine dialogue, genericity is of increasing importance. Several parts of the technology are concerned, from spoken language recognition to extensional reference resolution. There are two reasons for this, first is quantitative: the adaptation and development cost of dialogue systems, the second is qualitative: some researchers consider that a system that has to be rewritten to cover a new task does not implement a theoretically coherent model of interaction. This paper examines the issue of *extensional resolution of referential expressions* in the context of a practical dialogue system, *extensional resolution of reference* is defined in (Byron, 2001) as the research of *elements of the mediated software*<sup>1</sup> referred to by an utterance. The process of finding objects from words is often reduced to either repeated selection based on first-order predicates, or built around specific heuristics (Winograd, 1973). These approaches to extensional resolution do not allow a generic framework definition for extensional reference resolution in practical dialogue systems<sup>2</sup>.

This paper presents a reference resolution model, which come as an extension of the frame defined in (Salmon-Alt, 2001). Our model aims to provide a way to describe detailed referential extractors, while staying at an abstract level, thus permitting to access deep structures of the mediated software within a formalism that is a direct extension of a natural language processing system. In Salmon-Alt's words, parts of a *referring expression* are considered to act as several *differentiation criteria* in the element selection process of the reference resolution. These criteria are used to organize sets called *reference domains*. Reference domains store possible candidates to a reference at each step of the resolution process, they also express a point of view about the situation, thus defining a reference frame or specifying a differentiation criterion for future resolutions. Contrary to Salmon-Alt's position about the representation of differentiation criteria, we adopt a procedural view on the meaning of referential extractors. They are then represented as active operators of element selection and sorting, and not as truth conditions used to choose candidates in a binary way. This paper examines the idea of using **differentiation functions** in place of logical criteria, we motivate this position at section 2, by showing that referential extractors' action viewed as logical predicates can't take context into account during the selection process. Thus we propose our model based on the view of the referent construction built on a context-dependent differentiation, supported by differentiation functions, section 3.

## 2 Intrinsic Referential Extractors

*Referential extractors* are the parts of a *referential expression* that play a role in the construction of the expression's extension, i.e. the objects of the "real world" referred to by

---

<sup>1</sup> The term "*mediated software*" represents the computer program with which we want to enable natural language communication. Dialogue system serves as a **mediator** between a human and a program, and our long-term goal is to propose a model that could uniformly cover the interpretation path between the two endpoints of the communication.

<sup>2</sup> Indeed, predicate approach, while declarative and thus generic in some respect, is not suited for processing some reference phenomenon, which require ad hoc heuristics.

the expression. Typically, in a sentence like “Take the **blue** box”, the word **blue** serves as a selector for objects having a color that can be characterized as blue. In this sentence, ‘**box**’ and ‘**the**’ are also referential extractors, the behavior of ‘**Le**’ and others determinants viewed as referential extractors is specifically studied for French language by Salmon-Alt.

In dialogue systems, the issue of representing objects<sup>3</sup> in the mediated software and their link with referential expressions has not been widely explored. Most dialogue models consider that they have access to an adequate representation, coherent with the way the user will refer to object in natural language. This can be observed even in late researches, for instance, (Salmon-Alt, 2001:149), (Byron, 2001) or (Dale et Reiter, 1995). Typically, (Byron, 2001) expresses restrictions due to the interpretation of a referring expression with a form of first-order predicate logic like (*color x RED*).

We propose to determine the minimal necessary expressiveness of the representation needed in order to represent referential extractor action in the process of referring. As (Dale et Reiter, 1995), we will start by distinguishing three kind of components used to refer to an element: intrinsic properties (e.g. size, color), relational properties (e.g. spatial position), and typological properties (e.g. triangle, square). In this paper, we will focus on the issue of representing the action of referential extractors that rely on intrinsic properties.

## 2.1 Definition

Intrinsic referential extractors rely on intrinsic properties of elements. These properties are characterized by their stability (they are always present), and their context-independence. We will demonstrate that even if the characteristics needed to *calculate* the right referent are intrinsic to the objects, and do not depend on any context, the intrinsic referential extractors actually need to use extrinsic information in order to resolve a referring expression. So, this will prove that whatever the referential extractors rely on, their action is always dependant on the context.

In the following, we will designate by **operating context** the current state of the mediated software, while **ontological context** represents the state of the graph of relations between categories and the knowledge that is attached to them.

## 2.2 Size and Color

In practical dialogue systems, semantics of size adjectives like *big* or *small* are often reduced to predicates. However, as (Dale and Reiter, 1995) notice, it seems obvious that the element referred to by an expression containing such an adjective is not supposed to own a property saying if it is absolutely big or small. The size of an element is an intrinsic characteristic, but the way natural language is used to refer to it makes use of extra information. Thus, referential extractors relying on size of elements, depend on **operating context**.

---

<sup>3</sup> We distinguish on one hand algorithmical representations of the mediated software, which we refer as **objects**, composed of a **type** and several **attributes**; on the other hand, entities perceived by the user are referred by us as **elements**, defined by their **category** and their **characteristics**.

One could consider that this is an exception to the rule, and that reference about size should actually be classified as a case of relational reference. We however believe that the general case of reference resolution is context-dependent, and that the rare cases of almost context-independent reference are actual exceptions to the rule.

In order to support this hypothesis, we will show, from experiments based on fig. 1, that color adjectives are context-dependent too. On case 1 of this figure, “the blue leaf” refers to the leaf on the left. On the other hand, on case 2, if one asks a user to select no blue squares, she will choose the three squares in the bottom of the figure. Finally, if one asks for the color of the car on the case 3, users will preferably answer “black”, even if they answer yes to the question “Is the car blue?”. In these three cases, however, the leaf, the bottom-left square and the car are of the same color.

We explain this behavior by the fact that a leaf has a typical color of green, yellow, red or brown, and thus can be categorized as blue as soon as its color get out of these typical colors and tends toward blue. On the other hand, cars like the one shown are likely to be of particular color, and a dark color appeal the user to see it as being black. This phenomenon is exactly what we call *dependence to the ontological context*. On case 2, we illustrate the fact that the verbal categorization of colors depends on the environment, and thus that even if the bottom-left square would be considered as blue if it is seen isolated, it is categorized as non blue when surrounded by more blue elements. This is what we call *dependence to the operating context*.

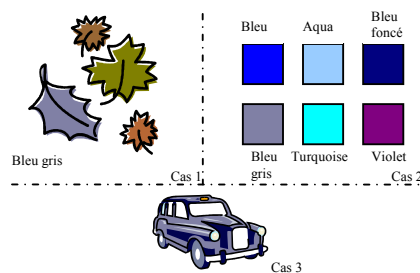


FIG. 1 : Categorization of colors depending on the environment and the type of elements

These observations lead us to make two propositions. The first is that reference resolution process should rely on the use of functions, not on predicates. The hypothesis is supported by conclusions made by (Pateras *et al.*, 1995) in a paper dealing with dialogue cases where a database access is not sufficient to handle reference resolution. Pateras *et al.* propose to use functions based on fuzzy sets to choose the right referent. However these functions are defined relative to a particular task, and are not designed to take ontological or operating context into account, because they have no influence in this very task. The same situation can be found in (Lammens & Shapiro, 1993) where the authors build color categorization functions with a learning algorithm. The second of our propositions is that the functions we use to represent the semantics of referential extractors should be able to take the context as one of their arguments.

### 3 Functional Differentiation Model

We propose to represent referential extractors with a **structure of functions** that will be used by a generic resolution algorithm. For the purpose of illustrating our view, we define a situation as the one described in FIG. 3. In order to focus the discussion on the issue of resolving intrinsic references, we will consider that elements of the mediated software are *objects* (in the object-oriented algorithmic sense). These objects are defined by a type and two attributes: color (a 3-uple of values in  $[0,1]$ ) and a size (a value in  $]0, +\infty[$ ).

#### 3.1 Resolution Algorithm

The resolution algorithm is considered to be triggered during the natural language analysis process when a referring expression has been isolated and is supposed to refer to a specific element (that is, it does not refer to a generalization). The rules leading the algorithm follow the guidelines defined in (Salmon-Alt, 2001), so that the type (definite, indefinite, demonstrative, pronominal) of the referring expression determines the way the initial reference domain is built, and how the extractors are used to return the final result of the reference resolution in the *restructuration* phase.

We consider each referential extractor as a sub-process of the following form:

1. Transform the original reference domain, sorted and partitioned, through a point of view that would make the extractor able to process the domain.
2. Use the  $f_{\text{Simil}}$  function of the extractor to sort the domain.
3. Use the  $f_{\text{Excl}}$  function to partition the domain among possible and impossible objects.
4. Use the  $f_{\text{Pref}}$  function to partition the possible part among preferred and possible objects.

Compared to Salmon-Alt's differentiation criteria, our algorithm makes use of two different partition functions ( $f_{\text{Excl}}$  and  $f_{\text{Pref}}$ ) in order to produce a ternary partition instead of a binary one. Indeed, as we introduce a total order relation in the differentiation process, we are faced with an issue that was hidden when using simple predicates: when referring to several elements, like in "*remove blue squares*", the user refers to several elements in the top of the list sorted by blueness; however, squares less blue but still blue must be distinguished from non-blue squares, because if the user asks for "*remove big blue squares*", and big squares are not in the top of the list sorted by blueness, they must be passed from *blue* extractor to *big* extractor in the list of possible candidates for the referring expression. So it is necessary to distinguish between best candidates, needed to answer to group references, and possible and impossible candidates in order to respond "no match" or to near ambiguity references.

### 3.2 Referential extractors representation

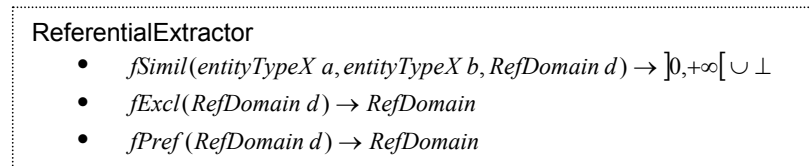


FIG. 2 : Referential extractor representation

The ***fSimil*** function is used to sort a set of elements among a given characteristic (say, *blueness* or *bigness*). For instance, the *fSimil* function of the referential extractor “blue” should give the ratio between the colors of two elements projected along the blue axis<sup>4</sup>. If *a* is more blue than *b*, then the returned value is above 1, while the contrary returns a value between 0 and 1 (exclusive), and equality returns 1. If one of the two colors is outside the area considered to be the limit of blue colors, returned value is  $\perp$ .

The ***fExcl*** function serves to select in the sorted list (produced by the resolution algorithm using *fSimil* function) objects that will be excluded from the candidates list that is going to be submitted to the next extractor. All extractors do not select the same way, for instance extractor *big* will select objects of the same size as the littlest one (that is, the last one in the list). For any extractor for *color*, this function selects all objects after the first similarity ratio equals to  $\perp$ .

The ***fPref*** function selects the preferred objects for a given extractor. Here again, there is no general rule for the function. The extractor *big* could select it based on any heuristic (either the 30% upper sizes, or all the objects from the beginning to the first point where the secondary differential coefficient is over a given threshold).

### 3.3 Practical example

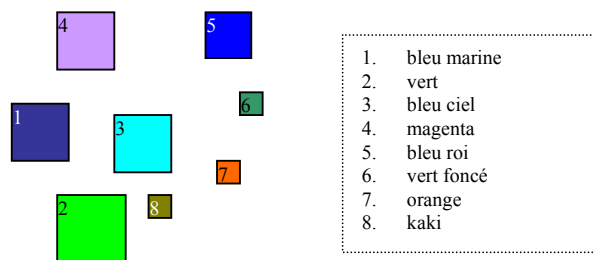


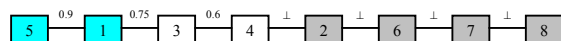
FIG. 3 : Situation de test pour la représentation fonctionnelle des couleurs et des tailles.

From the situation presented FIG. 3, we illustrate how to use differentiation functions in the case of resolving the referring sub-expression “big blue square”.

---

<sup>4</sup> Do not confuse with the blue component of a red, green and blue decomposition of color. The *blue axis* or *dimension* tends mostly toward a mathematical correlate of a cognitive dimension.

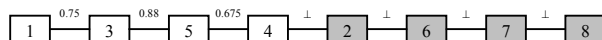
We take as a starting point for reference resolution the referential domain of all the squares of the scene. First we apply the referential extractor for *blue*, we will argue later for the ordering of the execution of each extractor. Theoretically, this function will produce, when applied to a set of  $n$  objects, a matrix of  $n \times n$  elements. For all practical purpose, one can consider that in many cases,  $fSimil_G(x,z) = fSimil_G(x,y) \times fSimil_G(y,z)$ , and there are some relatively low-cost algorithm to build a sorted list from this function. This point could be subject to discussion, but we will assume that it is a reasonable assumption. Using the  $fSimil$  function, we construct an ordered list whose links between elements are annotated with similarity ratios, and partition it with the two functions  $fExcl$  (blue cells on the left) and  $fPref$  (gray cells on the right).



Next, we apply the extractor *big*. For the example, we do not consider that excluded elements are eliminated from the resolution process.



Combination function is:  $simil(x,y) = \alpha.simil_{Bleue}(x,y) \times \beta.simil_{Grand}(x,y)$  ( $x \times \perp = \perp$  and  $x.\perp = \perp$ ) with  $\alpha, \beta$  coefficients depending on the relative weight of each dimension in the determination of the final result, that should be calculated from a psychological experimentation. We obtain ( $\alpha = \beta = 1$ .)



The result of reference resolution for “big blue square”, in the situation of FIG. 3 gives square number 1 as the best candidate. Same method applied to “small blue square” produces the square number 5. In these two cases, the result seems to corroborate the basic intuition and the human validation<sup>5</sup>.

### 3.4 Arranging extraction operations

In order to justify the order for executing extractors, we rely on propositions from (Dale et Reiter, 1995) made about the relative importance of objects’ characteristics in the framework of referential expressions generation. The authors mainly base their research on the implication of Grice’s maxims (Grice, 1975) on reference generation, following the principle of maximum economy in language generation. Authors roughly propose to take characteristics into account in the following order: type > intrinsic property > relational property.

From this proposition, we propose to generalize saying that the cost of treating a given characteristic in a reference is related to the quantity of contextual information that is necessary to take into consideration.

---

<sup>5</sup> We have made a little experimentation with a dozen of volunteers.

*The more the operating context is important in the computation related to an extractor, the later it should be executed in the resolution chain.*

As the perception of colors is only a little dependent on the situation, whereas the perception of the size is highly related to the other objects present in the scene, the arrangement of extractor execution is *color* then *size*. This approach offer the advantage to explain also why the type (the noun) is always the most important of the characteristics and is always considered before other properties. So the expression “*the big blue square*” is disambiguated as: “*the big ones among the blue ones among the squares*”.

## 4 Current work and perspectives

We propose a fonctionnal representation of intrinsic referential extractors, in a formalism that could integrate resolution models like the one of Salmon-Alt. This approach aims to cover as well relational and typological extractors as well, but there is still some important work to do in order to reach this goal. One of the issues raised by this extension to the preliminary model is the need for a mechanism to measure the distance between two elements in any dimension, and this implies to perform some kind of metaphorical transformation of reference domains from one dimension to another.

The AMI team at LIMSI-CNRS is currently developping this mechanism inside the INTERVIEWS project (Sansonnet, 2002) (Sabouret, 2001), built around the concept of conversational agents for assistance to ordinary people. The formalism presented in this paper, as well as the overall model of natural language analysis integrating it, is under development in this framework. There is no evaluation of this system currently, but we plan to reach a working platform in order to be able to compare results from our reference resolution method with the results of psychological experimentations.

## References

- Byron D.K., Allen J.F. (2002) *What's a Reference Resolution Module to do? Redefining the Role of Reference in Language Understanding Systems*, Proc. DAARC2002
- Dale R., Reiter E. (1995), *Computational Interpretation of the Gricean Maxims in the Generation of Referring Expressions*. Cognitive Science
- Dzikovska M.O., Byron D.K. (2000), *When is a union really an intersection? Problems interpreting reference to locations in a dialogue system*, Proc. GOTALOG'2000
- Dzikovska M.O., Allen J.F., Swift M.D. (2002), *Finding the balance between generic and domain-specific knowledge: a parser customization strategy*. Proc. LREC 2002
- Dzikovska M.O., Swift M.D., Allen J.F. (2003), *Constructing custom semantic representations from a generic lexicon*. Proc. 5th IWCS
- Grice H.P. (1975). *Logic and conversation*. In P. Cole and J. Morgan, editors, *Syntax and Semantics: Vol 3, Speech Acts*, pages 43--58. New York, Academic Press

Lammens J.M., Shapiro S.C., *Learning Symbolic names for Perceived Colors*, in Machine Learning in Computer Vision: What, Why and How?, AAAI-TR FSS93-04

Ozkan N. (1993), *Vers un modèle dynamique du dialogue : analyse de dialogues finalisés dans une perspective communicationnelle*, PhD Thesis, INP Grenoble

Pateras C., Dudek G., Mori R. D. (1995), *Understanding Referring Expressions in a Person-Machine Spoken Dialogue*. Proc. ICASSP'95, Detroit, MI

Sabouret N., Sansonnet J.P. (2001), *Un modèle de requêtes sur le fonctionnement de composants actifs*, Proc. MFI'01

Sansonnet J.P., Sabouret N., Pitel G. (2002), *An Agent Design and Query Language dedicated to Natural Language Interaction*, Poster AAMAS 2002

Salmon-Alt S. (2001), *Référence et dialogue finalisé : de la linguistique à un modèle opérationnel*, PhD Thesis, Université H.Poincaré - Nancy 1, France. Mai 2001.

Schang D. (1995), *Application de la notion de cadre aux énoncés de positionnement et de référence*, Research Report n° 2529, Unité de Recherche INRIA Lorraine.

Winograd T. (1973), *A Procedural Model of Language Understanding*, Computer Models of Thought and Language, Roger Schank & Kenneth Colby éditeurs., W. H. Freeman Press.