

Attempt on the elaboration of good expression principles for information retrieval problem

Stéphane Gorla, Amos David, Jean-Claude Derniame, Philippe Geffroy

► **To cite this version:**

Stéphane Gorla, Amos David, Jean-Claude Derniame, Philippe Geffroy. Attempt on the elaboration of good expression principles for information retrieval problem. 8th World Multi-Conference on SYSTEMICS, CYBERNETICS AND INFORMATICS - SCI 2004, Jul 2004, Orlando, United States. XV, pp.306-310, 2004. <sic_00001278v2>

HAL Id: sic_00001278

https://archivesic.ccsd.cnrs.fr/sic_00001278v2

Submitted on 24 Aug 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Attempt on the elaboration of good expression principles for information retrieval problem

Stéphane GORIA; Amos DAVID; Jean Claude DERNIAME

{ Loria, Campus Scientifique,
BP 239, 54506 Vandoeuvre-les-Nancy (France) }

and

Philippe GEFFROY
CEIS,
BP 632, 75367 Paris Cedex 08 (France)

Abstract

If simplified, every information retrieval problem can be solved when the information need implied by its expression has been identified. We are interested in the criteria used in realising a good information retrieval problem expression. We have listed these criteria through some principles where maxims, which first characterized the communication between two persons are applied. We choose to use the gricean maxims because they are the most favoured for this type of situation. Secondly, we have tried to identify some others principles that can be used to realise a good information retrieval problem expression. The principles by Grice can not resolve all forms of error associated with this particular form of communication. In our work, we defined three other principles namely: *adhesion principle*, *reformulation principle*, *memorization principle*. We give some examples of situations, where the principles we have formulated are not applicable and the consequences. We present the possible applications of our new model: MIRABEL, which can help in the description of information retrieval problem from. It also compels its user to use essential good expression principle implicitly.

Keywords: information retrieval, problem modelling, problem understanding, problem expression, problem formulation, maxims of conversation, pragmatics, MIRABEL model

INTRODUCTION

Every information retrieval problem (IRP) can be solved if the information need it implies is well identified by its expression. So, our concern in this paper is the principle of good IRP formulation that should be pragmatic. The formulation should be pragmatic because a problem depends highly on the context in which it is identified. In this paper we will present in the first part, our approach to solving an IRP, then we will define a list of principles that help in problem expression and finally we will show some of the consequences of not respecting these principles when solving an IRP. This will permit us to justify notably the usability of our model MIRABEL.

1. CONTEXT FOR DEFINING AN IRP EXPRESSION

1.1 SITE TEAM MODELS

Our research team studies economic and competitive intelligence environments, with a view to create a system that aids in understanding the real information needs of the users and in particular the resolution of IRPs. It is within this framework that we got interested in contextualising IRP. Several models have been proposed by members of this team for purpose. The DMP (Decision-Maker Problem) model by Najoua Bouaka [1] presents the IRP from a decisional point of view in terms of decision-maker's characteristics, his organisation's characteristics and environment parameters. The WISP model (Watcher Information Search Problem) by Philippe Kislin [8] presents an IRP from the process point of view between the decision-maker and the watcher for IRP resolution. The EQUA²te model (Explore, Query, Analyse and

Annotate) by Amos David and Odile Thierry [4] presents an IRP from an information system and user point of view. The MIRABEL model (Model for Information Retrieval query Annotations¹ Based on Expression Levels) presents, like its name indicates, an IRP from its expression point of view. MIRABEL represents a certain number of parameters for descriptions linked with the understanding and archiving of an IRP. These parameters are hierarchic (see section 4.1). In this paper, we will use the model MIRABEL but limit the actors implicated in an IRP to two: the PCS (the Person Charged with Solving an IRP) and the PEI (the Person who Expose the IRP to a PCS).

1.2 A PRAGMATIC APPROACH

We adopt here a pragmatic approach to define our expression principles of an IRP. Indeed, we are interested in IRP contextualisation essentially because of its influence on the formulation. Also, the pragmatic approach proposes that signification of one phrase does not depend only on its contents but also on other contextual information. In addition, we believe that the contextual information is not neutral. It permits some form inferences. In fact our approach will be linked with inferential pragmatics. We believe that the sense of a problem expression can be explained via inferential calculus. This inferential calculus is supposed to be generated by pragmatic rules. This joins notably the definition of J. Moeschler [10].

1.3 A STUDY BASED ON KNOWLEDGE TRANSFER

Many works have been carried out on user modelling and environment modelling in economic intelligence context. Our team works on user modelling too, but we believe that there is need to integrate the approach of knowledge transfer. In fact, rather than trying to model the knowledge and the beliefs of the two persons that collaborate to solve an IRP [9], we create a set of shared knowledge which is stored inside a particular model. That is why we search for the collaborative principles between PEI and PCS which will optimise the result of their collaboration.

2 KNOWLEDGE TRANSFER AND IRP UNDERSTANDING

We will adopt the basic assumption that to resolve an IRP, we must identify the information needs which it implies. We must choose one basic element to begin the identification of these needs. Reasonably, we can take the IRP formulation as the first need declaration. Thus, we first suppose the validity of the principle of expressability of John R. Searle: « whatever can be meant, can be said. » [12]. Of course, in our work, we will consider “to say” in a large sense.

2.1 KNOWLEDGE TRANSFER

In order to understand the IRP formulation process, it is normal to be interested in knowledge transfer. In summary, a minimal IRP formulation needs two actors: PCS and PEI. We modify the classical communication model and use this to describe some problem formulations that have to be solved:

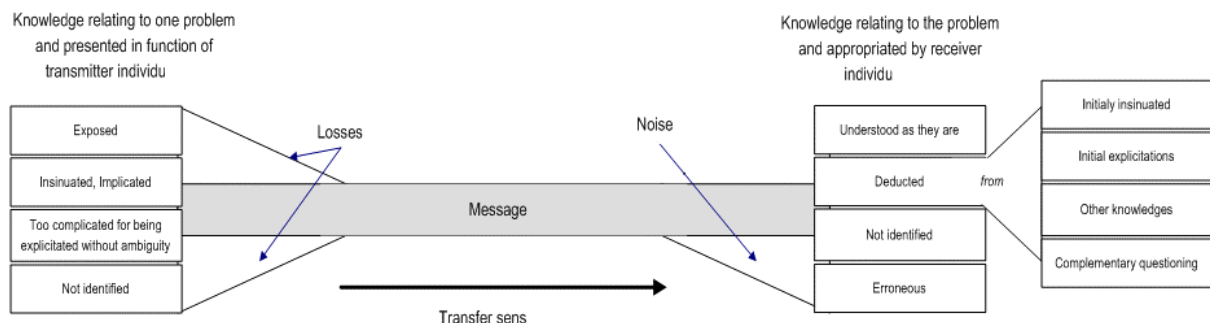


Figure 1. Knowledge transfer in the case of IRP formulation

¹ Annotation here means description

There are many variables that depend on the PEI cognitive environment (i.e. his competencies and knowledge), on the PCS cognitive environment, and also on the IRP environment itself. In addition, we are in fact faced with a classical case of knowledge transfer.

We should of course present the distinction between information and knowledge. That is why we refer to Dick Stenmark [15] who made a chart for this distinction according to different authors. We have adopted the approach which characterises knowledge as a set of experiences, values, insights and contextual information; while information is «a message meant to change the receiver’s perception» [3].

As shown in fig 1, we see that for PEI and PCS to understand each other, it is necessary for each of them to explain his knowledge and beliefs of the IRP. Consequently, we will note simply that the knowledge transfer process should be used to define the knowledge it conveys, that this process should obtain the adhesion of each party (particularly in order to validate the expressed knowledge), and that knowledge conveyed will be specific to the IRP context and the transfer context.

2.2 COOPERATIVE PRINCIPLE

Once we accept the basis above, we can focus on what is said and what is signified in the dialogue. Consequently, we can distinguish two senses for one phrase [13]: one sense is out of context and the other is within context. To expose a phrase that is out of context is almost impossible. We think that it is reasonable to accept the idea of conversational implications as proposed by HP Grice [7]. In fact, like him we suppose that the sense of one phrase depends on a set of circumstances that we can identify by an inference process. Also, we adopt the *cooperative principle* of H.P. Grice which treats certain non conventional implications linked to discourse. Indeed, in our case, we can imagine that to begin to solve an IRP, PEI and PCS should contribute as much as they can to have a relevant dialogue that corresponds to a *cooperative principle*. H.P. Grice divided his *cooperative principle* into four categories called maxims:

- **Maxim of Quantity:**
 - 1 Make your contribution as informative as is required (for the current purpose(s) of the exchange).
 - 2 Do not make your contribution more informative than is required. (...)
- **Maxim of Quality:**
 - 1 Do not say what you believe to be false.
 - 2 Do not say that for which you lack adequate evidence. (...)
- **Maxim of Relation:**
 - 1 Be relevant. (...)
- **Maxim of Manner:**
 - 1 Avoid obscurity of expression.
 - 2 Avoid ambiguity.
 - 3 Be brief (avoid unnecessary prolixity)
 - 4 Be orderly. (...)

2.3 CORRECTION DURING KNOWLEDGE TRANSFER

To correctly understand a knowledge process between two actors, we should consider also the system’s errors correction possibilities. In this case, we use the schematic diagram of a correction system by C. E. Shannon [14] to present this facet of IRP expression:

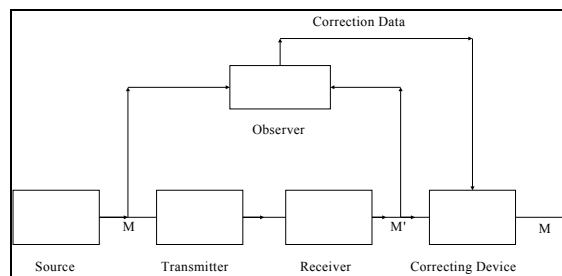


Figure 2. Schematic diagram of a correction system [14]

In this schema, the signal correction system (signal is the meaning of a message M) is perturbed by a noise during its transmission. But an auxiliary device (observer) helps the transmitter to limit the communication errors. In our approach, we suppose that during the transmission, there is only one transmitter and the receiver who is also the observer. Since human beings do not react as a simplified artificial system, and in virtue of the cooperative principle, we can only hope that they help themselves.

3. ADDITIONAL PRINCIPLES

3.1 TO VALIDATE THE KNOWLEDGE TRANSFER

Other principles can be added to the ones above. The first of these is the **Adhesion principle**:
The two principal actors, in a case of problem formulation have to mutually validate and aid themselves in their goal of understanding the problem.

This principle itself can be divided into two parts in form of maxims:

- **Additional shape maxim:**
Each principal actor must propose at least one different formulation of the problem for its improved understanding.
- **Consensus maxim:**
The principal actors of an IRP must validate each other's problem formulation.

The **additional shape maxim** allows the limitation of communication errors and the **consensus maxim** allows the validation of the problem formulation.

3.2 CASE OF EXPONIBLE PROPOSITIONS

We saw that Grice maxims can in a big measure describe many pragmatic implications of an IRP formulation. But sometime, an IRP formulation can be more informative than necessary for its interpretation. H.P Grice tried to minimise this category with the assistance of a new principle which is the **principle of modified Occam's razor**:
«*Senses are not to be multiplied beyond necessity*».

In addition, the classical philosophy has identified some other propositions. Port-Royal school [5] defined these propositions as: «*a statement which being grammatically simple includes in reality few different coordinate judgements. So it is necessary to expose it, i.e. to reveal the hidden propositions which constitutes its sense, if we desire to understand its signification and determine the valid reasoning which it can enter into*».

Thus, Grice principles did not treat clearly all the above propositions. We suppose that an additional principle to **maxim of manner** and **principle of modified Occam's razor** could help clarify the application of these particular propositions. We name this new maxim: **Limited inference maxim**:
1. *If a set of propositions is less ambiguous than only one proposition, use the set.*
2. *Use two sub-propositions if the use of only one, including two different sets of information presents an important risk for the interpretation to loose one of that information.*

This maxim considers one sentence or proposition as ambiguous, if in spite of the Grice principles being respected, the signification of the proposition is not rapidly easy to understand. Typical cases of this genre of propositions are propositions including often, an anaphoric pronoun or propositions including a conjunction or a sentence modifier. For example, if a PEI exposes an IRP to a PCS as:
«*We could find only few documents on subject S, consequently we need you for your qualities in information retrieval ...*».

This proposition can be declined in three sub-proposition:

- «*We could find a document about subject S*»,
- «*We think it must exist many more documents about subject S*»,
- «*We need you to find many other documents about subject S*».

(a) It erases the risk of loss of the signification of “*few documents*” for IRP treatment. (b) Prevents the risk of loss of the signification of “*only*”. (c) Prevents the risk of lost of the signification of “*we need you*”.

Of course, the *limited inference maxim* can be used during the IRP formulation dialogue phases, more especially in the debriefing and note taking phases.

3.3. COMPLEMENTARY HELP

It is trivial that to solve a problem we must understand it very well. Unfortunately as Y. Elissalde [6] remarked, all interpretation supposes three different things: the interpreter, what it is interpreted, the person to which it is interpreted. We suppose that the interpreter and the recipient, in their respective role, must propose the use “landmarks”. This use of “landmarks” helps the interpretation and consequently the validation of this interpretation. These are tools to necessary for identifying one problem representation or other knowledge representation between two persons. These “landmarks” express the point of view chosen to solve or pose a problem. They can be annotations, links, or every other signs, that one can imagine to help another person understand his point of view. Like with a compass card, these “landmarks” or simply marks exist to direct others to the objective we wish. We translate this assumption in the case of IRP resolution in an *orientation maxim*:

Choose some marks to help present your implicit point of views during your problem formulation.

We have now two additional maxims (limited inference and orientation) without a principle to handle them. What principle could include these maxims?

The use of these principles assures us of the capacity to express the IRP. The cooperative principle helps us with the discourse between PEI and PCS. The adhesion makes the IRP a consensus. But if we obtain this consensus, the principle wants us to give another IRP formulation. To help in this case we create another principle: the *Reformulation principle*:

All IRP complementary formulation should clarify the informational content about initial formulation without deforming it.

We can translate this principle into three maxims:

- *Orientation maxim*

- *Limited inference maxim*

- *Description maxim*:

1. *All formulation deduced from an initial formulation should minimize the part of implicit information contained in the initial formulation.*

2. *Don't reformulate an expression into a more complicated that makes it more difficult to understand.*

But if we should need to transcribe to a support, the result of the application of all these principles on an IRP, we should create a principle for this transcription: the *Memorization principle (or transcription principle)*

To memorize the IRP interpretation, you should keep the formulation identical to the initial formulation with the conditions of its expression.

We can also translate this principle to maxims:

- *Traceability maxim*:

1. *If we formulate one expression from another, add to it a trace that could help you follow easily the transition from the first to the second.*

2. *If, to expose one formulation, you create two other simpler and complementary formulations, then keep a trace of its composition.*

- *Proximity maxim*:

If one formulation can be deduced directly from another, these two formulations are neighbours; transcribe them with this same notion of nearness between them.

Now, we can suppose that we have every principle that can be used to express, validate and use an IRP interpretation. But we can also use these principles as the first steps to solve an IRP. This will be illustrated in the following section.

4. IRP SOLVING AND EXPRESSION PRINCIPLES

In this section we will show, with the aid of some examples, the principal utilities of the above formulation principles and the consequences if they are not respected. This reasoning by negation will be implemented using the model MIRABEL that is centred on IRP description and oriented towards the use of formulation that helps in understanding the IRP. In addition to this model, the user must implicitly make call to essential of principles expression. Unfortunately, even if MIRABEL is used, the ignorance of these principles do not always resolve the problems.

4.1 MIRABEL MODEL

The MIRABEL model was created to identify a set of templates to describe an IRP at different levels:

```
{decisional environment
  {problem to solve
    {works realised
      {information needs to be managed
        {temporal template
          {individuals implicated
            {decision maker}
            {intermediaries}
            {PCS}
            {recipient} }
          {IRP expression
            {complex expressions
              {simple expression
                {ask}
                {types of presentation
                  {contents}
                  {shape} }
                {subject
                  {key concept}}}}}}}}}}}}}}
```

Each element of this model can help structure an IRP and complete them with complementary description or serve as a basis for any definition of characteristics and parameters at the level desired.

4.2 EXAMPLES OF LACK OF RESPECT FOR EXPRESSION PRINCIPLES

4.2.1 LACK OF RESPECT OF *COOPERATIVE PRINCIPLE*

Situation:

PCS: “This is the synthesis of information that you ask me”

PEI: “Thank you, very much”

“Could you send it directly to X, my chief engineer, this is in fact for him”

Remarks:

The PEI did not respect the maxim of quantity. He did not say that he was an intermediary. The major risk is that the real recipient may not be completely satisfied. This particular risk is avoided by the use of the template recipient in MIRABEL model.

4.2.2 LACK OF RESPECT OF *ADHESION PRINCIPLE*

Situation:

PCS: “This is the study that you ask me”

PEI: “But it’s as big as a directory.

I would like something clear and light”

Remarks:

During their discussion, PEI and PCS do not agree on a particular type of result. This is why in the MIRABEL model we have made the distinction between contents and form for the types of result through the corresponding parameter.

N.B: The MIRABEL model does not completely erase this risk, even though the adhesion principle can only be used once MIRABEL is completed by the PCS. The PEI could judge then if the PCS interpretation is in adequacy with his.

4.2.3 LACK OF RESPECT OF REFORMULATION PRINCIPLE**Case 1: Situation t_i :**

PEI: *"I want to know more about the Canadian market of gas"*

Situation t_{i+n} :

PCS: *"This is the panel of information that you asked me"*

PEI: *"Thank you"*

PEI: *"But we already have this study of domestic fuel in Ontario and also this on evolution of gas consumption for Alberta since 1996 to 2000 too!"*

Remarks:

When the PCS took note of his discussion with PEI, he neglected the expression: "more" (see *limited inference maxim*), and also did not think that PEI already had some results on his IRP. The MIRABEL model can help to avoid this error if the adhesion principle is applied with the "work realised" parameter.

Case 2: Situation:

PCS: *"This is the synthesis of information that you ask for on the wood table"*

PEI: *"You didn't treat the question of tables in wood imitation?"*

Remarks:

The PCS and PEI did not define the real subject(s) to be treated for the resolution of the IRP and the *orientation maxim* was neglected.

4.2.4 NON ADHERENCE TO THE PRINCIPLE OF MODIFIED OCCAM'S RAZOR**Situation:**

PCS: *"This is the synthesis of information that you ask me on the civil planes market"*

PEI: *"Oh, you have treated civil Helicopters market too. It wasn't necessary. Only planes interest us"*

Remarks:

In this case, the PCS assimilated civil planes market to mean the market of civil helicopters and planes. Thus, he extrapolated the information need of the PEI. The MIRABEL model presents this information in the subject template, but it can't prevent this error.

4.2.4 LACK OF RESPECT OF MEMORIZATION PRINCIPLE**Situation:**

PCS A takes an old IRP by PCS B, because his recipient want the same treatment for his information needs as B did a year before.

A to B: *"Why did you work with paper technologies for this investigation? The subject was information on textile market?"*

B to A: *"I don't know. But if I remember the reason, I will let you know"*

Remarks:

In this case, the PCS B did not think of the possibility of reemploying the past result of his IRP. And now, without clue (traces) A will lose a lot of time to understand the old IRP result produced by B.

4.2 HIERARCHY OF IRP GOOD EXPRESSION PRINCIPLES

Now, we can propose an ordered hierarchy of our principles by their importance: *expressability, cooperation, adhesion, reformulation, modified Occam's razor, and memorization*. The first is applied automatically. The second is applied integrally, but each maxim which it contains can eventually be violated. The third is absolutely necessary to reduce the problem of inadequacy of the results produced in response to the IRP. We find this criterion in the application of elaboration of functional specification of a work. The fourth is a help necessary for the effectiveness of the *adhesion principle*. The last two principles permit the anticipation of some particular problems for fast as well as good interpretation and memorization of the IRP and its resolution.

CONCLUSION

We have seen that we can consider an IRP as a set of principles to which it is necessary to conform with. We started from one hypothesis: the *expressability principle* and the Grice's work. The cooperation principle could a fortiori help us define some rules that can guarantee the success of a discussion between two persons during an IRP formulation. We have completed it with four other principles. Finally we have shown some examples of erroneous responses to an IRP due to lack of respect of these principles. We have in parallel demonstrated the application of our MIRABEL model to aid in an IRP resolution. This model allows the implicit use of these principles for a lot of basic cases.

This work is continued with the elaboration of some criteria for a best definition of subjects identified with the MIRABEL model and according to a good IRP expression principle. In addition, another use of the principles exposed here will help in developing a user interface for information retrieval as an evaluated search engine. M. H. Sarnier and S. Carberry transposed cooperation principle to principle of usefulness [11] for man-machine dialogue. We hope to transform the set of knowledge shared between the actors using MIRABEL to help in the creation of a good IRP description. We also hope that the formulations collected using MIRABEL model will help improve an information retrieval system. We will then have a key element for a complete Economic Intelligent System that our team hope to develop [4].

REFERENCES

- [1] N. Bouaka and A. David, and O. Thiery. **Contribution to the understanding of explanatory factors for the decision-maker problem within the framework of economic intelligence**. In 6th World Multiconference on Systemics, Cybernetics and Informatics - SCI' 2002. (Orlando, Florida, USA). 2002. Information Systems development. vol XVIII.
- [2] J. Chu-Carroll, **Introducing a Multi-Dimensional User Model To Tailor Natural Language Generation**, Thesis, University of Waterloo, Ontario, Canada, 1991.
- [3] T. H. Davenport, **Information Ecology**. Oxford University Press, New York, 1997.
- [4] A. David and O. Thiery. **Application of "EQUA²te" Architecture in Economic Intelligence**. In **Information and Communication Technologies applied to Economic Intelligence - ICTEI'2002**. (Ibadan, Nigeria). 2002.
- [5] Oswald Ducrot : **Dire et ne pas dire : principes de sémantique linguistique**, 2^e édition, Paris éditions Hermann, 1985.
- [6] Y. Elissalde, **La critique de l'interprétation**, Paris Librairie Philosophique J. Vrin, 2000.
- [7] H. P. Grice, **Studies in the way of words**. Cambridge : MA, Harvard University Press, 1989.
- [8] P. Kislin et A. David, **De la caractérisation de l'espace-problème décisionnel à l'élaboration des éléments de solution en recherche d'information dans un contexte d'Intelligence Economique**, in IERA, Nancy, France, 2003.
- [9] D. Küpper and A. Kobsa, **Tailoring the Presentation of Plans to Users' Knowledge and Capabilities** – Proceedings of the 2003 German Conference on Artificial Intelligence, Hamburg, Germany, Springer Verlag, 2003. <http://link.springer.de>

- [10] J. Moeschler, **Théorie pragmatique et pragmatique conversationnelle**. Paris, A. Colin, 1996.
- [11] M.H. Sarnier and S. Carberry. **A new strategy for providing definitions in task-oriented dialogues**. In Proceedings of the International Conference on computational Linguistics, pages 567-572, Budapest, Hungary, 1988.
- [12] J. R. Searle, **Speech acts: An essay in the philosophy of language**. London: Cambridge University Press, 1969.
- [13] J. R. Searle, **Sens et expression : études de théorie des actes de langage**. Paris éditions de Minuit, 1982.
- [14] CE Shannon, **A Mathematical theory of communication**, The Bell System Technical Journal, vol 27, pp 379-423, 623-656, Juillet, 1948.
- [15] D. Stenmark, **Information vs. Knowledge: The Role of intranets in Knowledge Management**. In Proceedings of the 35th Hawaii International Conference on System Sciences, 2002.