

Codage des références et coréférences dans les DHM

Florence Bruneseaux
CRIN-CNRS&INRIA Lorraine
brunesea@loria.fr

Laurent Romary
CRIN-CNRS&INRIA Lorraine
romary@loria.fr

0. Introduction

Les avantages d'une normalisation des ressources textuelles en format électronique par l'utilisation de la TEI (Text Encoding Initiative) ont déjà été présentés dans de nombreux articles[1]. Cette application de la norme SGML propose en effet des directives pour le codage des textes en offrant plus de 500 éléments (et autant d'attributs) permettant de décrire un document. Nous voudrions nous intéresser ici à un type de document particulier, les dialogues homme-machine multimodaux (parole et geste). Un codage de base pour l'ensemble des dialogues, et qui peut être généralement réalisé automatiquement à partir d'une transcription initiale correcte, doit mettre en évidence un certain nombre d'informations parmi lesquelles le locuteur de chaque énoncé, les changements de tours de parole, les pauses[2]... A partir de ce codage stable et indépendant du type d'étude que l'on voudra réaliser, il serait souhaitable de mettre en évidence des phénomènes plus spécifiques au niveau du contenu. Parmi ceux-ci, on peut considérer le problème de la référence et plus généralement le problème de la relation pouvant exister entre différents types de syntagmes syntagmes (nominaux et verbaux). En effet, si on analyse un dialogue entre deux individus, il est important de pouvoir dire si un segment de discours réfère à un objet particulier et si son interprétation peut être faite directement ou si celle-ci dépend d'un autre segment. Nous parlerons de coréférence lorsque deux segments renvoient à un même objet dans une tâche[3]. Nous désirons faire ici quelques propositions concernant le codage de la référence, mais les choix que nous adopterons resteront malgré tout assez généraux afin de pouvoir être appliqués à une plus large catégorie de dialogues qui ne prennent pas nécessairement en compte le geste dans les transcriptions. C'est pourquoi nos propositions de codage seront illustrées à la fois par des exemples devenus classiques en linguistique et par des exemples issus d'expérimentations de type Magicien d'Oz dans le domaine des dialogues homme-machine. Après avoir présenté un codage de base pour la référence, nous montrerons que la sélection d'un référent peut poser un certain nombre de problèmes.

1. Un codage de base pour la référence

Si au cours d'une conversation l'énoncé "passe moi *le journal*" s'adresse à un individu précis, celui-ci devra être capable d'identifier l'objet désigné afin de pouvoir le transmettre à son interlocuteur. Le syntagme nominal "le journal" fait donc référence à un objet précis. Nous verrons ultérieurement que la référence ne se limite pas au domaine des objets mais concerne aussi entre autre les propriétés et les actions. Suivant les recommandations de la TEI, nous coderons la référence par l'utilisation de la balise <rs> et nous différencierons chacune d'elles, en utilisant un attribut "type". Si maintenant nous situons dans un DHM, et que nous désirons repérer automatiquement l'ensemble des objets cités au cours d'un dialogue, il ne sera pas suffisant de trouver l'ensemble des S.N. qui ont un sens lexical (descriptions définies ou indéfinies, S.N. démonstratifs...). Même s'ils sont - à la différence des pronoms - dotés d'une autonomie référentielle, cela ne nous indique pas s'il coréfèrent ou non à des S.N. déjà apparus au cours du dialogue. C'est pourquoi nous attribuerons une "key" à chacun des <rs> et indiquerons ensuite - grâce à une balise <link> - le lien qui existe entre deux objets[4]. Ainsi, nous pourrions à la fois marquer la référence et indiquer entre autre les liens de coréférence entre deux objets.

Ex.1 :

<u id="u1" who="Sujet"> Construire *la surface*
 <u id="u2" who="Compère"> Opération effectuée.
 <u id="u3" who="Sujet"> Modéliser *la surface*

<u id="u1" who="Sujet"> Construire <rs type="objet" key="O1">la surface</rs> </u>
 <u id="u2" who="Compère"> Opération effectuée. </u>
 <u id="u3" who="Sujet">Modéliser <rs type="objet" key="O2">la surface</rs> </u>
 <link type="coref" targets="O2 O1"></link>

Pour dire si les deux occurrences de "la surface" réfèrent au même objet ou non, on ne peut se limiter aux indications fournies par les énoncés. Il sera quelques fois nécessaire de prendre en compte ce qui apparaît à l'écran. Dans des exemples plus traditionnels, nous aurons plus facilement des emplois d'anaphores mais le lien entre un pronom objet et l'objet auquel il réfère pourra être indiqué de façon similaire. Reprenons le célèbre exemple de Brown & Yule (1983)[\[5\]](#):

Ex.2 :

Prenez *un poulet vif et bien gras*. Tuez-*le*, préparez-*le* pour le four, coupez-*le* en quatre et rôtissez-*le* avec du thym pendant une heure.

<s>Prenez <rs type="objet" key="O1">un poulet vif et bien gras</rs>. Tuez- <rs type="objet" key="O2">le </rs>, préparez- <rs type="objet" key="O3">le </rs>pour le four, coupez- <rs type="objet" key="O4">le </rs> en quatre et rôtissez- <rs type="objet" key="O5">le </rs> avec du thym pendant une heure.</s>
 <link type="coref" targets="O2 O1"></link>
 <link type="coref" targets="O3 O2"></link>
 <link type="coref" targets="O4 O3"></link>
 <link type="coref" targets="O5 O4"></link>

Dans "rôtissez-le", le pronom repéré par "key=O5" n'est pas seulement à mettre en relation avec l'objet désigné par "un poulet vif et bien gras". On sous-entend ici que le poulet a été tué, préparé, coupé. Ce choix permet de montrer l'évolution des objets au cours de l'énoncé. En effet, nous faisons référence à un objet tout en prenant en compte ses états successifs ; l'objet de départ pouvant toujours être récupéré (il correspond à la première occurrence, donc à la première "key" attribuée).

2. La complexité de la référence

Dans un dialogue multimodal, la référence à un objet ne se limite pas au repérage des anaphores et de leur antécédent (Prenez *un poulet*. Tuez-*le*). En effet, dans certains cas, un geste peut accompagner la parole. Il faudra alors par exemple pouvoir faire la différence entre des démonstratifs associés à une désignation et ceux qui ne le sont pas. Ainsi, un énoncé E tel que "... *la surface*. ...*Déplace cette surface*." sera analysé différemment selon qu'il est accompagné ou non d'un geste. Alors que dans le premier cas le recours à un antécédent ne s'impose pas pour définir l'objet référé, dans le second cas, l'objet a du être identifié précédemment.

Ex. 3:

<u id="u1" who="Sujet">Peut-on tracer la surface entre *la ligne que je désigne là* (+désignation) et *celle-ci* (+désignation)? </u>
 <u id="u2" who="Compère">... Quel nom dois-je donner à la surface ?</u>
 <u id="u3" who="Sujet">Diapir1</u>

Afin d'avancer dans la réflexion sur la possibilité de référence ou non entre deux objets, nous proposons d'utiliser une balise que nous nommons <univers> et qui désigne l'univers de référence dans lequel la référence sera faite. Dans l'exemple suivant, les trois énoncés sont à analyser dans un univers où trois ensembles distincts apparaissent : un ensemble d'objets (dans lequel se trouve la surface citée),

un ensemble de couleurs disponibles et un ensemble de propriétés possibles pour l'objet "surface" (*la couleur de la surface*).

Ex. 4 :

```
<u id="u1" who="Sujet"> Changer la couleur de la surface.  
<u id="u2" who="Compère"> Quelle couleur désirez-vous ?  
<u id="u3" who="Sujet"> Bleu.
```

Cet exemple met en évidence l'une des relations possibles entre N1 et N2 dans le syntagme [N1 de N2]. Il s'agit de la relation entre une propriété et un objet. Au niveau des syntagmes désignant des objets, nous avons vu le cas où le lien reliant deux objets était de type coréférence. Afin d'être exhaustif, la balise <link> devra permettre de décrire l'ensemble des liens possibles entre deux objets : relation entre une partie et un tout (*la mine du crayon*), entre un contenant et un contenu (*les points de la surface*)...

Les syntagmes nominaux qui nous ont permis d'illustrer la référence désignaient jusqu'ici des objets, or, le sens lexical de certains S.N. indique clairement qu'il ne s'agit plus d'une référence à des objets mais à des actions.

Ex. 5 :

```
<u id="u1" who="Sujet">Faire un zoom avant.  
<u id="u2" who="Compère"> Bien reçu. Opération effectuée. Le résultat vous convient-il ?  
<u id="u3" who="Sujet">Oui.  
  
<u id="u1" who="Sujet"><rs type="action" key="A1">Faire un zoom avant</rs>.</u>  
<u id="u2" who="Compère">  
<seg id="u2seg1"> Bien reçu.  
<seg id="u2seg2"> Opération effectuée.  
<seg id="u2seg3"> <ref type="action" key="R1" target="A1">Le résultat</ref> vous convient-il ?  
</u>  
<u id="u3" who="Sujet">Oui.</u>
```

Alors que dans cet exemple il est assez facile de faire la relation entre "le résultat" et la prédication précédente ("faire un zoom"), le problème se compliquera très vite avec des énoncés qui ne permettent pas de délimiter clairement la liste des actions à réitérer. Nous empruntons l'exemple suivant à Hill[6], lequel aborde le problème de "recommencer" avec l'exemple des instructions portées sur une bouteille de shampoing.

Ex. 6 :

"Pour de bons résultats, mouillez les cheveux à l'eau chaude.
Faites pénétrer doucement la première application.
Rincez abondamment et *recommencez*."

Lui aussi se demande à partir d'où il est nécessaire de recommencer ? Les cheveux sont déjà mouillés (on ne peut donc recommencer à partir de là). La première application a déjà été réalisée (on ne peut donc non plus la recommencer). Donnant ces arguments, il en déduit que le prédicat ne peut s'appliquer qu'à "rincer abondamment" et conclut que l'on obtient une boucle sans fin. Cependant, comme le remarque Falzon[7] (1986, p.26) " ce texte en langage naturel, en dépit des remarques de Hill, n'est certainement pas ambigu pour un être humain. " Au regard du contexte, seul le shampoing peut permettre de rendre propres des cheveux sales. De plus, certains éléments de l'énoncé (tel que le choix de l'adjectif "première") guident l'interprétation. Pour décider à quel ensemble d'actions un énoncé fait référence, il sera donc non seulement nécessaire de structurer les actions en groupes d'actions mais il faudra peut-être aussi de prendre en compte des éléments plus fins dans le dialogue et qui pourront guider le choix définitif.

3. Conclusion

Comme nous l'avons signalé déjà dans l'introduction, si une partie du codage peut être réalisé automatiquement, marquer la référence dans les dialogues demande un jugement de la part du codeur et doit donc être fait manuellement. Le souhait de Karttunen[8] de repérer automatiquement des individus d'un texte et les coréférences associées n'est pas encore réalisé.

La proposition que nous avons faite au niveau du codage se veut simple : toutes les chaînes de référence sont notées par la seule balise <rs>. Le lien existant entre deux <rs> est ensuite précisé dans une balise <link>. Nous distinguons les différents types de référence à l'intérieur de <rs> (ex. <rs type="objet/action...>) ainsi que les différents types de lien. L'idée la plus intéressante et qui nous a permis d'aller au-delà des cas simples (où l'antécédent d'un pronom est repérable dans un environnement textuel très proche), consiste à introduire une balise <univers> dans le codage. Tel qu'il a été défini ici, un univers a pour fonction de décrire la liste des individus ou objets accessibles au moment de l'analyse. Ainsi, si une référence (ou coréférence à un objet déjà nommé) est possible, le référent devra être puisé parmi la liste des candidats potentiels. Sans vouloir résoudre tous les cas d'ambiguïtés, ceci nous permettra au moins de proposer des candidats potentiels pour l'interprétation de certains groupes nominaux.

Les propositions faites dans ce document sont actuellement appliquées au codage de corpus de type dialogue homme-machine possédés au CRIN et en particulier au corpus GOCAD[9]. Cependant, noter par un codage spécifique les références dans un texte ne suffit pas s'il n'existe pas d'outil pour le manipuler. Pour répondre à ce besoin, le CRIN a développé un environnement *XCorpus*[10] permettant non seulement la manipulation de corpus textuels en SGML mais aussi l'édition et la gestion de corpus mono et multilingues afin d'effectuer entre autre des opérations d'alignement, de calculs statistiques, sur le lexique...

Notes

¹ Pour une introduction à la TEI, se référer aux *Cahiers de Gutenberg* n°24 juin 96 et pour plus d'informations, consulter Véronis J. et Ide N.(1995) *Text Encoding Initiative : Background and Context*, Kluwer Academic Publishers.

² Une balise minimale permettant d'introduire un énoncé sera de la forme :

<u id="u1" who="Sujet/Compère">... </u>

³ La notion de tâche est liée au fait que nous considérons des dialogues finalisés mais le codage est généralisable à d'autres types de dialogues.

⁴ La description des balises <rs> et <link> se trouvent respectivement dans la TEI aux paragraphes 6.4.1 et 14.1.1.

⁵ Brown G. & Yule G. (1983) *Discourse analysis*, Cambridge, Cambridge University Press.

⁶ Hill (1972) cité par Falzon (1986, p23)

⁷ Falzon P. (1986) *Langages opératifs et compréhension opérative*, Thèse, Paris 5.

⁸ Karttunen L. (1976) "Discourse referents" in McCawley J.D. (ed) *Syntax and Semantics 7 : Notes from the Linguistic Underground*. New-York, Academic Press, 363-385.

⁹ Corpus réalisé dans le domaine de la géologie sur la base du logiciel [Gocad](http://www.loria.fr/Projet/XCorpus) et dont certains exemples étaient extraits.

¹⁰ <http://www.loria.fr/Projet/XCorpus>