

Variation du contenu et de la forme dans la génération de descriptions d'itinéraires en métro

Lidia Fraczak[†], Guy Lapalme[‡] et Michael Zock[†]

[†] LIMSI-CNRS, BP 133, F-91403 Orsay Cedex, France, {lidia,zock}@limsi.fr

[‡] DIRO, Université de Montréal, CP 6128, Succ. Centre-Ville, Montréal
Québec H3C 3J7 Canada, lapalme@iro.umontreal.ca

Résumé

Se déplacer à pied ou par un autre moyen est une activité humaine courante. Cependant, trouver son chemin suppose souvent des aides de type verbal (descriptions d'itinéraires) ou iconique (croquis, cartes). Nous présentons dans cet article un système capable de produire des descriptions d'itinéraires en métro. Notre générateur est fondé sur un modèle cognitif de la production de telles descriptions et sur une analyse de corpus. Nous montrons comment des facteurs comme l'importance relative d'informations et les choix stylistiques peuvent faire varier à la fois le contenu et la forme de la description d'un itinéraire.

1. Introduction

La description d'itinéraires est d'un intérêt à la fois théorique, notamment pour les Sciences Cognitives, et pratique, pour la construction d'outils d'aide à la navigation (Denis, 1994). Plusieurs auteurs travaillant sur la génération de descriptions d'itinéraires se sont intéressés aux aspects cognitifs (Carstensen, 1992; Maaß, 1993; Herzog *et al.*, 1993; Gryl & Ligozat, 1995). Nous nous limiterons ici à la génération automatique de descriptions d'itinéraires en métro. En particulier, nous nous intéresserons aux variations du contenu et de la forme linguistique. C'est surtout par rapport à ces aspects que notre travail se distingue des autres travaux portant sur la description d'itinéraires.

Le contenu et la forme d'une description d'itinéraire dépendent avant tout de la nature des données : le type d'environnement et l'itinéraire particulier à décrire. Cependant, comme révèle l'analyse de notre matériel empirique, de grandes variations sont possibles, tant au niveau du contenu qu'au niveau de la forme, ce qui fait que de nombreuses descriptions peuvent être produites pour un itinéraire donné. Afin de pouvoir simuler cet éventail de variations, nous avons essayé d'isoler les facteurs pertinents. Deux d'entre eux semblent particulièrement importants : l'aspect pragmatique qui consiste à tenir compte de l'importance relative d'informations et l'aspect stylistique. On notera que ce premier aspect ne se résume pas à un couple de termes binaires (important vs non important), car une information peut être "plus ou moins" importante. Ainsi, l'importance relative d'une information ne détermine pas seulement son inclusion / exclusion au

sein du texte, mais également sa forme linguistique en cas d'inclusion. Quant aux facteurs stylistiques, ils impliquent à la fois le degré d'explicitation des informations (explicite vs implicite) et leur réalisation linguistique.

2. Modèle cognitif général

La production de descriptions d'itinéraires peut être décomposée en deux étapes : la *détermination* de l'itinéraire et la *formulation de sa description* (Carstensen, 1992; Maaß, 1993; Denis, 1994). La première tâche n'appartient pas au domaine du discours puisqu'elle peut être mise en œuvre pour des fins non-verbales (ex. dessiner un plan), ou même non-communicatives (effectuer soi-même un déplacement). La deuxième tâche, en revanche, est de nature discursive, car elle vise la production d'un texte.

La *détermination* de l'itinéraire consiste à identifier un parcours dans un espace, en utilisant d'une part des connaissances sur l'environnement (*connaissances référentielles*), et d'autre part des connaissances liées aux besoins de l'interlocuteur (*connaissances pragmatiques*). Le résultat de ce processus est une *représentation référentielle*. Celle-ci constitue le point de départ du processus de la *description* de l'itinéraire. Il est fondé sur des *connaissances discursives* relatives au domaine (aspects "conceptuels", "textuels" et "linguistiques") et sur d'autres *connaissances pragmatiques* concernant l'interlocuteur.

C'est notamment la partie "discursive" du traitement qui nous intéresse ici, c'est-à-dire la *description* de l'itinéraire. Elle consiste à "transformer" la représentation référentielle de l'itinéraire en sa *représentation discursive*. Ce processus se divise essentiellement en trois parties : *conceptuelle*, *textuelle* et *linguistique*. La structuration conceptuelle consiste à "découper" la représentation référentielle en unités verbalisables (informations). La structuration textuelle implique le choix des parties du contenu conceptuel à expliciter dans le texte. Enfin, la réalisation linguistique détermine la forme de surface. Notre modélisation des principes conceptuelles, textuelles et linguistiques – principes faisant partie des connaissances discursives – est fondée sur l'analyse de descriptions d'itinéraires en ville destinées à des piétons. Cette modélisation nous a servi ensuite de trame pour analyser des exemples de descriptions d'itinéraires en métro en vue de la génération automatique.

3. Analyse et modélisation de descriptions d'itinéraires en métro

Notre corpus est constitué d'une trentaine de descriptions d'itinéraires en métro parisien. Il a été recueilli par courrier électronique auprès de dix sujets qui décrivaient chacun trois itinéraires. Le corpus contient donc dix descriptions de trois itinéraires variant en termes de longueur et de complexité. Le premier itinéraire (Saint-Lazare – Jussieu) est le plus long (9 stations) et contient un changement. Le deuxième (Denfert-Rochereau – Place d'Italie) est de 4 stations et c'est le plus simple car il n'implique pas de changements. Enfin, le troisième itinéraire (Jussieu – Gare de Lyon) est de même longueur que le précédent (4 stations), mais c'est le plus complexe car il contient deux changements.

Notre analyse du corpus a porté sur trois aspects : le *contenu conceptuel*, le *contenu textuel* et la *forme linguistique*. Nous en présentons les résultats dans les trois sections suivantes. La distinction entre le *contenu conceptuel* et le *contenu textuel* n'est habituellement pas faite dans les modèles de génération de textes dont nous avons connaissance, dans lesquels la détermination du *contenu* ("quoi dire") est opposée à la détermination de la *forme linguistique* ("comment

dire”). D’après nous, le contenu *conceptuel* concerne l’ensemble des messages “à transmettre”, tandis que le contenu *textuel* concerne les informations “à exprimer” réellement dans le texte (les informations explicites par opposition aux informations implicites).

3.1. Contenu conceptuel

Un itinéraire est structuré en *segments* et en *relais*. Un *segment* est un fragment d’itinéraire qui se caractérise par un même “chemin” et par une suite de “lieux” différents, tandis qu’un *relais* correspond à un lieu qui coïncide avec un changement de chemin. La description de l’exemple 1 ci-dessous reflète la structuration de l’itinéraire en deux segments et un relais, comme illustré dans la figure 1.

Ex. 1 À Saint-Lazare, prendre la direction Gallieni. Descendre à Opéra (deux stations plus loin). Prendre alors la direction Mairie d’Ivry/Villejuif jusqu’à Jussieu (7-ème station).

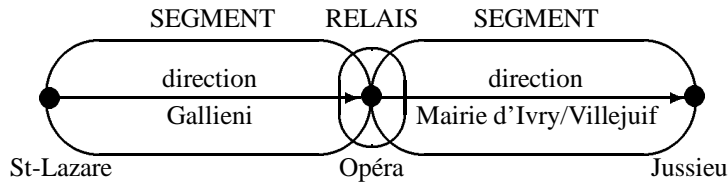


FIG. 1 - Structuration conceptuelle “globale” de l’itinéraire décrit dans l’exemple 1.

L’analyse de descriptions d’itinéraires en ville nous a permis de constater qu’un même itinéraire peut être structuré de différentes façons car la détermination d’un “chemin” peut se faire selon un ou plusieurs critères, comme la direction, le type ou le nom de la voie, des caractéristiques de l’environnement immédiat, etc. Cependant, dans le cas du métro, les principes concernant la structuration conceptuelle d’itinéraires sont plus stricts. Cela est dû au fait qu’en métro les “chemins” sont toujours définis par la direction de la ligne à emprunter. Ainsi, un *segment* correspond à un fragment d’itinéraire avec la même direction et un *relais* à un lieu de changement de direction, c’est-à-dire à une station de correspondance.

Au sein des segments et relais, nous distinguons quatre types d’étapes. Dans le tableau 1, nous énumérons ces étapes, et nous en donnons des exemples tirés du corpus. Comme nous le montrerons plus loin, toutes ces étapes n’ont pas nécessairement à être explicitées, du fait que les principes de la structuration conceptuelle d’itinéraires font partie intégrante des connaissances partagées par les interlocuteurs. La représentation conceptuelle de l’ensemble des étapes consécutives est néanmoins nécessaire, aussi bien pour le locuteur afin qu’il puisse produire une description, que pour l’interlocuteur afin qu’il puisse la comprendre.

Étape	Exemple
début-segment	<i>prendre la direction Gallieni</i>
transfert-segment	<i>il y a deux stations</i>
fin-segment	<i>descendre à Opéra</i>
transfert-relais	<i>changer à Opéra</i>

TAB. 1 - Étapes d’itinéraire et descriptions correspondantes.

La structuration conceptuelle concerne également la détermination de *repères*. Nous en distinguons deux principaux types : les *repères-simples* et les *repères-chemins*. Les premiers sont associés aux relais, tandis que les seconds sont associés aux segments. Dans le cas des descriptions

d’itinéraires en métro, le premier type de repères correspond aux stations, et le deuxième type correspond aux sections de lignes de métro allant dans une même direction. Nous définissons les *repères* en termes de couples attribut-valeur. Nous avons spécifié les attributs pour chaque type de repères, comme montré dans le tableau 2.

Repère	Attribut	Valeur	Statut	Exemple textuel
repère-chemin	entité	“ligne”	obligatoire	“prendre la <i>ligne</i> 3 ”
	nom	nom de la ligne	optionnel	“prendre la <i>ligne</i> 3”
	direction	direction de la ligne	obligatoire	“prendre la <i>direction Gallieni</i> ”
	dimension	nombre de stations	optionnel	“il y a <i>deux</i> stations”
repère-simple	entité	“station”	obligatoire	“descendre à la <i>station Opéra</i> ”
	nom	nom de la station	obligatoire	“descendre à la <i>station Opéra</i> ”
	ordre	ordre de la station	optionnel	“descendre à la <i>deuxième</i> station”

TAB. 2 - Représentation de repères.

L’analyse du corpus nous a également permis de déterminer les attributs obligatoires et optionnels (voir tableau 2). Nous considérons comme “obligatoires” les attributs apparaissant dans toutes les descriptions du corpus : *direction* de repère-chemin et *nom* de repère-simple (nom d’une station de correspondance). Bien qu’il n’apparaisse pas dans toutes les descriptions, nous considérons également l’attribut *entité* pour les deux types de repères comme obligatoire car ses valeurs sont stables dans le contexte du métro (“ligne” pour un repère-chemin et “station” pour un repère-simple). Par ailleurs, ces informations sont connues de l’interlocuteur, si bien qu’il n’est pas nécessaire de les expliciter. Les attributs optionnels sont ceux qui n’apparaissent pas dans toutes les descriptions du corpus : *nom* et *dimension* de repère-chemin et *ordre* de repère-simple.

Les attributs obligatoires doivent être représentés au niveau conceptuel. Ils font donc partie intégrante du contenu conceptuel des descriptions, même quand ils ne sont pas explicités dans le texte (c’est le cas de l’attribut *entité*). Quant aux attributs optionnels, leur représentation conceptuelle et leur forme linguistique dépendent de l’importance relative que leur donne le locuteur. La valeur informationnelle peut être définie en fonction d’indices linguistiques. Selon Combettes (Combettes, 1992), les propositions principales relèvent du “premier plan” du texte (informations primaires), tandis que les structures subordonnées relèvent du “second plan” (informations secondaires). En ce qui concerne l’importance relative des informations exprimées, nous pensons que les structures subordonnées peuvent avoir un statut différent. En effet, l’analyse du corpus nous amène à croire que, par exemple, une proposition relative indique une importance plus grande de l’information exprimée qu’une apposition entre parenthèses.

Pour les attributs optionnels *dimension* et *ordre*, qui véhiculent une même valeur informationnelle (ex. 3 pour “trois stations” et pour “troisième station”), nous avons pu dégager des facteurs “contextuels” qui influencent leur importance relative. Il s’agit notamment du nombre de stations dans un segment (distance partielle) et de la partie de l’itinéraire. Nous avons remarqué qu’une plus grande importance est accordée à l’information concernant la distance s’il n’y a qu’une seule station et/ou s’il s’agit du dernier segment de l’itinéraire. L’importance relative du troisième attribut optionnel : *nom* de repère-chemin (nom d’une ligne) n’est pas affectée par ce type de différences ; l’importance relative sera la même pour tous les noms des lignes pertinentes pour la description d’un itinéraire.

À la suite de l’analyse du corpus, et en vue de la génération automatique, nous avons défini quatre stratégies d’attribution d’importance pour les attributs *dimension* et *ordre* et trois stra-

tégies pour l'attribut *nom* de la ligne, ainsi que les formes linguistiques correspondantes. Ces résultats sont présentés dans les tableaux 5 et 6 plus loin.

3.2. Contenu textuel

La détermination du *contenu textuel* consiste à sélectionner, à partir du contenu conceptuel, les *étapes* d'itinéraire et les *attributs de repères* à expliciter dans le texte. Comme nous l'avons déjà signalé, toutes les *étapes* ne sont pas nécessairement exprimées dans la description, surtout lorsqu'il s'agit d'un itinéraire composé de plusieurs segments. L'étape la moins souvent mentionnée dans les descriptions d'itinéraires en métro est le *transfert-segment*. En effet, à la différence des descriptions d'itinéraires en ville, on ne rencontre pas ici d'expressions comme : "continuer" ou "aller tout droit". Cela est dû à la situation du voyageur de transport en commun qui, une fois embarqué, n'est plus tout à fait maître de son déplacement. Cependant, il faut noter que même si les verbes d'action n'apparaissent pas dans notre corpus pour exprimer un *transfert-segment*, ce dernier est parfois mentionné en termes de distance : "il y a *n* stations" (expression de l'attribut *dimension* de repère-chemin). En ce qui concerne les autres *étapes*, il s'avère que le *début-segment* est toujours explicité. Ceci peut être expliqué par le fait que le type d'expression lui correspondant exprime également l'attribut obligatoire de repère-chemin : *direction* (ex. "prendre direction Gallieni"). En revanche, l'expression des autres *étapes* est variable. Afin d'illustrer ce point, nous analysons dans le tableau 3 les exemples 2 et 3 ci-dessous référant à un même itinéraire.

Ex. 2 *Prendre direction Gallieni (ligne 3). Sortir à Opéra (2 stations). Prendre direction Mairie d'Ivry/Villejuif (ligne 7). Descendre à Jussieu (7-ème station).*

Ex. 3 *À partir de Saint-Lazare, prendre la ligne 3 en direction de Gallieni. S'arrêter à la station Opéra et changer pour prendre la ligne 7 en direction de Mairie d'Ivry. La station Jussieu est la 7-ème après Opéra.*

Étapes	Exemple 2	Exemple 3
début-segment	<i>prendre direction Gallieni</i>	<i>prendre la ligne 3 en direction...</i>
transfert-segment		
fin-segment	<i>sortir à Opéra</i>	<i>s'arrêter à la station Opéra</i>
transfert-relais		<i>changer</i>
début-segment	<i>prendre direction Mairie d'Ivry</i>	<i>prendre la ligne 7 en direction...</i>
transfert-segment		
fin-segment	<i>descendre à Jussieu</i>	

TAB. 3 - Analyse des exemples 2 et 3 par rapport à l'expression des étapes.

Dans l'exemple 2, les mêmes étapes sont exprimées pour les deux segments de l'itinéraire : *début-segment* et *fin-segment*, tandis que l'étape *transfert-relais* n'est pas exprimée. L'exemple 3 est différent : pour le premier segment on mentionne les étapes *début-segment* et *fin-segment*, on mentionne aussi l'étape *transfert-relais*, et pour le deuxième segment seulement l'étape *début-segment* est exprimée. La dernière phrase de ce deuxième exemple n'est pas représentée dans le tableau car elle n'évoque pas d'étape. L'analyse du contenu textuel des descriptions en termes d'*étapes* nous a permis de spécifier des schémas textuels, appelés *blocs*, utilisés par notre générateur.

Comme déjà mentionné, l’analyse du contenu textuel du corpus traite également du problème de l’explicite/implicite de certains attributs de *repères*. Il s’agit des attributs “obligatoires” aux valeurs stables, c’est-à-dire de l’attribut *entité* de repère-chemin (valeur : “ligne”) et de l’attribut *entité* de repère-simple (valeur : “station”). Concernant l’entité “ligne”, nous avons constaté qu’elle est mentionnée explicitement uniquement pour introduire l’attribut *nom* de repère-chemin (nom d’une ligne). Ainsi, nous trouvons des énoncés comme : “prendre *la ligne 3* en direction de Gallieni”, mais non : “prendre *la ligne* en direction de Gallieni”. Quant à l’entité “station”, son expression n’est pas nécessaire pour introduire l’attribut *nom* de repère-simple (nom d’une station) ; nous trouvons aussi bien “descendre à *Opéra*” que “descendre à la *station Opéra*”. Il apparaît que la décision d’expliciter ou pas l’information concernant l’entité “station” s’applique généralement à toute la description, c’est-à-dire à toutes les stations mentionnées.

3.3. *Forme linguistique*

En ce qui concerne la *forme linguistique*, nous distinguons deux unités de base : *séquences* et *connexions*. Les premières ont pour fonction d’exprimer des informations concernant les étapes et les repères, tandis que les secondes servent à “connecter” les *séquences* et les *blocs* de séquences avec la structure conceptuelle de l’itinéraire. Les *séquences* sont divisées en *noyaux* et *ajouts*. Dans le tableau 4, nous donnons un exemple d’une description d’itinéraire analysée en termes d’unités linguistiques.

Connexion	Séquence	
	Noyau	Ajout
<i>à partir de Jussieu</i>	<i>prendre la ligne 10</i>	
<i>pour</i>	<i>aller à Gare d’Austerlitz</i>	<i>qui est juste la station suivante dans ce sens</i>
<i>à Gare d’Austerlitz</i>	<i>changer</i>	
<i>afin de</i>	<i>prendre la ligne 5 en direction de Bobigny</i>	
	<i>descendre à Bastille</i>	<i>2 stations plus loin</i>
<i>enfin</i>	<i>prendre la ligne 1 en direction de Château de Vincennes</i>	
<i>et</i>	<i>descendre à la station suivante</i>	<i>c’est la gare de Lyon</i>

TAB. 4 - *Analyse d’une description d’itinéraire en termes d’unités linguistiques.*

Les *noyaux* sont réalisés par des propositions indépendantes. Elles expriment les informations obligatoires et des informations optionnelles auxquelles on attribue une plus grande importance. Les *ajouts* correspondent à des structures “subordonnées” (relatives, appositions, propositions “anaphoriques”). Elles expriment des informations additionnelles concernant les *repères* : attributs de moindre importance (généralement des attributs optionnels). Dans les tableaux 5 et 6 plus bas, nous présentons les règles induites à partir du corpus qui sont utilisées par notre générateur de descriptions d’itinéraires. On y trouve une spécification des stratégies concernant l’importance relative des attributs optionnels en relation avec les formes linguistiques.

4. Génération de descriptions d’itinéraires en métro

Le générateur est programmé en GNU Emacs Lisp. Son architecture est fondée sur notre modèle cognitif de la production de descriptions d’itinéraires et sur les résultats de l’analyse du

Stratégie	Condition d'inclusion	Importance	Forme linguistique
0	pour aucun cas	aucune	aucune
1	si seulement une station	1	ajout : apposition entre parenthèses
2	si seulement une station si le dernier segment	2 1	ajout : proposition relative ou anaphorique ajout : apposition entre parenthèses
3	si seulement une station si le dernier segment pour tous les autres cas	3 2 1	séquence : proposition indépendante ajout : proposition relative ou anaphorique ou expression "il y a n stations" ajout : apposition entre parenthèses

TAB. 5 - *Stratégies concernant l'importance relative des attributs dimension et ordre et leurs formes linguistiques correspondantes.*

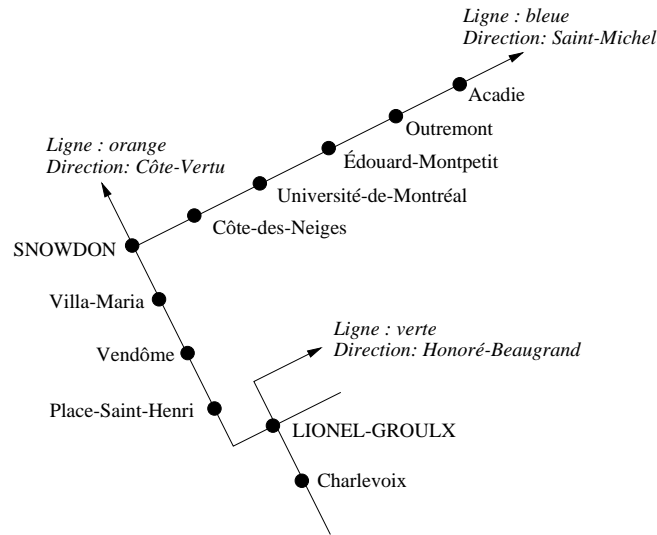
Stratégie	Condition d'inclusion	Importance	Forme linguistique
0	pour aucun cas	aucune	aucune
1	pour tous les cas	1	ajout : apposition entre parenthèses
2	pour tous les cas	2	séquence : proposition indépendante

TAB. 6 - *Stratégies concernant l'importance relative de l'attribut nom de ligne et ses formes linguistiques correspondantes.*

corpus que nous venons de décrire. Le générateur est composé de deux modules : (1) un module *référentiel* qui détermine les itinéraires, et (2) un module *discursif* qui engendre les descriptions correspondantes. Pour valider notre modèle, nous avons utilisé les données du métro de Montréal. Le fonctionnement du générateur décrit ici est illustré dans la figure 2 pour l'itinéraire "Charlevoix – Acadie" ayant deux changements.

La sortie du module référentiel et l'entrée du module discursif est constituée par la *représentation référentielle* de l'itinéraire (voir figure 2). Cette représentation est une liste de nœuds qui correspondent à des quais, encodés par des couples dont le premier élément indique la station et le deuxième la direction.

Le module discursif est divisé en deux parties. La première est la partie *conceptuelle* dont le rôle est de préparer les informations à transmettre. Elle est réalisée par deux fonctions principales. La première prend en entrée la *représentation référentielle*, pour donner en sortie une première *représentation conceptuelle* symbolisant la structuration de l'itinéraire en unités "globales" (segments et relais) et "locales" (étapes). La deuxième fonction prend en entrée la représentation précédente, les données référentielles pertinentes, ainsi que deux paramètres spécifiant les "stratégies d'importance relative" des attributs optionnels : *dimension* de repère-chemin, *ordre* de repère-simple (premier paramètre, valeurs possibles : 0-3) et *nom* de repère-chemin (deuxième paramètre, valeurs possibles : 0-2). Ces paramètres sont responsables des variations du contenu conceptuel et de la forme linguistique. Dans notre exemple (figure 2), ces valeurs sont fixées respectivement à 1 et à 2. La fonction produit en sortie, sous forme de couples attribut-valeur, la deuxième *représentation conceptuelle*, concernant les repères. Sont représentés ici tous les attributs obligatoires des repères ainsi que leurs attributs optionnels (*dimension*, *ordre* et/ou *nom* de ligne) selon les valeurs de leurs paramètres d'importance. Si la valeur du paramètre d'importance est 0, l'attribut concerné n'est pas représenté. Si cette valeur est supérieure à 0, l'attribut et son importance relative sont représentés. Ce dernier type d'information est encodé par les



Représentation référentielle :

```
((Charlevoix Honore-Beaugrand) (Lionel-Groulx Honore-Beaugrand)
(Lionel-Groulx Cote-Vertu) (Place-Saint-Henri Cote-Vertu) (Vendome Cote-Vertu)
(Villa-Maria Cote-Vertu) (Snowdon Cote-Vertu) (Snowdon Saint-Michel)
(Cote-des-Neiges Saint-Michel) (Universite-de-Montreal Saint-Michel)
(Edouard-Montpetit Saint-Michel) (Outremont Saint-Michel) (Acadie Saint-Michel))
```

Représentation conceptuelle - structuration de l'itinéraire :

```
(( (Charlevoix Honore-Beaugrand) nil (Lionel-Groulx Honore-Beaugrand))
((Lionel-Groulx Cote-Vertu)
((Place-Saint-Henri Cote-Vertu) (Vendome Cote-Vertu) (Villa-Maria Cote-Vertu))
(Snowdon Cote-Vertu))
((Snowdon Saint-Michel)
((Cote-des-Neiges Saint-Michel) (Universite-de-Montreal Saint-Michel)
(Edouard-Montpetit Saint-Michel) (Outremont Saint-Michel))
(Acadie Saint-Michel)))
```

Représentation conceptuelle - repères de l'itinéraire (paramètres d'importance en entrées : pour attributs *dimension* et *ordre* = 1, pour attribut *nom* de ligne = 2) :

```
(( (entite . ligne) (direction . "Honoré-Beaugrand") (nom . "verte")
(degre-nom . 2) (dimension . 1) (degre-dim . 1))
((entite . station) (nom . "Lionel-Groulx") (ordre . 1) (degre-ordre . 1))
((entite . ligne) (direction . "Côte-Vertu") (nom . "orange") (degre-nom . 2))
((entite . station) (nom . "Snowdon"))
((entite . ligne) (direction . "Saint-Michel") (nom . "bleue") (degre-nom . 2))
((entite . station) (nom . "Acadie")))
```

Représentation discursive :

"À partir de Charlevoix, prendre la ligne verte en direction de Honoré-Beaugrand jusqu'à Lionel-Groulx (station suivante). Puis, changer pour prendre la ligne orange en direction de Côte-Vertu et descendre à Snowdon. Enfin, prendre la ligne bleue en direction de Saint-Michel et sortir à Acadie."

FIG. 2 - Illustration du traitement pour l'itinéraire "Charlevoix – Acadie".

attributs : *degré-dim*, *degré-ordre* et *degré-nom*. Dans la représentation des repères donnée en exemple (figure 2), les attributs *dimension* et *ordre* sont représentés uniquement pour la distance d'une station (donc seulement pour le repère-chemin et le repère-simple du premier segment) car l'importance spécifiée dans le paramètre en entrée est faible (valeur 1). Le paramètre d'importance pour l'attribut *nom* de repère-chemin a la valeur 2 (importance forte) donc cet attribut est représenté en sortie.

La deuxième partie du module discursif est la partie *linguistique*. Elle utilise un dictionnaire de formes verbales et nominales employées dans les descriptions d'itinéraires en métro. La réalisation linguistique est basée sur des fonctions qui encodent les schémas de *séquences* variant en fonction des types d'étapes à décrire. Ces fonctions prennent en entrée les informations concernant les attributs de repères pertinents, ainsi que la forme verbale à utiliser (ex. *inf*: “descendre”, *prés-2*: “tu descends”). La version de schéma d'une séquence est choisie en fonction des attributs “spéciaux” : *degré-dim*, *degré-ordre* et *degré-nom* qui encodent le degré d'importance des attributs optionnels, en accord avec les règles données dans les tableaux 5 et 6. Ainsi, dans le texte de notre exemple (figure 2), la première *séquence* contient un *ajout* sous forme d'apposition entre parenthèses exprimant la valeur de l'attribut *ordre* (“station suivante”). S'agissant de la même valeur informationnelle que pour l'attribut *dimension*, ce dernier n'est pas exprimé (bien qu'il aurait pu l'être à la place de l'attribut *ordre* : “une station”, ce choix étant aléatoire). Puisque les attributs *dimension* et *ordre* des autres repères ne sont pas contenus dans la représentation conceptuelle, ils ne sont pas exprimés dans les *séquences* décrivant ces repères. Les attributs *nom* des repères-chemins (ex. “ligne verte”) sont exprimés dans les *noyaux* des séquences (propositions principales) du fait de leur importance élevée (valeur 2).

Le module discursif contient aussi une partie *textuelle*. Une des fonctions de cette partie encode les schémas de *blocs* textuels qui sont des combinaisons des schémas des *séquences*. Deux autres fonctions encodent respectivement des schémas de *connexions* et des schémas de combinaisons possibles de *blocs* au sein d'une description, ces combinaisons variant en fonction du nombre de segments de l'itinéraire. La description dans la figure 2 contient trois *blocs* textuels. Le premier et le troisième *blocs* utilisent le même schéma (avec des variations stylistiques “locales”) qui consiste à exprimer les étapes *début-segment* et *fin-segment* (pour le premier *bloc* : “prendre la ligne verte... jusqu'à Lionel-Groulx” ; pour le troisième “prendre la ligne bleue... et sortir à Acadie”). Le deuxième *bloc* est engendré à partir d'un autre schéma qui consiste à exprimer les étapes *transfert-relais*, *début-segment* et *fin-segment* (“changer... prendre la ligne orange... descendre à Snowdon”). Le schéma de *connexions* utilisé pour “connecter” les *blocs* est le suivant : “à partir de (station), ... puis, ... enfin, ...”.

Afin d'illustrer la façon dont le contenu conceptuel, le contenu textuel et la forme linguistique peuvent varier pour le même itinéraire, nous donnons ci-dessous deux autres exemples des descriptions d'itinéraires engendrés automatiquement par notre générateur.

De Charlevoix, prendre la direction Honoré-Beaugrand et changer à Lionel-Groulx. À Lionel-Groulx, prendre la direction Côte-Vertu jusqu'à Snowdon. À Snowdon, changer pour prendre la direction Saint-Michel et descendre à Acadie.

D'abord, prendre la direction Honoré-Beaugrand (ligne verte) et s'arrêter à la station suivante qui est Lionel-Groulx. Ensuite, changer pour prendre la direction Côte-Vertu (ligne orange) et descendre à Snowdon (4-ième station). Finalement, prendre la direction Saint-Michel (ligne bleue), il y a 5 stations et on arrive à Acadie.

La première description ne contient pas d'informations concernant les attributs *dimension* et *ordre* ni ceux de *nom* de lignes car les deux paramètres indiquant l'importance relative de ces informations ont été fixés à 0. Le schéma des *blocs* textuels utilisé prévoit la mention explicite des étapes suivantes : *début-segment* et *transfert-relais* (premier bloc) ; *début-segment* et *fin-segment* (deuxième bloc) ; *transfert-relais*, *début-segment* et *fin-segment* (troisième bloc). Le schéma de *connexions* entre les *blocs* est le suivant : “de (station), ... à (station), ..., à (station), ...”.

L'importance relative des attributs *dimension* et *ordre* est maximale dans le deuxième texte (valeur 3) ; en conséquence, le nombre de stations est donné pour tous les cas. L'expression linguistique de cette information varie cependant pour chaque segment selon les règles définies dans les tableaux 5 et 6, c'est-à-dire selon qu'il s'agit d'une seule station, du dernier segment ou autre. Le paramètre spécifiant l'importance relative de l'attribut *nom* de ligne a été fixé à 1 (importance faible). Par conséquent, cette information est exprimée par des *ajouts* du type appositions entre parenthèses. Le schéma de *blocs* textuels utilisé ici exprime les *étapes* suivantes : *début-segment* et *fin-segment* (premier bloc) ; *transfert-relais*, *début-segment* et *fin-segment* (deuxième bloc) ; *début-segment*, *transfert-relais* et *fin-segment* (troisième bloc). Le schéma de *connexions* entre les *blocs* utilisé est : “d'abord, ... ensuite,... finalement, ...”.

5. Conclusion et perspectives

Nous avons présenté la mise en œuvre d'un générateur de descriptions d'itinéraires en métro, fondé sur un modèle cognitif de la production de descriptions d'itinéraires et sur une analyse de corpus. Des facteurs d'importance relative d'informations et des facteurs stylistiques, observés dans le corpus, nous permettent de mieux contrôler le contenu conceptuel, le contenu textuel et la forme linguistique pour un itinéraire donné. Les descriptions engendrées automatiquement sont comparables à celles produites par les sujets.

Notre générateur tourne actuellement sur deux ensembles de données, celui du métro de Montréal et celui du métro parisien. Nous envisageons d'élargir la portée du générateur à d'autres types de transport en commun et à la navigation routière.

Références

- CARSTENSEN K.-U. (1992). Finding adequate routes for the generation of route descriptions. In G. GÖRZ, Ed., *Proc. of KONVENS'92 (Konferenz "Verarbeitung natürlicher Sprache")*, p. 309–318.
- COMBETTES B. (1992). *L'organisation du texte*. Didactique des textes. Université de Metz.
- DENIS M. (1994). *La description d'itinéraires : Des repères pour des actions*. Notes et documents du LIMSI 94-14, LIMSI, Orsay, France.
- GRYL A. & LIGOZAT G. (1995). Route descriptions: a stratified approach. In *IJCAI'95 Workshop on spatial and temporal reasoning*, p. 57–64, Montreal, Canada.
- HERZOG G., MAASS W. & WAZINSKI P. (1993). Utilisation du langage naturel et de représentations graphiques pour la description d'itinéraires. In *Images et langages. Multimodalité et modélisation cognitive. Actes du colloque interdisciplinaire du CNRS*, p. 243–251, Paris.
- MAASS W. (1993). A cognitive model for the process of multimodal, incremental route descriptions. In *Spatial information theory*, volume 716 of *LNCS*, p. 1–13: Springer-Verlag.