« Un modèle pour l'alternance de langue sous la contrainte d'équivalence »

David Sankoff et Sylvie Mainville

Revue québécoise de linguistique, vol. 15, n° 2, 1986, p. 233-245.

UN MODÈLE POUR L'ALTERNANCE DE LANGUE SOUS LA CONTRAINTE D'ÉQUIVALENCE

David Sankoff et Sylvie Mainville

1. Introduction

Dans le discours des bilingues, il n'est pas rare d'entendre des éléments de différentes langues à l'intérieur d'une seule phrase. Parmi ces juxtapositions figurent des sons, des mots ou même des structures syntaxiques propres à chacune des deux langues. De tels mélanges relèvent de divers processus déclenchés par le contact linguistique, dont l'emprunt lexical, l'interférence phonologique et l'alternance de langue (*code-switching*).

Les phénomènes du bilinguisme posent un problème intéressant aux modèles formels courants en linguistique. La spécification des paramètres d'un modèle général afin de décrire une langue donnée implique l'identification de restrictions, de contraintes et de conditions particulières à cette langue qui ne sont pas forcément compatibles avec des composantes d'autres langues. On se demande donc jusqu'à quel point les modèles des grammaires unilingues peuvent être utilisés pour engendrer des productions bilingues et quels mécanismes devraient être ajoutés aux modèles existants pour qu'ils puissent s'appliquer au cas du bilinguisme.

Dans cet article, nous abordons ce problème d'un point de vue mathématique en ne traitant que d'un aspect du comportement bilingue à savoir l'alternance intra-phrastique entre langues qui diffèrent au niveau de l'ordre des mots. Notre approche suppose un certain nombre d'hypothèses et de simplifications propres à assurer que nos analyses mathématiques aboutissent à des conclusions significatives. Loin de prétendre à la validité stricte de ces hypothèses pour les langues naturelles, nous espérons être éventuellement en mesure d'étendre notre analyse à des modèles plus adéquats.

Le modèle de grammaire unilingue que nous utilisons est relativement simple. Il s'agit de la grammaire indépendante du contexte qui offre le double avantage d'être bien comprise d'un point de vue mathématique et de tenir compte de plusieurs des propriétés des langues naturelles liées à l'ordre des mots. Dans ce cadre, notre définition de l'alternance de langue s'appuiera sur la contrainte d'équivalence (Poplack 1978, 1980) dont la pertinence pour l'alternance intra-phrastique a été vérifiée dans diverses communautés bilingues.

Dans un premier temps, nous construirons un modèle formel capable d'engendrer les phrases bilingues qui satisfont à la contrainte, ainsi que toutes les phrases grammaticales unilingues de chacune des deux langues. Ce modèle aura lui-même la capacité générative d'une grammaire indépendante du contexte. Certains éléments de ce projet figurent déjà dans Sankoff et Poplack (1981) et dans Rivas (1980), bien que les aspects mathématiques n'y soient pas développés.

Dans un second temps, nous explorerons le rôle de la typologie de l'ordre des mots dans l'identification des points qui permettent ou empêchent les alternances de langue dans des phrases types. Pour ce faire, nous ajouterons un aspect probabiliste à des grammaires rudimentaires correspondant à quelques-uns des types dans la typologie établie par Greenberg (1966). La quantité de points d'alternance sera ensuite prédite au moyen d'un théorème (Sankoff 1971) sur les grammaires probabilistes.

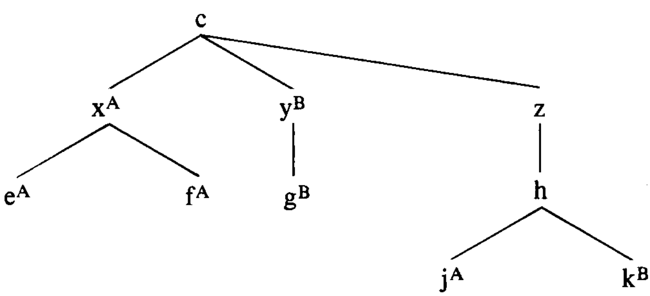
2. Les grammaires indépendantes du contexte et la contrainte d'équivalence

Considérons une grammaire indépendante du contexte composée d'un ensemble de symboles non-terminaux, d'un ensemble de symboles terminaux, et d'un ensemble R de règles de réécriture de la forme $c \rightarrow v_1v_2 \ldots v_n$ où le membre de gauche c est non-terminal et où les éléments de la suite $v_1v_2 \ldots v_n$ dans le membre droit sont terminaux ou non-terminaux.

Nous emploierons la représentation arborescente traditionnelle pour la dérivation et la structure d'une phrase. Tous les symboles apparaissant en cours de dérivation correspondent à un nœud de l'arbre. Chaque symbole (y compris les symboles terminaux) domine le constituant dont il est le sommet et on peut se servir de ce symbole pour désigner le constituant.

Une condition pour l'analyse de l'alternance de langue est l'existence de liens entre les catégories des deux grammaires qui engendrent ces langues. Plutôt que de nous attaquer à la question controversée, soit de savoir si ces liens sont valables dans le cas de deux langues naturelles (question qui a relativement peu d'incidence sur l'ordre des mots), nous nous permettrons dans un premier temps de poser l'hypothèse d'une correspondance complète entre les catégories de la langue A et celles de la langue B et nous utiliserons même des symboles identiques pour les catégories des deux langues. De plus, nous supposerons une correspondance bijective entre les règles de A et de B — si la langue A a une règle $c \rightarrow v_1v_2 \ldots v_n$ la langue B en aura une de la forme $c \rightarrow u_1u_2 \ldots u_n$, où chaque symbole de la suite $v_1v_2 \ldots v_n$ a sa contrepartie dans la suite $u_1u_2 \ldots u_n$, et vice-versa, bien qu'en général l'ordre des termes diffère dans les deux suites. Finalement, nous supposerons que l'ordre des mots est fixe; si une règle d'une grammaire donnée réécrit c en une suite $v_1v_2 \ldots v_n$, il n'y aura pas d'autre règle dans cette grammaire qui puisse réécrire c en une permutation des mêmes éléments $v_1,\ldots,v_n$. (Un ordre variable compliquerait notre présentation mais ne changerait guère notre analyse.)

Nous sommes donc en mesure d'énoncer formellement la contrainte d'équivalence. Considérons une structure de phrase dont certains nœuds (ou tous) sont étiquetés quant à la langue d'appartenance, comme l'illustre la figure suivante :



La contrainte ne s'applique qu'entre deux constituants adjacents, c'est-à-dire dont les sommets dépendent directement du même nœud. Ainsi, elle est pertinente seulement pour les paires x et y, y et z, e et f et j et k.

Donc pour qu'une phrase bilingue soit dite grammaticale, chaque paire de constituants adjacents (ex : x et y) doit satisfaire aux conditions suivantes : considérons l'ensemble E des descendants immédiats du nœud qui domine les deux constituants en question, c'est-à-dire les deux nœuds ainsi que tout autre nœud qui dépend immédiatement du même ancêtre (ex : x, y et z). Si les deux nœuds ont des étiquettes différentes, ou si l'un ou l'autre n'est pas étiqueté, la contrainte exige que l'ensemble E concorde avec le membre droit d'une règle de réécriture dans chacune des deux langues, de la façon suivante : *les symboles correspondants aux nœuds de l'ensemble E situés à gauche de la frontière entre les deux constituants doivent précéder les symboles correspondants aux nœuds de l'ensemble E situés à droite de la frontière et ce, dans les deux règles correspondantes dans les deux langues.* Ainsi si la règle dans la langue A est c → x y z et celle dans la langue B, c → x z y, les nœuds x^A et y^B satisfont à la contrainte tandis que y^B et z ne la respectent pas. Par contre si les deux nœuds adjacents ont la même étiquette, disons A, alors la contrainte demande simplement que l'ensemble E concorde de cette façon avec le membre droit d'une règle de la langue A.

Cette description de la contrainte d'équivalence exige que nous ayons une méthode pour identifier des nœuds non-terminaux comme appartenant à une ou l'autre langue. La façon de réaliser cette identification fait l'objet d'une controverse. Notre approche consistera à engendrer à la fois la phrase bilingue et les étiquettes sur les nœuds non-terminaux de l'arbre afin de nous assurer de la consistance mutuelle de la grammaire, de l'étiquetage et de la contrainte d'équivalence. Comme l'illustre la figure ci-dessus, nous admettons des symboles étiquetés et non-étiquetés sur les nœuds.

Avant d'engendrer des phrases bilingues, nous examinons chaque paire de règles correspondantes dans les deux grammaires. En alignant les membres droits de ces règles on remarque que des points d'équivalence (là où l'alternance de langue est permise selon la contrainte d'équivalence) arrivent aux mêmes positions dans les deux règles. Entre deux points d'équivalence successifs, et aux deux extrémités des membres droits, il y a un seul symbole (singleton) ou une région de non-équivalence. Par exemple :

points d'équivalence

Langue A : $c \rightarrow t\ u \downarrow v\ x\ y \downarrow z$

Langue B : $c \rightarrow \underbrace{u\ t}\ \ \underbrace{y\ v\ x}\ \ z$

singleton

régions de non-équivalence

La génération d'une phrase procède à partir de la réécriture du symbole S, S^A ou S^B. En réécrivant un symbole non-terminal, on se sert d'une paire de règles correspondantes des deux grammaires. Les symboles dans chaque région de non-équivalence doivent être ordonnés et étiquetés selon la règle d'une de ces grammaires. La grammaire utilisée pourra différer d'une région à l'autre, par exemple :

$$c \rightarrow u^B\ t^B\ v^A\ x^A\ y^A\ z$$

Les symboles hors de ces régions de non-équivalence (les singletons) peuvent être étiquetés A ou B, ou demeurer non-étiquetés. Les régions et les singletons se suivent dans le même ordre que dans les deux règles. L'étiquetage de l'ensemble des descendants d'un nœud doit satisfaire aux conditions suivantes :

i) si le symbole non-terminal est non-étiqueté, tous les descendants ne peuvent posséder la même étiquette;
ii) si le symbole non-terminal a l'étiquette A, alors au moins un de ses descendants doit porter l'étiquette A. De même pour B;
iii) tous les symboles terminaux doivent recevoir une étiquette.

Notons que s'il n'existe aucune paire de règles de réécriture d'un symbole non-terminal qui satisfait à la condition i), alors ce symbole devrait être étiqueté A ou B partout où il apparaît dans une dérivation.

Notre façon d'engendrer des phrases bilingues possède les propriétés suivantes. Premièrement, elle satisfait à la contrainte d'équivalence — négativement et positivement. Négativement, car il n'y aura pas d'alternance là où c'est interdit mais positivement aussi car les alternances pourront survenir partout ailleurs. Deuxièmement, tout constituant unilingue d'une phrase est grammaticalement correct par rapport à la grammaire unilingue appropriée.

Troisièmement, si un constituant possède une étiquette, alors au moins un des symboles terminaux qu'il domine héritera de cette étiquette (condition ii); tout modèle plus simple engendrerait des phrases unilingues non-grammaticales à cause de l'absence de cette propriété et conduirait à d'autres conséquences contraires à l'intuition. Plusieurs théories de l'alternance, telles que celles de Sankoff et Poplack (1981), Rivas (1980) et Muysken, di Sciullo et Singh (ms), contiennent aussi des conditions semblables d'hérédité, quoiqu'on y spécifie, de façons différentes, quels sont les symboles qui doivent hériter de l'étiquette.

Une quatrième propriété, conséquence des conditions i) et ii), est que si tous les symboles terminaux d'un constituant possèdent une même étiquette, alors la catégorie dominant ce constituant possède aussi cette étiquette. L'absence de cette propriété dans un modèle serait contraire à l'intuition.

Cinquièmement, il n'y a pas d'étiquette non motivée. Les propriétés précédentes nous assurent de la présence des étiquettes sur a) les paires de constituants entre lesquels on interdit une alternance; b) au moins une ligne de descendants de tout nœud étiqueté; c) les nœuds non-terminaux de tout constituant dont les nœuds terminaux sont uniformément étiquetés. Aucune de ces étiquettes ne peut être éliminée sans sacrifier à la crédibilité du modèle. Cependant, aucun autre étiquetage n'est exigé. Ce conservatisme se justifie par des données sur des bilingues équilibrés (Sankoff et Poplack 1981) où l'affectation d'étiquettes à tous les nœuds de l'arbre d'une phrase bilingue s'avère souvent arbitraire et non-naturelle.

Sixièmement (en fait une conséquence directe de la contrainte d'équivalence), aucun constituant, même s'il est bien formé, ne peut apparaître mal ordonné dans le constituant qui le contient, à cause d'une alternance. Cette propriété n'est pas partagée par des modèles dans lesquels on utilise plus librement des règles des deux grammaires pour réécrire les symboles non-terminaux (Woolford 1983). Dans un travail qui sera publié prochainement, nous mettrons en évidence des justifications de cette propriété à l'aide de données empiriques sur l'ordre verbe-objet où les deux langues des bilingues sont de type SOV et SVO, et aussi sur les phrases adpositionnelles où les deux langues sont de type prépositionnel et postpositionnel, respectivement.

Septièmement, ce mécanisme de génération des phrases est essentiellement une grammaire indépendante du contexte, quoique nous ne l'avons pas explicitement formalisé comme tel. Nous l'avons plutôt formulé en termes de règles identiques ou différentes pour les catégories correspondantes dans les deux grammaires unilingues afin de mettre en évidence les rapports entre celles-ci et le mécanisme d'alternance. Néanmoins, d'un point de vue technique, nous avons engendré les phrases bilingues au moyen d'une seule grammaire indépendante du contexte.

3. Fréquence des points d'alternance

Si l'ordre des termes du membre droit des règles correspondantes de deux langues diffère beaucoup, la contrainte d'équivalence permettra peu d'alternance dans les phrases. Par contre, si deux langues ont des règles identiques on peut concevoir que dans une phrase la langue alterne entre chaque paire de symboles terminaux. Dans cette section, nous explorons le rôle de la typologie de l'ordre des mots sur la fréquence des points d'alternance. Nous ajouterons à notre modèle des paramètres quantitatifs qui déterminent les taux relatifs d'utilisation des différentes règles de réécriture d'un même symbole non-terminal. Comme sous-produit, nous dériverons la dépendance quantitative entre l'abondance des points d'alternance et la fréquence des différentes règles récursives et non-récursives de la grammaire.

Nous construisons un modèle de performance en spécifiant les probabilités des règles d'une grammaire indépendante du contexte. À un symbole non-terminal donné, pour lequel il existe N règles de réécriture, nous assignons aux règles 1, 2, ..., N des nombres positifs $p_1, p_2, \ldots, p_N$ dont la somme est 1. Nous pouvons alors déterminer entièrement le comportement statistique de la langue engendrée par la grammaire. Plus particulièrement, on détermine ainsi le nombre moyen de symboles terminaux de chaque type dans une phrase.

Dans cette section, nous voudrions prédire la fréquence moyenne des points possibles d'alternance. Or, ces points ne sont pas eux-mêmes des symboles terminaux. On peut contourner ce problème en construisant une grammaire enrichie de deux nouveaux symboles terminaux : σ et τ. Le symbole σ représente un point possible d'alternance tandis que le symbole τ représente un point où l'alternance est interdite. En comparant les paires de règles correspondantes de R^A et R^B; par exemple $c \rightarrow v_1v_2 \ldots v_n$ et $c \rightarrow u_1u_2 \ldots u_n$ on crée une nouvelle règle $c \rightarrow v_1v_2 \ldots v_n\sigma \ldots \sigma\tau \ldots \tau$, où il y a autant de σ que de points possibles d'alternance permis par la contrainte d'équivalence et où le nombre de τ égale le nombre de frontières où on ne permet pas l'alternance. Par exemple, si les règles correspondantes dans R^A et R^B sont $c \rightarrow w\ x\ y\ z$ et $c \rightarrow w\ x\ z\ y$ la nouvelle règle sera $c \rightarrow w\ x\ y\ z\ \sigma\sigma\tau$ puisqu'il peut y avoir alternance après le w et après le x et qu'il y a un point défendu entre le y et le z.

Remarquons que puisqu'on ne s'intéresse ici qu'au nombre de σ et de τ dans les phrases, une fois qu'on a déterminé combien il y en a dans le membre droit de chaque nouvelle règle en comparant les deux règles unilingues correspondantes, ni l'ordre des termes dans le membre droit de la nouvelle règle, ni les étiquettes sur les nœuds non-terminaux, n'ont d'importance.

Nous nous servons d'une version probabiliste de la nouvelle grammaire pour calculer le nombre moyen de σ et de τ par phrase, à partir d'une méthode de l'algèbre linéaire développée par Sankoff (1971, 1972).

Pour illustrer les résultats de cette méthode, nous construisons cinq grammaires indépendantes du contexte. Elles diffèrent entre elles selon les quatre aspects de l'ordre des mots qui caractérisent la typologie du langage de Greenberg (1966) : SVO versus SOV versus VSO, prépositions versus postpositions, adjectif-nom versus nom-adjectif et génitif-nom versus nom-génitif.

Voici les cinq grammaires ;

TYPE 1 (ex : langues celtes)	TYPE 9 (langues romanes)
S → vNO	S → NvO
O → P	O → P
O → N	O → N
N → n	N → n
N → na	N → na
N → ng	N → ng
N → nag	N → nag
N → NS	N → NS
P → pN	P → pN

TYPE 11 (langues scandinaves)	TYPE 15 (finlandais)
S → NvO	S → NvO
O → P	O → P
O → N	O → N
N → n	N → n
N → an	N → an
N → gn	N → gn
N → gan	N → gan
N → NS	N → NS
P → pN	P → Np

TYPE 23 (langues dravidiennes)

$$
\begin{aligned}
S &\rightarrow NOv\\
O &\rightarrow P\\
O &\rightarrow N\\
N &\rightarrow n\\
N &\rightarrow an\\
N &\rightarrow gn\\
N &\rightarrow gan\\
N &\rightarrow SN\\
P &\rightarrow Np
\end{aligned}
$$

Nous combinons deux à deux ces cinq grammaires afin de produire des grammaires bilingues; entre les types 11 et 9, seules des différences au niveau des modificateurs de noms sont impliquées; entre les types 11 et 15 interviennent des contrastes au niveau des adpositions; entre les types 11 et 23, ce sont des différences aux niveaux de la phrase et des adpositions; entre les types 1 et 9, ce sont des différences uniquement au niveau de la phrase; finalement les types 1 et 23 diffèrent de toutes les façons possibles.

La grammaire bilingue pour la combinaison des types 11 et 23 se décrit comme suit :

$$
\begin{aligned}
S &\rightarrow NvO\sigma\tau\\
O &\rightarrow P\\
O &\rightarrow N\\
N &\rightarrow n\\
N &\rightarrow an\sigma\\
N &\rightarrow gn\sigma\\
N &\rightarrow gan\sigma\sigma\\
N &\rightarrow NS\tau\\
P &\rightarrow pN\tau
\end{aligned}
$$

Rappelons qu'à ce moment, nous sommes indifférents à l'ordre des termes du membre droit des règles; en effet la première règle aurait bien pu s'écrire $S \rightarrow N\sigma v\tau O$ ou même $S \rightarrow ON\sigma\tau v$.

Pour probabiliser cette grammaire, nous associons des probabilités, de somme égale à 1, aux règles $O \rightarrow P$ et $O \rightarrow N$; et des probabilités, de somme égale à 1, aux cinq règles de réécriture du symbole N. (Nous ne devons pas donner une trop forte probabilité aux règles récursives telles que $N \rightarrow NS\tau$ afin d'éviter les dérivations qui ne se terminent jamais. Nous omettons ici les détails mathématiques de cette condition.) Dans le tableau 1, nous donnons les résultats de l'assignation des probabilités 0,7 à $O \rightarrow P$, 0,3 à $O \rightarrow N$, 0,41 à $N \rightarrow n$, 0,18 à $N \rightarrow NS\tau$, 0,25 à $N \rightarrow an\sigma$ et 0,08 à $N \rightarrow gn\sigma$ ainsi qu'à $N \rightarrow gan\sigma\sigma$.

Paire de langues	Symboles terminaux	Points d'alternance prohibés	Points d'alternance possibles
11 / 9	9,2	2,6	5,6
11 / 15	9,2	1,2	7,0
11 / 23	9,2	3,8	4,4
1 / 9	9,2	1,8	6,4
23 / 1	9,2	8,2	0

Tableau 1. Nombre moyen de points possibles d'alternance par phrase pour diverses combinaisons d'ordre de mots

La longueur moyenne d'une phrase est de 9,2 symboles terminaux (en excluant les σ et les τ). En moyenne, il y a 4,4 paires de symboles terminaux entre lesquels on admet un point d'alternance et 3,8 paires entre lesquels ils sont interdits. Le tableau montre aussi que pour les paires de langues qui diffèrent seulement par leur type adpositionnel, le nombre de points où l'alternance est interdite diminue radicalement à 1,2 en moyenne, tandis qu'il n'y a aucun point d'alternance possible si l'on combine deux langues complètement différentes.

Or ces résultats dépendent de l'assignation des probabilités aux règles de la grammaire. Nous illustrons au tableau 2 cette dépendance pour trois des probabilités.

Comme on aurait pu s'y attendre, la dépendance entre le nombre de σ et la probabilité assignée à la règle O $\rightarrow$ P est positive lorsque les deux langues sont prépositionnelles, alors qu'il y a dépendance entre le nombre de τ et cette probabilité lorsqu'une langue est postpositionnelle et l'autre prépositionnelle.

Paire de langues	Dépendance sur p pour les règles :					
	O $\rightarrow$ P		N $\rightarrow$ n		N $\rightarrow$ NSτ ou N $\rightarrow$ NSσ	
	τ	σ	τ	σ	τ	σ
11 / 9	0	1,8	-10,5	0	-10	41
11 / 15	1,8	0	0	-10,5	8	23
11 / 23	1,8	0	0	-10,5	30	1
1 / 9	0	1,8	0	-10,5	11	19
23 / 1	1,8	0	-10,5	0	31	0

Tableau 2. La dépendance de fréquence de points d'alternance possibles et prohibés, sur les probabilités des règles. Les chiffres indiquent le changement de nombre de σ et de τ par changement de probabilité de règle.

Les changements provoqués par une modification de la probabilité de la règle N → n sont tout à fait différents. Lorsque cette probabilité croît, le nombre de groupes nominaux plus complexes ainsi que le nombre de τ décroissent pour les paires de langues qui diffèrent quant à la position des modificateurs de nom. Par contre, pour les paires de langues partageant le même ordre dans le groupe nominal, l'accroissement de la probabilité de la règle N → n provoque une décroissance dans le nombre de points possibles d'alternance.

Finalement, lorsque la probabilité de la règle récursive N → NSτ (ou N → NSσ) augmente, nous trouvons des changements plus subtils qui accompagnent l'augmentation de la longueur des phrases. Le cas le plus favorable à l'alternance est le pairage (type 11 / type 9) où seul diffère l'ordre dans le groupe nominal, suivi des cas type 11 / type 15 où seuls les adpositions diffèrent, type 1 / type 9 où l'ordre des mots dans la phrase diffère, type 11 / type 23 où tout diffère sauf la position du nom et finalement le cas type 23 / type 1 où tout diffère et où seulement des τ viennent s'ajouter quand la probabilité de la règle augmente.

4. Discussion

Le débat autour de la contrainte d'équivalence a souffert d'un manque de précision quant à son interprétation dans les cas particuliers. Nous croyons avoir réussi ici à formuler la contrainte de façon explicite et rigoureuse, quoique dans un cadre restreint — celui des grammaires indépendantes du contexte. Cette formulation exige une attention particulière à l'étiquetage des constituants d'une phrase quant à la langue d'appartenance.

À partir de cette définition de la contrainte, nous avons construit un mécanisme capable d'engendrer l'ensemble des phrases bilingues qui la satisfont, ce qui implique que la structure qui accompagne chaque phrase soit étiquetée de la bonne façon.

Les étiquettes satisfont à deux conditions supplémentaires : chaque nœud non-terminal qui possède une étiquette domine au moins un symbole terminal avec la même étiquette et chaque constituant où les symboles terminaux ont tous la même étiquette doit être dominé par un nœud avec cette même étiquette. On peut, de plus, considérer que notre procédé de génération est une grammaire indépendante du contexte.

On remarque qu'une phrase bilingue peut correspondre à plus d'un étiquetage des nœuds non-terminaux. Généralement, la grammaire bilingue est ambiguë en terme d'étiquetage des nœuds non-terminaux même lorsque tout autre aspect de l'arbre de dérivation est déterminé sans ambiguïté. Ce fait peut mener à des difficultés dans l'analyse syntaxique des phrases bilingues mais cela reflète très bien la difficulté, souvent rencontrée, à déterminer clairement le niveau de la structure hiérarchique où s'est produite l'alternance.

Notre discussion s'appuie sur deux hypothèses simplificatrices qui facilitent l'exposition. Il est possible de leur substituer des hypothèses plus faibles. On peut facilement affaiblir l'hypothèse d'ordre unique des mots dans chacune des langues en permettant des ordres variables dans l'une ou l'autre des langues. Cela occasionne une augmentation du nombre de phrases bilingues potentielles puisque l'alternance n'est interdite que si les langues ne possèdent pas le même ordre. On peut laisser tomber l'hypothèse d'identité de catégories entre les deux langues en admettant des catégories et des symboles dans une langue sans qu'ils aient de contrepartie dans l'autre langue ou en se dispensant de la bijection entre les ensembles R^A et R^B. Ce changement serait plus difficile à effectuer et aurait pour effet de réduire le nombre de points d'alternance.

Notre utilisation des grammaires indépendantes du contexte est sans doute l'aspect le moins réaliste de notre théorie. Il n'est pas déraisonnable cependant d'espérer que plusieurs aspects de notre approche pourront se généraliser aux grammaires admettant les effacements, les catégories vides et les transformations.

David Sankoff
Centre de recherches mathématiques
Université de Montréal

Sylvie Mainville
Département de mathématiques
Collège Militaire Royal

Références

GREENBERG, J. (1966) «Some Universals of Grammar, with Particular Reference to the Order of Meaningful Elements», dans *Universals of Language.* Cambridge : MIT Press.

MUYSKEN, P., A.-M. Di Sciullo et R. Singh (ms) «Code-mixing and Government», Université de Montréal.

POPLACK, S. (1978) «Syntactic Structure and Social Function of Code-switching», dans *Latino Discourse and Communicative Behavior.* New-Jersey : Ablex.

POPLACK, S. (1980) «Sometimes I'll Start a Sentence in Spanish y termino en español», *Linguistics 18,* pp. 581-618.

RIVAS, A.M. (1980) «On the Application of Transformations to Bilingual Sentences», *Umass Working Papers on Hispanic Linguistics and Bilingualism.*

SANKOFF, D. (1971) «Branching Processes with Terminal Types : Application to Context-free Grammars», *Journal of Applied Probability 8,* pp. 233-240.

SANKOFF, D. (1972) «Context-free Grammars and Non-negative Matrices», *Linear Algebra and Applications 5,* pp. 277-281.

SANKOFF, D. et S. Poplack (1981) «A Formal Grammar for Code-switching», *Papers in Linguistics 14,* pp. 3-46.

WOOLFORD, E. (1983) «Bilingual Code-switching and Syntactic Theory», *Linguistic Inquiry 14,* pp. 520-536.