

L'approche syntaxique des dynamiques langagières : non-standard et variation

Katia Ploog*

La description de la variabilité des productions langagières telle qu'elle est proposée par la sociolinguistique variationniste s'appuie sur l'identification d'un trait structural qui donne lieu à des réalisations multiples, conditionnées par des facteurs internes et/ou externes que l'on cherche à déterminer dans un corpus. Or, dans des productions non standard, l'existence d'un trait variable n'est pas prévisible ni même toujours visible ; nous proposons une réflexion sur l'identification même de cette variable, qui constitue dans ce cas l'enjeu de la démarche.

The sociolinguistic analysis of language usually works out the internal and/or external parameters of several variants relied to one structural variable in a corpus of speech data. In non standard speech corpora, the existence of a variable is not yet known and even may not be visible immediately. Our aim is to show how we can initiate the description of this kind of language data : The first step of the approach is to identify by which feature(s) the non standard structure differs from the standard one.

* ERSSAB, CNRS UMR 5610 - Bordeaux III.

1. Introduction : le non-standard dans les dynamiques langagières

1.1. Problématique

Les situations de contact génèrent chez les locuteurs des représentations spécifiques des langues impliquées, représentations qui se répercuteront dans des pratiques linguistiques différentielles, qui peuvent, dans certaines conditions, affecter le système lui-même. La variation structurale et l'hétérogénéité de la langue sont parties prenantes des mécanismes du changement linguistique. Toutes les langues sont soumises à ces enjeux de restructuration, que nous approcherons ici par les manifestations *non standard* telles que les discours les révèlent.

Le niveau syntaxique reste l'un des moins explorés dans le domaine des études sur la variation linguistique. L'objet d'étude est en effet suffisamment peu établi pour nous inciter à affiner sa délimitation, puis à examiner les outils permettant son approche. Avant de pouvoir affirmer que les conditions d'apparition de ces caractéristiques laissent envisager un changement syntaxique, il convient ainsi de s'interroger sur la nature même du non-standard.

1.2. Corpus

Nous illustrerons nos réflexions à l'aide d'exemples tirés d'un corpus de français parlé, recueilli à Abidjan, métropole ivoirienne, où le français est devenu langue principale au cours des quarante dernières années. Praticué par une population peu exposée dans sa majorité à la norme standard, le français des Abidjanais comporte aujourd'hui des caractéristiques structurales non standard dont seulement une partie est partagée avec le français parlé en France. Les particularités des niveaux lexical et phonétique ont fait l'objet de nombreuses descriptions et sont aujourd'hui reconnues au titre d'appropriations socio-sémantiques de la langue française en contexte africain. En revanche, les spécificités des niveaux morphosyntaxique et discursif ont presque toujours été reléguées au rang d'erreurs voire d'impuretés, à celui d'hypercorrections dans le meilleur des cas. Nous discuterons dans quelle mesure l'approche variationniste permet d'appréhender les phénomènes observés.

L'étude systématique d'un corpus de productions d'enfants peu scolarisés¹ a permis de dresser un premier inventaire des dispositifs syntaxiques investis et de dégager un certain nombre de particularités, *plus* ou *moins* innovantes, relatives à la microstructure du premier actant. Il est apparu par la suite que d'autres actants étaient affectés par ces restructurations.

¹ L'analyse porte sur près de 10000 unités propositionnelles (microsyntaxiques) manuellement (systématiquement mais partiellement) étiquetées.

L'objet de cet article est de remettre en perspective ces observations : le type de données analysées et l'origine des locuteurs qui les produisent fournissent certes quelques éléments de compréhension sur les mécanismes à l'œuvre (i.e. interférences des langues substrats et/ou paramètres liés aux dynamiques discursive et interactionnelle). La question de l'intégration de ces restructurations au système global reste néanmoins entière, en partie à cause de la lenteur des dynamiques langagières, car un changement ne peut être constaté qu'« après coup ».

1.3. Le non-standard : définition

Le *non-standard* est constitué par la somme des traits non compatibles avec les principes d'une grammaire standard mais relevés dans le discours.

Trois aspects de cette définition seront développés dans cet article : dans la mesure où le non-standard n'existe qu'en creux du standard, nous aborderons la dialectique entre celui-ci et le non-standard ; nous proposerons ensuite une typologie des structures pouvant se produire en marge et au-delà de la grammaire ; la problématique de son apparition non systématique dans le discours nous conduira enfin à raisonner l'hétérogénéité du système, en pointant quelques problèmes majeurs liés aux contraintes méthodologiques de la démarche descriptive.

2. Articulation entre non-standard et standard

Le terme *non-standard* désigne son objet par la négative. L'avantage immédiat consiste à rappeler au lecteur à chaque moment la spécificité de cet objet, par définition imbriqué au standard mais exclu de celui-ci. Traditionnellement, le terme *non-standard* traduit également l'implication sociale (« involvement » : Chafe 1982, Cheshire 1987) dans les formes manipulées, existant autant chez le locuteur que chez l'observateur.

Afficher clairement le point de vue de l'observateur, empreint de la Norme et du métalangage qui l'accompagne, semble prudent dès lors que l'objet n'est pas convocable à souhait. *A priori*, le non-standard échappe à la compétence du linguiste, techniquement, et souvent aussi socialement. Paradoxalement, ceci aura un effet « positif » sur la description, dans la mesure où l'on s'attend à ce que la sensibilité de l'observateur au non-standard soit proportionnelle à sa compétence du standard.

2.1. Coexistence et changement

La tradition variationniste pose comme point de départ la coexistence de deux termes X et Y, dans notre cas l'un non standard, l'autre standard, dans une relation d'« équivalence communicative » (la variable). L'on cherche alors à dégager les facteurs internes et externes déterminant la réalisation

concrète de cette variable. Nous reprenons ici la schématisation des différents cas de coexistence proposée par Garcia (1997 : 31) :

- A. Disjonction : aucun environnement où X apparaît n'est partagé par Y (absence d'intersection, distribution complémentaire) ;
- B. Inclusion : tous les environnements où X apparaît permettent Y, mais non *vice versa*.
- C. Intersection :
 - i. totale : tous les environnements où X apparaît permettent Y et *vice versa*,
 - ii. partielle : une partie des environnements où X apparaît permettent aussi Y et *vice versa*.

Dans le cas A, chaque terme est doté de suffisamment de caractéristiques structurales propres pour qu'aucun contexte d'emploi ne soit partagé. Il n'y a donc pas d'équivalence et pas lieu non plus de parler de *variation* : il s'agit d'un cas de figure de réorganisation aboutie du système. Néanmoins, les pratiques, telles que nous les observons dans les discours, s'articulent en synchronie sous forme de continuum : la situation sociale conduit tous les locuteurs à investir dans un même discours des formes standard et des formes non standard. Les propriétés des unités linguistiques n'en sont que plus délicates à établir. Tout comme le standard, qui est composé d'éléments plus ou moins normés/normatifs, le non-standard n'a pas le statut d'une variété. Il se raisonne comme une somme de traits en décalage avec le standard, tel qu'il est validé par les usages valorisés et/ou les grammaires.

En synchronie, le cas A semble donc constituer davantage une modélisation qu'une réalité discursive.

2.2. Inclusion et intersection

Dès lors qu'une séquence alterne avec une autre dans le même contexte, il y a intersection. Mais l'existence d'une intersection apparente ne garantit pas l'équivalence. Il convient ainsi de mettre en évidence la nature fonctionnelle de l'intersection.

Garcia (1997) précise que le cas B est un cas particulier de C ; ici aussi, le continuum des usages rend difficile la distinction des deux cas. L'établissement de contraintes nettes se heurte à une alternance tout sauf systématique des formes considérées : que faut-il entendre par « les environnements *permettent* » (*cf.* ci-dessus) ? On sait combien il est hasardeux de prétendre connaître tous les environnements possibles à partir d'un corpus quelle que soit sa taille.

Par ailleurs, le niveau d'analyse choisi pour l'observation influe sur la notion d'*environnement* : il se pose la question d'une définition opérationnelle de l'équivalence communicative de deux termes au-delà du niveau phonétique. Une séquence syntaxique ne peut être dénuée de sens (ou de fonction pragmatique) et il paraît illusoire de penser que la référenciation

du corpus puisse prétendre à la restitution exhaustive de la diversité des forces opérant sur le discours.

D'autres problèmes se posent à l'application de cette grille de description, car il n'y a pas toujours deux termes : le microsystème peut en comporter plusieurs, ou investir une même forme dotée d'une valeur tantôt standard, tantôt non standard ; l'équivalence peut s'établir entre séquences entières ou entre traits structuraux (qui ne constituent pas des termes immédiatement identifiables) ; l'alternance d'une séquence phonologique et d'une marque « zéro » peut difficilement être décrite de la même manière que celle de deux termes effectivement observés, même si l'on restitue sur le plan interprétatif un élément, dont la présence sous forme de corps phonétique est localisée en référence au standard.

Ce premier « déblayage » du domaine nous aura permis de constater qu'il est indiqué d'observer la plus grande prudence lorsqu'on envisage la mise en contraste de non standard et standard. Il apparaît ainsi que l'opposition de deux termes constitue une simplification peu opérationnelle de la diversité structurale observée. Nous retrouverons ces problèmes d'interprétation de la variation dans les paragraphes traitant de la distribution des séquences non standard.

3. Structure des séquences non standard

Nous faisons le pari que les séquences non standard sont dotées d'une structuration interne, tout comme les séquences linguistiques du standard, les hapax étant exclus des considérations. La structure est descriptible comme étant dotée d'une forme et d'une valeur. La *forme* est la face linéaire d'une séquence repérée à un endroit donné de la chaîne de parole ; elle est immédiatement observable, et descriptible en termes de caractéristiques morpho-phonologiques. Sa *valeur* est la face empirique de l'objet ; elle doit être dégagée par le biais d'une double approche quantitative et qualitative, qui révélera les caractéristiques sémantiques - référentielles et relationnelles - de la séquence.

Cette première approche tente d'envisager le non-standard « en lui-même et pour lui-même », avant de recourir au crible imposé par le métalangage, inévitablement fondé sur le standard. Le tableau 1 schématise les quatre types structuraux que l'on obtient à partir du croisement de ces deux critères et des objets d'étude non-standard et standard :

Tableau 1

TYPE	FORME	VALEUR	EXEMPLES DE TRAITS NON STANDARD EN ABIDJANAIS
NS	non-std	standard	la marque « Ø » est une nouvelle forme pour la valeur sujet
NN	non-std	non-std	- un constituant non nominal peut se trouver en position sujet, - de nouveaux constituants intègrent la catégorie nominale (et peuvent désormais remplir la fonction du sujet) ; - la fonction sujet s'articule avec de nouvelles contraintes
SN	standard	non-std	redistribution fonctionnelle des clitiques objet de la 3 ^e personne
SS	standard	standard	double-marquage du sujet, omission du <i>ne</i> de négation

3.1. La forme non standard

Le recours à la transcription phonétique permet de déceler un certain nombre de séquences non standard. Si la transcription objective est un leurre – elle restera toujours impressionniste en raison du crible phonologique du transcritteur – elle évite néanmoins d'anticiper sur la description structurale que comprend l'orthographe, standardisée. Par exemple, la transcription révèle la forme [lɪvilapaseləbak] :

- (1) [oui de la tête] | nɔ̃ | dotfreak̃t | epɪija ləplygrã |
lɪvilapaseləbak | pijalezotkitravaj

(*Que font tes frères et sœurs, ils travaillent ?*) Non, certains vont à l'école, le plus grand a passé son bac, d'autres travaillent (B45Y : II.02/049)

Plusieurs représentations orthographiques seraient ici imaginables :

- (a) lui il a passé le bac (pronom fort et clitique sujets + verbe perfectif),
- (b) lui l'a passé le bac (pronom fort sujet + clitique objet direct + verbe perfectif),
- (c) lui-là a passé le bac (pronom sujet + postdétermination + verbe perfectif),
- (d) lui-là passait le bac (pronom sujet + postdétermination + verbe au passé/imparfait),

(e) lui [l] a passé le bac (pronom sujet + [consonne euphonique] + verbe perfectif ?

Une stratégie euphonique (e) paraît ici peu vraisemblable, car la tendance générale de l'abidjanais consiste plutôt à réduire les consonnes intervocaliques. Si la valeur (qualité morphologique) de la séquence transcrite ne peut être établie de façon univoque, c'est que tous les correspondants structuraux sont attestés à travers des formes non ambiguës. L'actualisation du pronom fort (c/d) existe, mais comporte un renvoi situationnel (comme en standard le démonstratif « celui-là »), ce qui n'est pas le cas ici. Du point de vue cotextuel, on relève une opposition directe de plusieurs frères et soeurs dans la même intervention, cadre dans lequel on peut trouver le pronom fort comme unique sujet (b) ; le double-marquage de l'objet par un clitique et un SN est également attesté. Du point de vue purement quantitatif, les séquences comportant un double-marquage du sujet (a) par le pronom fort et le clitique, demeurent les plus fréquentes. On pencherait alors pour l'interprétation (a), sans d'autre argument que l'empirie, et sans tenir compte de la redondance entre le constituant nominal et le pronom fort. En attendant d'effectuer les analyses nécessaires à la description, la représentation phonétique du discours permet de maintenir l'ambiguïté contenue dans une séquence sonore.

Si son interprétation structurale – la valeur qui lui revient dans le système – est souvent difficile à déterminer, une forme divergente du standard peut être aisément repérable dans la transcription. Considérons :

(2) mwajkõnesepalaba | pusmadidøveni | sakømadonedezabi
MOI(i) MOI-sj(i) CONNAITRE-passé neg LA-BAS-obj
POUR-ÇA Ø-sj² MOI-obj sg-perf-DIRE comp VENIR
POUR-ÇA comp Ø-sj MOI-obj sg-perf-DONNER pl-HABITS-obj

Moi je connaissais pas cet endroit ; c'est pour ça qu'il/on m'a dit de venir, qu'il/on m'a donné des habits. (C12X : I.02/005)

(3) evreunõ
Ø-sj sg-ETRE VRAI OU NON

C'est vrai ou non ? (C09X : VI.17/119)

En français standard, le sujet est la seule position syntaxique obligatoirement construite auprès du verbe tensé. Ces deux exemples comportent un constituant Ø pour la fonction sujet. Il s'agit donc de séquences à forme non standard. Qu'en est-il de leur valeur ? L'approche cognitive de la mécanique pragmatique (Givon 1976 et 1992) décrit la taille linguistique accordée à un référent comme inversement proportionnelle à son activation discursive ; le taux très faible d'ambiguïtés référentielles pour les sujets-Ø

² Selon les besoins de l'argumentation, les transpositions morphologiques matérialisent ou non des constituants « zéro » dans une position précise.

corrobore cette hypothèse. Or, si la proximité entre la marque \emptyset et le constituant clitique n'est pas seulement formelle, mais aussi référentielle, cela peut nous inciter à opposer les deux « formes » comme variantes concurrentes. Sont-elles pour autant équivalentes, dès lors que \emptyset n'exprime aucune des oppositions prises en charge par le paradigme clitique ? Le statut du sujet- \emptyset est-il phonologique ou syntaxique ?

3.2. La valeur non standard

De nombreux obstacles peuvent entraver l'identification de la valeur non standard d'une séquence donnée. Pour établir l'existence d'une valeur divergente de celle du standard, l'approche empirique des environnements structuraux s'impose.

Dans notre exemple, l'absence de tout constituant sujet semble être facilitée par la présence d'autres éléments dans la même zone - avant le verbe - qu'il s'agisse d'un autre clitique et/ou d'un constituant tonique antéposé (exemple 2). Le tableau 2 résume les environnements structuraux des occurrences de sujets- \emptyset dans le corpus :

Tableau 2

<i>facteurs phonologiques et syntaxiques</i>	introduceur	clitique	complétive	série verbale	aucun des facteurs
introduceur	39 ⁽³⁾	71	9	1	--
clitique	71	88	6	10	--
complétive	9	6	27	1	--
série verbale	1	10	1	31	--
Total	119	173	43	42	165
%	25,93	37,69	9,15	9,37	35,95

La présentation choisie souligne la présence fréquente de plusieurs facteurs facilitateurs⁴. Par ailleurs, le tableau révèle que dans plus d'un tiers des occurrences, aucun des environnements structuraux facilitateurs ne peut expliquer cette séquence, comme en témoigne l'exemple 3 ci-dessus. Au cours de l'examen des occurrences, il s'est avéré que l'occurrence d'un sujet- \emptyset peut également être relié à des facteurs prédictifs : la construction avec le

³ Présence du seul critère, ici *introduceur*, de même pour les cases clitique/clitique, complétive/complétive et SV/SV.

⁴ Comme nous n'avons pu tenir compte d'un éventuel troisième facteur dans ce tableau, la somme des taux dépasse les 100%.

verbe *être* ou alors le référent [-animé] (neutre, déictique) sont tous deux également propices à l'occurrence d'un sujet Ø.

Quoi qu'il en soit, aucun de ces facteurs ne permet de prédire l'occurrence d'un sujet-Ø, ce qui s'explique au moins en partie par l'hétérogénéité plus ou moins forte de chaque production, effet concret du continuum. Faute d'arguments structuraux convaincants – faute de *contraintes* – on doit conclure (du moins provisoirement) que la fonction sujet elle-même n'est pas affectée par le phénomène de réduction/omission de son constituant.

La séquence Ø s'insère dans un ensemble syntaxique plus vaste, dont la spécificité peut faire changer de perspective et soulever d'autres questions. Considérons :

- (4) isieruj l episie ve
ICI sg-ETRE ROUGE coord ICI sg-ETRE VERT
Ici c'est rouge et puis ici c'est vert (C06Y : VI.09/032)

Trois des facteurs précédemment cités sont réunis ici : la présence d'un introducteur, celle du verbe *être*, et le référent neutre du sujet. Un seul exemple suffit pour caractériser cette structure dont toutes les occurrences dans le corpus sont parfaitement symétriques : un référent locatif, dont le constituant linguistique est, en standard, catégorisé comme syntagme adverbial ou prépositionnel (i.e. *dans la rue n'est pas bon*), se positionne devant l'unité prédicative, composée de la copule et d'un adjectif qualificatif. La régularité de sa composition nous permet d'identifier cette séquence comme véritable innovation structurale, à forme *et* à valeur non standard (structure de type NN). Cependant, il n'est pas certain que cette construction soit descriptible en termes de sujet-Ø.

3.3. Identifier la valeur non standard dans une forme standard

L'approche empirique de la séquence dotée d'un trait NS implique que le premier pas, son identification, ait eu lieu, ce qui ne pose pas de problème pour les types NS et NN. En revanche, la spécificité structurale reste tout d'abord imperceptible lorsque la séquence possède une apparence conforme au standard : le type SN ne peut être appréhendé en opposition à un terme standard, puisqu'il ne s'en distingue pas formellement ; la distinction ne peut se faire que sur le plan fonctionnel. Or, celle-là même constitue l'objet de l'approche variationniste : la circularité rend caduque une telle démarche.

La valeur non standard d'une séquence peut être détectée à travers l'étude du microsystème auquel elle appartient, ou alors à l'occasion de l'étude d'un microsystème environnant. Par exemple, l'étude de la fonction sujet et l'observation de la possibilité d'un sujet-Ø nous ont conduit à examiner plus en détail toute la « zone » clitique préverbale, suivie de

l'auxiliaire verbal. Il est apparu que la distribution des clitiques objet de 3^e personne fait l'objet de quelques flottements entre *le* et *lui* :

- (5) safamladəmānde si | sōmarivapa māj
 (a) 3poss-FEMME(i)-sj **ELLE(i)-sj** sg-perf-DEMANDER (comp 3poss-MARI(k)-sj sg-prosp neg MANGER)j
 (b) 3poss-FEMME(i)-sj **LE(j)-obj** sg-perf-DEMANDER (comp 3poss-MARI(k)-sj sg-prosp neg MANGER)j
 (c) 3poss-FEMME(i)-sj **LUI(k)-obj** sg-perf-DEMANDER (comp 3poss-MARI(k)-sj sg-prosp neg MANGER)j
Sa femme [L] a demandé si son mari n'allait pas manger. (GM3Z: III.05/124)

Les cas (5) et (6) témoignent du fait que nous sommes confronté à un nouveau problème, dans la distinction des objets directs/indirects, qui s'ajoute aux ambiguïtés dues à l'opacité de la forme [l] pour l'analyse des formes sujet. L'opacité absolue du [l] est dans un premier temps interprétée comme mentionnée en (a) dans la mesure où le double-marquage d'un objet direct (le discours intégré par *si*) semble peu fréquent en standard – qui constitue le repère du descripteur. La « suspicion » est venue de l'observation du cas contraire, illustré en (6) :

- (6) ðlyiäv waje | pwale | pwale plānte | lemais
 ON-sj LUI-obj ENVOYER-passé comp-ALLER (bis) PLANTER det-MAIS
On [LUI] a envoyé planter le maïs. (GW2Y: III.06/062)

La particularité tout d'abord identifiée était l'absence de l'objet direct (i.e. *quel genre d'outil lui a-t-on envoyé pour planter ?*). Dans ce cas, le contexte permet d'établir sans équivoque qui envoie quoi à qui : [lYi] représente non pas le bénéficiaire mais le protagoniste, ici patient (lui-même a été envoyé pour planter le maïs)⁵.

Deux enseignements sont à tirer de ces observations. D'une part, que la référence puisse être restituée en contexte ne signifie pas pour autant qu'il s'agisse d'un phénomène de discours, où tout serait possible : la construction est diathétique, et le phénomène grammaticalement pertinent. D'autre part, l'approche s'expose à tout moment au risque de ne pas l'interpréter à sa juste valeur en isolant un point du système et sans connaître la « géographie » environnante. Le micro-système à envisager doit alors être celui du marquage

⁵ Claude Muller me suggère d'interpréter cette séquence comme une « fusion prédicative » entre la construction causative en *faire* et celle de l'infinitif *envoyer*, « on lui a fait planter le maïs » ; si une telle interprétation n'infirme pas l'hypothèse d'une restructuration diathétique, elle ne s'applique pas à toutes les occurrences de ce type et il ne me semble pas acquis par ailleurs que les locuteurs du corpus maîtrisent la construction causative.

morphologique (clitique) des compléments verbaux dans son intégralité, le sujet y compris.

4. Distribution des structures non standard

4.1. Le point de vue de l'observateur

On est tenté de voir dans ces difficultés d'identification et d'interprétation l'une des raisons pour lesquelles les particularités morphosyntaxiques de variétés marginales sont généralement traitées sur le seul plan de l'interlangue : si tout le monde peut se tromper de forme, cette explication paraît d'autant plus plausible lorsqu'on n'a pas à faire à un locuteur légitime. Les linguistes n'ont pas toujours conscience de décrire plutôt leur propres représentations que des faits observables et observés :

« L'analyse de la variation du français en Afrique fait apparaître une dichotomie : français littéraire / français véhiculaire. D'une façon générale, le français littéraire reste encore nettement dominant dans les usages quelle que soit la situation de communication. Ce français littéraire s'oppose à celui, véhiculaire et vivant pratiqué par une large majorité de locuteurs au marché, au stade, dans les bureaux et autobus. Cette langue vivante et dynamique, colorée et adaptée à toutes les situations de communication est partout pourchassée et farouchement combattue parce qu'elle est porteuse d'incorrections et d'impuretés qui menaceraient la norme. » (Tabi-Manga 1992 : 115)

Les notions descriptives employées révèlent tout d'abord une confusion entre *français littéraire* (paradigme genre ou texte) et *français standard* (paradigme formel et/ou social). Puis, au lieu de se contenter de dresser un inventaire d'observations, l'auteur émet des jugements pour le moins ambigus : il juge le français "littéraire" dominant dans toutes les situations, tout en constatant que le français "véhiculaire" est omniprésent : si les deux variétés conviennent dans toutes les situations de communication, l'appellation *véhiculaire* pour l'"autre" variété semble inappropriée. Dans cette folklorisation du français véhiculaire comme langue « vivante » et « colorée », l'attribut « dynamique » reste le qualificatif le plus juste, en opposant le véhiculaire à la Norme, nécessairement plus figée.

Si le métadiscours scientifique (!) peut ainsi témoigner d'une certaine partialité, il ne suffit pas d'en avoir conscience pour aller de l'avant : le recours à un corpus conséquent et édité de manière à rendre les données quantifiables en constituent le préalable. Il convient notamment de « protéger » les ambiguïtés formelles - au lieu de les lisser (pendant l'étiquetage) ou les éliminer (pendant l'analyse) - entre autres en répertoriant les différentes dimensions contextuelles des productions. C'est là une entreprise lourde et coûteuse, dont la rentabilité n'est en rien garantie, ce qui explique en partie le manque de données non standard.

4.2. Le contexte discursif

Le non-standard possédant quelques caractéristiques sociales qui le rapprochent du vernaculaire, il faut se résoudre à accepter que l'on n'y accède pas par introspection : il n'est présent que dans *certain*s discours, généralement parlés.

Le tableau 3 montre que la fréquence d'une structure non standard telle le sujet-Ø dépend du genre discursif :

Tableau 3

Locuteur	N propositions	N sujet Ø	% récit	% interview	% interaction entre pairs
C06Y	1323	66	4,7	4,0	6,2
C09X	812	35	2,1	2,1	5,2
C11Y	624	35	2,5	6,3	6,2

Sans grande surprise, on constate que les caractéristiques non standard apparaissent plus fréquemment lorsque la situation exerce peu de pression normative. Les écarts les plus importants apparaissent dans les interviews : la forte proportion de sujets-Ø chez C11Y peut s'expliquer ici par la finalité des échanges, qui étaient liées dans l'environnement matériel immédiat (et dont les référents étaient donc hautement activés), ainsi que par le débit particulièrement élevé de cet enfant (propice à la réduction d'éléments faibles) ; le locuteur C09X en revanche possède un niveau de scolarisation élevé, qui lui procure une capacité d'adaptation supérieure au standard, mise à profit dans les situations plus formelles telles le récit et l'interview (productions à l'intention de l'enquêteur, face au microphone).

On notera ainsi que l'augmentation générale de sujets Ø dans les interactions entre pairs résulte ou bien de la pression normative réduite, ou bien de l'objet du discours plus ancré dans le contexte spatial. Il s'y ajoute l'influence de caractéristiques d'élocution individuelles. Tout comme pour les facteurs structuraux, il convient de souligner l'impossibilité de relier l'origine d'un phénomène à un seul facteur, ce qui rend délicate la généralisation des observations.

4.3. Non-standard et norme

Le comportement du locuteur C09X (cf. paragraphe 4.2. ci-dessus) témoigne de ce que les locuteurs du non-standard ont bien conscience de l'opposition entre « ce qu'il faut faire » et leur comportement effectif. Le *monitoring* génère des ajustements vers la norme, dont la maîtrise reste néanmoins toute relative, toujours lacunaire.

Le tableau 4 illustre à nouveau la dialectique entre standard et non-standard, qui se manifeste particulièrement dans le type structurel SN, illustré par les exemples (5) et (6) déjà cités :

Tableau 4

Nombre	TYPE D'ACTANT	N TOTAL	RESTRUCTURATION	N	DYNAMIQUE	REMARQUES
singulier	objet (ex.6)	47	l > lui	10	pseudo-standard	1 ambiguïté
	bénéficiaire (ex.5)	46	lui > l	14	non-standard	2 ambiguïtés
pluriel	objet	13		0		6/13 auxiliaires [E] ⁶
	bénéficiaire	5	leur > les	3	non-standard	4/5 auxiliaires [E]

Ces décomptes se réfèrent au sous-corpus de récits d'écoliers.

En (5), l'actant bénéficiaire (objet indirect) est réalisé sous la forme [l(')] :

- (5) safamladəmānde si l sōmarivapa māj
 3poss-FEMME(i)-sj **LUI(k)-obj** sg-perf-DEMANDER comp 3poss-MARI(k)-sj sg-prosp neg MANGER
Sa femme lui a demandé si son mari n'allait pas manger. (GM3Z: III.05/124)

Près d'un tiers des réalisations maintiennent la structure non standard dans une situation formelle – en contexte scolaire, ce qui illustre à la fois la compétence restreinte du standard et la restructuration massive à cet endroit du système.

En corollaire, le phénomène inverse - où des objets (directs) sont réalisés par la forme [lYl] - est attesté dans plus de 20% des cas :

- (6) ðlyiāv waje l pwale l pwale plānte l lemais
 ON-sj LE-obj sg-perf-ENVOYER comp ALLER (bis) PLANTER det-MAIS
On l'a envoyé planter le mais. (GW2Y : III.06/062)

L'opacité des formes elles-mêmes (notamment dans des situations plus spontanées) empêche ici aussi la confrontation binaire de non-standard et

⁶ Nous ne discuterons pas ici la question de l'auxiliaire (cf. Ploog 2003) ; il n'est pas établi cependant que la forme relevée [E] corresponde au verbe *être*.

standard : [lYi] a pour concurrent une seule forme [l(')] qui possède deux valeurs, mais dont les valeurs s'organisent sur des axes distincts, le premier fonctionnel – l'opposition direct/indirect – le second « socio-stylistique » – l'opposition standard/non-standard.

On peut ainsi penser que [l] a perdu une partie de sa valeur interne en non-standard. Mais comme dans l'exemple 6, il s'agit régulièrement – à deux exceptions près qui présentent d'autres irrégularités - d'un argument doté du trait [+humain]. L'opposition fonctionnelle [±direct] serait ainsi résorbée par un seul paradigme réservé aux seuls les constituants [+animé] et véhiculant l'information [±pluriel]. Les constituants clitiques des arguments secondaires de la 3^e personne s'aligneraient ainsi sur les 1^e et 2^e personnes.

Cette hypothèse est corroborée par une autre observation, faite sur le fonctionnement du constituant ÇA : la majeure partie des référents nominaux inanimés se réalise par ÇA au lieu de respecter le genre grammatical standard ; ÇA en fonction d'objet n'est pas relayé par le clitique [l] ; enfin, ÇA représente le constituant sujet le plus fréquemment marqué Ø. Tout se tient...

5. Perspectives : l'évaluation des données

La généralisation d'observations faites sur un seul corpus reste délicate.

Les exemples précédents indiquent que la fréquence de séquences non standard pouvait être conditionnée par des facteurs autant externes (socio-interactionnels) qu'internes (structuraux). C'est sur ces deux plans qu'il convient d'évaluer leur représentativité à l'échelle du système linguistique auquel elles appartiennent. Avec un référencement consciencieux du corpus, la représentativité sociale, ou discursive, d'une forme s'établit sans trop de heurts. Elle doit cependant être distinguée de sa valeur linguistique, ou systémique, dont l'identification soulève des questions auxquelles il semble plus délicat de trouver des réponses.

5.1. Représentativité

Si un trait non standard possède un terrain social (propre ou non), son emploi doit être doté d'une *certaine* stabilité. Comme nous l'avons vu en 4.2, sa distribution sur un groupe de locuteurs ou un type de discours peut indiquer son périmètre d'emploi.

Le nombre d'occurrences d'une structure relevées dans un corpus peut fournir une première indication sur son degré de normalité. Mais cette mesure reste très peu fiable, à moins d'en constater la présence écrasante. Il convient de relativiser la fréquence absolue par des indices proportionnels. Si la distribution observée dans le corpus se raisonne dans un continuum des pratiques – par un nombre d'occurrences plus ou moins élevées, ressortant de circonstances d'élocution plus ou moins spécifiques - la description doit intégrer une articulation graduelle des phénomènes. La représentativité

interactionnelle d'une structure au sein du corpus peut se schématiser comme suit :

Tableau 5

DEGRE DE REPRESENTATIVITE	NOMBRE D'OCCURRENCES	NOMBRE DE LOCUTEURS	TYPES DE DISCOURS	STATUT DU TRAIT NON STANDARD
0	1	1	1	accidentel (hapax)
1	≥1	1	1 à n	idiolectal (psycholing)
2	≥1	≥1	1 à n	restreint (interactionnel)
3	≥1	≥1	≥1	généralisé (social)

Pertinents pour l'étude du non-standard sont les degrés de représentativité 2 et 3.

Le critère minimal, le plus facile à établir, semble être la répartition des occurrences sur les locuteurs. La division du nombre total d'occurrences par le nombre de locuteurs ayant produit la structure fournit un indice de représentativité global. Par exemple, on recense 40 occurrences de la structure investie dans l'exemple cité sous (4) :

(4) isieruj l episie ve
 ICI sg-ETRE ROUGE coord ICI sg-ETRE VERT
Ici c'est rouge et puis ici c'est vert (C06Y:VI.09/032)

Produite par 17 locuteurs différents, cette structure voit son existence dans le continuum formel largement validée. Les caractéristiques interactionnelles des occurrences permettent ensuite de la situer dans le continuum social, toujours à l'échelle du corpus constitué.

Le dernier niveau de représentativité consiste à situer le corpus d'étude par rapport aux caractéristiques du corpus social : la projection des caractéristiques des locuteurs sur les caractéristiques générales de l'échantillonnage du corpus peut permettre d'évaluer sa représentativité réelle à l'échelle de la communauté linguistique. Ce cas est illustré par le dernier type structurel recensé dans le tableau 1 (*cf. supra*, type SS) :

(5) mēnã | ləmœsjø iletepa øø
 ALORS det-MONSIEUR(i)-sj LUI(i)-sj ETRE-passé neg HEUREUX
Alors le monsieur il était pas heureux. (GM1X : III.01/162)

Il correspond au cas de structures largement répandues en français commun, mais liées dans des situations d'emploi restreintes – et exclues des

considérations normatives (un échelon supplémentaire de représentativité serait l'intégration de ce non-standard généralisé au standard, par voie de décision ou par son utilisation indifférente, non marquée). Ce type fonctionnel n'est cependant pas envisagé comme une construction non-standard dans notre perspective.

5.2. Variation et syntaxe

Une structure non standard se caractérise par la divergence d'au moins un trait constitutif du standard. L'apparition d'un trait non standard peut sous-tendre différents mécanismes, qui se combinent entre eux :

- lorsqu'une nouvelle forme émerge dans le système, on la qualifiera tout d'abord de trait non standard « de parole » (type NS) : l'ajout d'un terme ne constitue pas en soi une restructuration, mais peut entraîner des réanalyses pour aboutir à une structure NN ou SN ; ceci est illustré par les séquences à sujet-Ø ;
- le trait non standard restructurant ; l'apparition d'une nouvelle opposition formelle dans le paradigme réorganise un micro-système entier : la distribution des clitiques objet selon le trait ±[animé] ;
- le trait non standard innovant n'est pas construit sur un élément du micro-système standard ; il consiste en un apport formel à l'un des paradigmes du système : le cas du « locatif-sujet ».

Les rares travaux en la matière⁷ montrent un désaccord notoire sur ce qu'on aurait à appeler la *variation syntaxique*. Winford (1984) par exemple propose de distinguer quatre types de variation morphosyntaxique, en plus de deux purement syntaxiques, en croisant les paramètres phonologique / syntaxique / catégoriel ; par exemple, l'omission du complémenteur serait « surface-structural », d'ordre phonologique en français montréalais et d'ordre syntaxique en anglais. A cet égard, Cheshire (1987) juge prématurée la proposition d'une classification aussi précise des structures, les fondements méthodologique pour leur approche n'étant pas établis. En effet, les niveaux d'analyse sont imbriqués et leur articulation dépend largement de la langue considérée.

La catégorisation de la structure « locatif-sujet », illustrée en (4), reste à faire. La définition du *sujet* en français est d'ordre syntaxique ; or, on y fait référence en réalité à quatre niveaux d'analyse distincts, largement congruents dans le cas précis du sujet français (en tout cas tel qu'il est conçu par la grammaire), comme illustré dans le tableau 6 :

⁷ Winford (1984), Romaine (1984), Cheshire (1987).

Tableau 6

NIVEAU D'ANALYSE	VALEUR	INDICE	REPERAGE (ET DIFFICULTES)	UNITE D'OBSERVATION
morpho- logique	actant nominal, « privilégié »	accord verbal	immédiat (oral : oppositions réduites)	unité phonologique (mot verbal)
syntaxique	cas	position, marquage syntaxique obligatoire	immédiat (parlé : fluctuations dans la linéarisation)	unité microsyntaxique (et macrosyntaxique)
prédicatif	agent, typiquement +hum/+animé	traits sémantiques, commandés par le prédicat	empirique	unité prédicative (composantes lexicales)
pragmatique	thème, ±focus	hiérarchisation informationnelle	empirique	configuration, discours, contexte spatio-temporel

La terminologie grammaticale du français prévoit ainsi deux interprétations structurales potentielles, toutes deux incompatibles avec la grammaire du français standard - un sujet-Ø (a) ou un sujet adverbial (b) :

(4) isieruj l episie ve

(a) ICI Ø-sj sg-ETRE ROUGE coord ICI Ø-sj sg-ETRE VERT

(b) ICI-sj sg-ETRE ROUGE coord ICI-sj sg-ETRE VERT

La problématique se situe-t-elle sur le plan catégoriel ou relationnel ?

L'exemple suivant avec LA est symétrique au précédent avec ICI, à la différence près qu'il comporte un sujet phonétique, le [s] :

(7) laseruj

LA(i?) ÇA(i?)-sj ETRE-sg-pres ROUGE

Là c'est rouge (C09X:VI.04/082)

ICI et LA sont des référents inanimés, neutres, très proches de ÇA, et on peut poser que la pronominalisation des locatifs s'opère de manière analogue à celle des objets inanimés. Il n'y a alors plus guère de raison de traiter ces locatifs comme adverbiaux ; en tant que nominaux, ils peuvent fonctionner comme sujet. La restructuration serait alors d'ordre catégoriel (morphologique) et non relationnel (syntaxique). Mais l'analyse ne s'achève pas là, car deux problèmes persistent.

Si l'existence d'un deuxième terme, supposé *équivalent*, nous a permis ici d'approcher les contours structuraux du phénomène, elle permet aussi d'établir le « poids » de la variante sur le système (cf. section 2) ; or, pourquoi la séquence (4) serait-elle plus exactement proportionnelle à *ici c'est rouge* qu'à *ceci est rouge* ? L'opposition binaire de deux termes s'avère à nouveau peu réaliste.

Si le sujet en français standard représente l'élaboration grammaticalisée d'un référent thématique, convient-il de traiter les phénomènes observés en termes de *sujet* ? La limitation à un objet donné constitue l'intérêt même de la terminologie ; l'appliquer à d'autres objets revient à canaliser leur description. La terminologie, qui n'est pas *a priori* inappropriée pour rendre compte de ces structures non standard, influe du moins directement sur l'interprétation des structures.

5.3. Vers la révision des outils d'analyse

Contrairement à une opinion assez répandue, *tout* n'est pas possible en non-standard. Si les traits les plus caractéristiques apparaissent plus fréquemment dans des situations de parole à pression normative réduite, les apparentes libertés s'articulent bien dans un environnement structurel, *plus* ou *moins* restreint, mais identifiable et descriptible.

Les outils de description étant conçus pour d'autres chantiers, la détermination des conventions d'annotation du corpus est une étape majeure dans la description des données non standard. Tout comme le recours à la transcription phonétique permet de préserver une partie inestimable (dans les deux sens du terme) de la variation, la catégorisation des structures ne peut se faire qu'à l'issue de la description.

La terminologie relative aux inventaires structuraux employée ici reste très *standard* : ce choix d'utilisation reflète la grille de lecture qui reste - qu'on le veuille ou non - le point de départ de la description. Il semble préférable par ailleurs, pour éviter une opacité accrue, de recourir aux termes partagés, quitte à en dégager les limites.

La première description est contrastive, inévitablement, par le constat que telle ou telle caractéristique du standard ne s'applique pas. Mais la terminologie investie doit être considérée comme provisoire, révoquant à l'issue de l'analyse : c'est une contrainte de travail d'avoir à envisager le non-standard en le raisonnant « en creux » du standard ; or, le tout est de dépasser ce prisme imposé *a priori*. L'approche peut être vue comme correspondant formel de la notion de *diglossie* au sens de Wald (1986), en tant que perception différentielle d'une distance interlinguistique « élastique », ou ce serait plutôt le standard qui possède un statut spécifique parmi les pratiques multiples et variables.

En l'occurrence, il convient de s'interroger sur la pertinence d'étiqueter comme *sujet* des séquences qui n'en suivent pas les contraintes définitoires. La catégorisation repose sur l'identification des niveaux d'analyse concernés

par un phénomène observé. Leur impact relatif n'étant pas prévisible, seul un étiquetage multi-niveaux peut rendre justice à l'hétérogénéité des données, en exploitant les indices des différents niveaux pour les combiner ensuite :

- un premier niveau linéaire (la succession phonique et l'ordre des « mots »), comme l'illustre l'influence de l'environnement linguistique immédiat pour le sujet \emptyset ;
- un second niveau relationnel (la variation morphologique, les contraintes prédicatives et la coréférenciation), illustré par les facteurs favorisant le « locatif-sujet » ;
- un troisième niveau pragmatique (la hiérarchisation informationnelle, la référence), comme les valeurs [\pm animé] et [\pm actif] pour le constituant \emptyset en général.

Dans la pratique, la démarche se heurte à la variété des paramètres d'emploi, dont le panachage très variable multiplie de manière exponentielle les approches nécessaires. Les séquences « pseudo-standard » (comme l'exemple 6) témoignent de la vigilance nécessaire lorsqu'on isole un point du système, car la cohésion des formes n'apparaît pas immédiatement

La réflexion doit intégrer autant l'observation d'occurrences concrètes que celle de schèmes structuraux, lier dans la description les formes aux contextes, et alterner entre caractéristiques non-standard et standard. Bon nombre des propriétés structurales restent opaques dans un premier temps, une démarche par « boucles » s'impose. L'analyse est rendue difficile en outre par l'impossibilité d'une annotation automatique du corpus ; on ne peut étiqueter avant d'avoir effectué l'analyse et l'on ne peut effectuer l'analyse qu'en connaissance de cause (des données). L'étiquetage doit s'affiner au fur et à mesure des spirales argumentatives effectuées ; il restera lacunaire, pendant plus ou moins longtemps.

6. Et si l'hétérogénéité était un objet de choix pour la linguistique générale ?

L'objectif de ces réflexions était d'illustrer dans quelle mesure le non-standard, malgré toutes les difficultés méthodologiques qui lui sont propres, peut constituer un objet d'étude fédérateur.

La mise à l'épreuve sur un terrain original - non pas nouveau, mais en friche - d'outils anciens donnera une impulsion au renouvellement de la linguistique générale. Les limites de toute description linguistique – l'imbrication des niveaux d'analyse, le poids normalisateur du métalangage et de l'orthographe – constituent le point de départ pour l'approche du non-standard. En tant qu'approche « réaliste » des langues naturelles, la réflexion sur le non-standard apportera sa contribution au développement des bases théoriques pour le traitement automatique des corpus oraux, qui en est encore à ses balbutiements.

Références bibliographiques

- Chafe, Wallace L. (1982), « Integration and involvement in speaking, writing and oral literature », in: Tannen, Deborah (ed.), pp. 35-53.
- Cheshire, Jenny (1987), « Syntactic variation, the linguistic variable, and sociolinguistic theory », in *Linguistics* 25, pp. 257-282.
- Garcia, Erica (1997), « La portée de la variabilité », in *Langue française* 115, pp. 30-47.
- Givon, Talmy (1992), « The grammar of referential coherence as mental processing instructions », in *Linguistics* 30, pp. 5-55.
- Givon, Talmy (1976), « Topic, pronoun, and grammatical agreement », in Li (ed.), pp. 149-187.
- Li, Charles (ed.) (1976), *Subject and Topic*. New York, Academic Press.
- Ploog, Katja (à paraître), « La restructuration de la diathèse en français abidjanais : l'exemple des auxiliaires verbaux », in *Actes du XXIIIe colloque de linguistique et de philologie romane (CILFR)*, Salamanca, sept.2001. Tübingen, Narr.
- Ploog, Katja (2002a), « Non-standard et variation du français en Côte d'Ivoire. Entre descriptions et pratiques », in *Traverses* 3, pp. 73-94.
- Ploog, Katja (2002b), « Une promotion pour les locatifs en abidjanais ? », in *Revue romane* 37/1, pp. 31-43.
- Ploog, Katja (2002c), « Pour une analyse syntaxique de corpus non-standard : ellipses et structures prédicatives », in Pusch, Claus D. / Raible, Wolfgang (eds), *Romanistische Korpuslinguistik. Korpora und gesprochene Sprache / Romance Corpus Linguistics. Corpora and Spoken Language*. Tübingen: Gunter Narr (coll. ScriptOraIia), pp. 411-424.
- Romaine, Suzanne (1984), « On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory », in *Folia Linguistica* 18, pp. 409-437.
- Tabi-Manga, Jean (1992), « Le français d'Afrique (bilan, évaluation et stratégie prospective) », in *Etudes linguistique appliquée*, 88, pp. 111-116.
- Tannen, Deborah (1982), « The oral/literate continuum in discourse », in Tannen, Deborah (ed.). pp. 1-16.
- Tannen, Deborah (ed.) (1982), *Spoken and Written Language: Exploring Orality and Literacy*. (Advances in Discourse Processes 9). Norwood : Ablex.
- Wald, Paul (1986), « Diglossia applied: Vernacular mixing and functional switching with Bangui Yakomas. », in Fishman, Joshua (et al., eds.) *The Fergusonian impact*, New York : Mouton de Gruyter, vol. 2, pp. 417-430.
- Winford, Donald (1984), « The linguistic variable and syntactic variation in creole continue », in *Lingua* 62, pp. 267-288.