

Projet de dictionnaire électronique : nouvelles perspectives pour la lexicographie africaine

Logiciels Opera et OmegaT

Mohomodou Houssouba,
Ecrivain, Bâle, Suisse

Résumé

Comment harmoniser l'orthographe d'une langue africaine à multiples dialectes qui n'a connu qu'une courte histoire d'écriture ? Le projet de dictionnaire électronique en langue songhay utilise les schémas bien établis pour accélérer la standardisation orthographique et faciliter la création d'outils de correction automatique. Mais, la lexicographie informatisée dépasse le cadre des mots ; elle joue un rôle central dans le dispositif de localisation des langues « peu dotées » en outils informatiques. La traduction du code souche du logiciel Opera en songhay est retenue comme une approche pragmatique pour atteindre plusieurs objectifs à court et moyen terme.

1. Introduction

Le projet d'un dictionnaire électronique en langue songhay a germé en mi-2004 avec une présentation théorique [Fad Seydou et Mohomodou Houssouba] à l'édition précédente du Symposium malien sur les sciences appliquées (MSAS 2004). La session « Education et langues » aurait connu un afflux inattendu grâce à la participation de l'Académie africaine des langues basée à Bamako et au réseau de chercheurs maliens travaillant sur la traduction des termes mathématiques et la construction d'un prototype de traducteur multilingue avec le bambara comme langue de référence. L'engouement suscité par les résultats-avouons le, jusque-là, rudimentaires dans le traitement informatisé des langues maliennes-démontre quand même un double enjeu : un intérêt pour la promotion de ces langues et une méconnaissance palpable du mécanisme d'application des nouvelles technologies à celles-ci. La poursuite de l'expansion des échantillons traités a fait une percée avec l'utilisation de logiciels de traduction.

Après différentes tentatives et avec l'aide de l'informaticien linguiste canadien Chris Harvey (www.languagegeek.com), l'idée de traduire les codes d'un logiciel libre polyvalent a pris forme; le traducteur Mozilla testé à partir de février 2005 a été abandonné en faveur d'un format plus simple et portable, en l'occurrence le logiciel Opera qui combine un outil de navigation, un client de messagerie et un support pour forum. Opera présente un avantage supplémentaire pour le néophyte : le code-souche peut être récupéré et traduit en fichier texte-ce qui le rend extrêmement transportable. Le travail initial débute avec le code de la version Opera 2.0 en avril 2005. La traduction concerne plus de 3500 segments requis pour générer une version « beta » de l'interface songhay du logiciel Opera (Version 9.25, 2006).

Le projet lexicographique intervient dans un contexte caractérisé par l'existence de multiples orthographes, d'où le besoin d'harmoniser les formes pour faciliter la création d'outils de correction automatique pour la langue songhay qui couvre un grand espace régional le long du fleuve Niger, même si sa population de locuteurs, estimée à quatre millions, dont la moitié au Niger, reste modeste. Cette population occupe également une aire francophone constituée de langues « peu dotées » en instruments d'accès aux nouvelles possibilités offertes par l'expansion de la Toile mondiale. Le projet de dictionnaire de référence songhay-français - français-songhay est une initiative citoyenne d'une part et d'autre part

un cadre de coopération sous-régionale à moindre coût pour produire un document d'utilité collective sous forme de livres, d'interfaces totalement « localisées » dans la langue africaine et un pont solide entre cette langue et la langue française.

2. Tâches principales

En premier lieu, il s'agit (i) de terminer la traduction du code souche du logiciel Opera (Versions 2.0 à 9.25) en songhay, (ii) de générer une interface expérimentale à l'aide du logiciel de traduction OmegaT qui permet aussi de créer la base d'un dictionnaire électronique à multiple usage : impression ou mise en ligne et (iii) de tester la plate-forme « beta » en songhay-français-anglais sur le terrain.

3. Étapes du travail

Le travail initial entamé par deux traducteurs continue dans la phase de développement au sein du groupe de travail international constitué de locuteurs qui ont le songhay comme première langue. Le groupe élargi a pour missions : (i) la relecture et correction de la traduction de base, (ii) l'examen des contributions des groupes de travail régionaux, surtout le tri des néologies des NTIC, (iii) la compilation du dictionnaire général, et (iv) l'édition finale du dictionnaire général.

La traduction de base sera complétée au cours du premier trimestre 2008 ; l'organisation du travail sur le terrain - dont les ateliers à Gao et Niamey - est prévue pour la période août-décembre 2008.

4. Résultats attendus

L'objectif principal est la production (i) d'un dictionnaire de référence songhay-français - français-songhay, (ii) d'un lexique technique pratique et moderne avec des articles songhay-français-anglais, et (iii) de création d'une expertise pouvant servir au niveau sous-régional à l'harmonisation des orthographes et méthodes d'enseignement et de recherche en langue songhay. Le lexique des termes actuels est inspiré, entre autres, par les *Vocabulaire des techniques de l'information et de la communication. Enrichissement de la française* (2005) et *Vocabulaire de l'audiovisuel et de la communication. Enrichissement de la langue française* (2006), édités par la Commission générale de terminologie et de néologie, attachée au bureau du Premier ministre français, en collaboration avec les instances de la Francophonie. Pour un début, le lexique NTIC s'est construit à travers la traduction du logiciel de navigation, de messagerie et support de forum Opera ; c'est à dire, 3500 segments qui contiennent l'essentiel des termes que l'utilisateur de l'ordinateur, surtout dans l'expérience de la navigation et de la messagerie, rencontre sous forme de boîtes de dialogue. Cette base sera élargie par d'autres codes-souches de logiciels libres à traduire. La traduction manuelle initiale est contrôlée par le logiciel de traduction OmegaT qui, sans automatiser la traduction, arrive à la rendre intuitive à partir d'une certaine base de segments traduits.

Ce dictionnaire moderne songhay-français - français-songhay sera élaboré à partir de corpus existants qui appartiennent à deux catégories principales : les descriptions de la langue et lexiques détaillés développées par les administrateurs et missionnaires intéressés par les idiomes songhay durant toute la période coloniale. Parmi ceux-ci, le livre d'Antoine Prost, *La langue songhay et ses dialectes*¹ reste l'un des plus cités. L'autre catégorie regroupe les manuels et lexiques produits dans des instituts de recherche linguistique du Niger, Mali, Burkina et Bénin, en général par des locuteurs du zarma-songhay, souvent dans le cadre des programmes d'alphabétisation ou de promotion des langues nationales comme langues d'instruction à l'école. En tout, une trentaine

¹ IFAN, Dakar, 1956.

de publications en songhay forment la base d'extraction du corpus élargi.

Bien sûr, d'autres dictionnaires récents comme la série dialectale de Jeffrey Heath (Djenné-Tombouctou-Gao) ressemblent au premier groupe car ils sont produits par des linguistes occidentaux qui ont passé beaucoup de temps à faire un inventaire diversifié du patrimoine linguistique songhay. Ces documents sont très utiles et seront utilisés comme références-clés durant le travail de traduction.

Le document final sera composé à partir des lexiques, glossaires et dictionnaires existants aussi bien que de nouvelles traductions et créations de mots réalisés sur le terrain, d'où la nécessité de conduire les ateliers pour corriger les traductions et harmoniser les orthographes. Cette étape prolonge la collaboration intermittente déjà existante entre des pays comme le Mali et le Niger. L'harmonisation transfrontalière, dans la mesure du possible, sera effectuée pour faciliter la création d'un clavier adapté à la langue dans son environnement ; c'est-à-dire, en considérant les langues partageant les territoires à populations « songhayophones ». Par exemple, le zarma et le haoussa au Niger, le songhay et le bambara au Mali. L'harmonisation facilite la création de dictionnaires électroniques et de multiples programmes de traitement de la « langue naturelle ».

La spécificité de ce projet est de faire de la lexicographie avancée un élément déterminant de l'introduction d'une langue africaine dans l'espace Internet, selon le procédé explicite développé par Marcel Kidi-Dikiri². Chaque étape de composition résulte en produit à usage dual : conventionnel (physique) et électronique.

5. Évaluation

Le lexique, le dictionnaire, aussi bien que les interfaces songhay pour ordinateurs seront testés et évalués dans les quatre (4) centres d'activité des groupes de travail. Le projet aura abouti si, en premier lieu, les élèves et enseignants, pourront disposer des outils suivants :

- un dictionnaire de référence songhay-français - français-songhay d'environ 20 000 articles, accompagnés d'un précis de prononciation et de grammaire et un aperçu des variations dialectales, en vue de réaliser l'harmonisation orthographique la plus aboutie jusque-là
- un lexique de poche des termes scientifiques et techniques, avec un glossaire des NTIC élaboré dans OmegaT
- une interface opérationnelle en songhay pour le traitement de texte (dictionnaire électronique), la navigation sur Internet, la messagerie électronique, et le support pour forum électronique (Opera 9.25).
- quelques accessoires localisés dont les claviers Unicode configurés pour le songhay et les langues voisines
- un site interactif en ligne et à jour pour répondre aux questions courantes d'usage et de traduction songhay-français-songhay

6. Collaboration

Le projet engage plusieurs groupes de personnes ressources mandatées sur différents sites de compilation et de traduction. Un volet permettra de rassembler les collaborateurs en ateliers pour harmoniser les traductions. Il s'agit des quatre groupes régionaux ainsi constitués : 1. Groupe de Bamako, 2. Groupe de Niamey, 3. Groupe de Tonka (Tombouctou), Groupe de Gao.

Ces groupes régionaux travaillent avec un groupe de chercheurs (linguistes, informaticiens, spécialistes de l'éducation et de la communication) qui supervisent les aspects scientifiques et

² Diki-Kidiri, Marcel, *Comment assurer la présence d'une langue dans le cyberspace ?*, UNESCO, Paris, 2007.

technologiques du projet. Mais, au-delà de cette coopération explicite, les différents groupes établissent des réseaux propres selon leurs besoins immédiats ou futurs. Cette démarche reflète une quête de décentralisation au sein d'un projet de recherche basé sur la gestion consensuelle d'importantes variations dialectales et idiomatiques. Le test de l'approche durant la phase de développement-allant de la mise en ligne de l'interface localisée en songhay à la production du lexique et du dictionnaire électroniques-doit confirmer sa pertinence ou son inadaptation. Tout de même, l'on ne peut trop accentuer l'importance du consensus au niveau régional, entre les instituts spécialisés de pays comme le Mali et le Niger, pour effectuer une harmonisation pérenne de l'écriture songhay.

7. Conclusion

L'actuel projet illustre la versatilité et l'imprévisibilité de la recherche en langues africaines. Dans l'ensemble, ces langues manquent encore de codification compréhensive ; ce qui complique énormément la création d'outils de correction électronique, même au seul niveau du mot. L'orthographe aussi bien que la grammaire sont récentes et souvent peu connues des locutrices et locuteurs même. Par comparaison, les langues internationales d'aujourd'hui ont des siècles de développement derrière elles. Les questions de forme et de structures ont été posées, discutées et répondues par étapes. L'entrée dans le nouvel ordre virtuel pose un choc moindre dans la mesure où de nombreuses ressources existent déjà et peuvent être mises en ligne et consultées par de nouveaux utilisateurs, souvent à distance de la source.

Cependant, on remarque que beaucoup de langues moins fournies en ressources arrivent à « squatter » l'Internet à travers des actions simples comme la lexicographie et la traduction de logiciels. L'effort de localisation du logiciel Opera nous a amené à essayer d'autres logiciels libres comme Mozilla, en mesurer les complexités et choisir les chemins les moins tortueux. S'il favorise un usage polyvalent aux mains des maîtres et élèves, il pourrait servir de repères pour le développement d'autres langues africaines pour occuper l'espace virtuel qui devient chaque un enjeu stratégique et économique central pour toutes les langues, cultures et nations.