



# Validité du corpus et évolution constatée, un exemple en basque,

Jean-Baptiste Coyos

► **To cite this version:**

Jean-Baptiste Coyos. Validité du corpus et évolution constatée, un exemple en basque,. La Linguistique Journal of the International Society for Functional Linguistics, 2011, pp.51-66. <artxibo-00660819>

**HAL Id: artxibo-00660819**

**<https://artxiker.ccsd.cnrs.fr/artxibo-00660819>**

Submitted on 17 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# VALIDITÉ DU CORPUS ET ÉVOLUTION CONSTATÉE, UN EXEMPLE EN BASQUE

par Jean-Baptiste Coyos  
*Iker UMR 5478-Bayonne*

*In this article a point is made about a research which relates to the variation and the evolution of the indication of the syntactic functions for certain transitive verbs out of souletin Basque. The principal question relates to the nature and the validity of the corpus built in order to update the postulated evolution, in diachrony as in synchrony. It is a reflexion of methodological and epistemological nature which is proposed concerning a descriptive and empirical-type research.*

## INTRODUCTION

Dans cet article, nous allons faire un point sur les dernières avancées d'une recherche qui concerne la variation et l'évolution de l'indication des fonctions syntaxiques pour certains verbes transitifs en basque souletin<sup>1</sup>. Le principal questionnement va porter sur la nature et la validité du corpus construit afin de mettre à jour l'évolution postulée. C'est donc une réflexion de nature méthodologique et épistémologique qui est proposée concernant une recherche à caractère essentiellement descriptif et empirique.

1. Ce travail vient à la suite d'une communication faite en 2010 au XXXIII<sup>e</sup> Colloque international de linguistique fonctionnelle à Corfou pour le thème Dynamique et évolution en syntaxe. Celle-ci a été elle-même précédée de plusieurs communications et articles publiés. Pour une synthèse en français voir Jean-Baptiste Coyos, « Transitivity et changement linguistique en basque souletin : le cas de certains verbes bivalents », *Dynamique et changement en syntaxe*, Françoise Guérin et Denis Costauvec (dir.), Bruxelles, EME Éditions, 2009, 95-110.

Le dialecte souletin est l'un des deux dialectes basques parlés à l'heure actuelle en France dans la partie ouest du département des Pyrénées-Atlantiques. La Soule est la plus petite des sept provinces basques la plus à l'est. Elle compte un peu plus de 13 000 habitants et environ 55 % de basco-phones, contre 25,7 % de basco-phones dans l'ensemble du Pays basque (665 700 locuteurs en tout)<sup>2</sup>. Dans la tradition, le souletin est reconnu comme l'un des quatre dialectes littéraires avec le labourdin, le guipuscoan et le biscayen. Jusqu'au XIX<sup>e</sup> siècle, il s'agit d'une littérature essentiellement à caractère religieux. L'Académie de la langue basque, *Euskaltzaindia*, a à partir des années 1960 élaboré une forme standard de l'écrit, *euskara batua*, le basque unifié, qui s'est diffusée assez largement des deux côtés de la frontière dans l'enseignement, les grands médias et l'administration de la communauté autonome basque en Espagne. Le basque unifié cohabite sous forme orale également, de façon plus ou moins harmonieuse, avec les parlers locaux là où ils n'ont pas disparu. Il a une légère influence en Soule, surtout dans l'enseignement, les médias et le parler de certains locuteurs lettrés.

#### LE PHÉNOMÈNE À OBSERVER ET À DÉCRIRE : SYNCHRONIE DYNAMIQUE ET DIACHRONIE

Le phénomène à observer, à décrire et ensuite à tenter d'expliquer concerne une trentaine de verbes transitifs qui se construisent actuellement en basque souletin de deux façons différentes, parfois trois, pour un sens identique selon les informateurs<sup>3</sup>. Tous les informateurs n'utilisent

2. D'après l'enquête sociolinguistique réalisée en 2006 par le gouvernement autonome basque : *IV<sup>e</sup> enquête sociolinguistique 2006*, [www.euskara.euskadi.net](http://www.euskara.euskadi.net).

3. Voir la liste de ces verbes dans Jean-Baptiste Coyos, 2009, p. 97-99. L'analyse fine de la valeur sémantique de chacune des constructions montrerait peut-être, pour certaines d'entre elles, qu'elles ne sont pas totalement synonymiques. Il s'agit donc d'un postulat qui autorise la recherche que nous présentons. Concernant la recherche du conditionnement de cette variation dans la construction de ces verbes et de l'évolution postulée voir Jean-Baptiste Coyos, 2009. Il ne s'agit pas d'innovation, mais d'une

pas les différentes constructions pour tous les verbes. Dans le cas d'un verbe pouvant être construit de plusieurs façons avec un même sens par un même locuteur, on parlera de fluctuation de construction. Cette variation observée en synchronie, à l'heure actuelle, a amené à s'interroger sur la forme que pouvait avoir ce phénomène au cours de l'histoire de la langue. Dit autrement, à observer comment étaient construits ces verbes dans les textes écrits anciens connus à ce jour, s'ils se construisaient différemment, s'il y avait variation de construction et dans ce cas quelle construction prédominait.

#### *Différentes constructions pour un même verbe transitif*

Le basque est une langue à cas avec un marquage de type ergatif majoritaire. En relation avec les déterminants nominaux du verbe porteurs d'un cas, des indices de personne, distincts des pronoms personnels, sont affixés à l'auxiliaire ou au verbe synthétique. Ils peuvent référer à l'agent (cas ergatif ou absolutif), au patient (cas absolutif ou autre) et au destinataire ou autre objet secondaire (cas datif). Parfois s'ajoute un indice qui réfère à l'interlocuteur (indice d'allocutivité). En souletin, l'indice de personne coréférent du nominal au datif n'est pas obligatoire dans le verbe.

Nous appelons dans ce travail verbes transitifs, les verbes qui impliquent deux déterminants nominaux. C'est une définition large de la transitivité que nous utilisons. En syntaxe, on a des déterminants nominaux et un prédicat verbal, en sémantique des arguments et un procès. Un procès qui nécessite deux arguments sera réalisé généralement en syntaxe par un prédicat verbal et deux déterminants nominaux. Bien sûr, chaque argument a un caractère plus ou moins agentif, et le procès peut ainsi être classé dans une échelle empirique de transitivité sémantique.

substitution de construction pour certains des verbes considérés par attraction analogique en première analyse. Plus largement, on a un ensemble de facteurs linguistiques et extralinguistiques qui peuvent expliquer le phénomène et dont le poids respectif reste à définir au cas par cas, verbe à verbe.

Voici pour illustrer le phénomène, l'exemple de deux verbes, *orit* « se rappeler » et *beha* « écouter », qui peuvent se construire actuellement chacun de deux façons différentes, parfois par un même locuteur<sup>4</sup> :

- |       |   |                         |  |
|-------|---|-------------------------|--|
| (1 a) | <i>historia-z</i><br>l'histoire + Inst  | <i>orit</i><br>rappeler | <i>n-iz</i><br>1Abs + être                       |
| (1 b) | <i>historia</i><br>l'histoire + Abs (Ø) | <i>orit</i><br>rappeler | <i>d-ü-t</i><br>3Abs + avoir + 1 <sup>er</sup> g |

Le sens est le même pour les deux phrases : « Je me souviens de l'histoire. »

En (1 a), le verbe *orit* « se rappeler » est construit avec l'auxiliaire verbal *izan* « être » et deux déterminants nominaux aux cas absolutif et instrumental. En (1 b), il est construit avec l'auxiliaire verbal \**edün* « avoir » et deux déterminants nominaux aux cas ergatif et absolutif.

- |       |   |                                      |  |
|-------|---|--------------------------------------|--|
| (2 a) | <i>txori-a-ri</i><br>oiseau + Déf + Dat | <i>beha-tzen</i><br>écouter + Inacc  | <i>n-iz</i><br>1Abs + être                       |
| (2 b) | <i>txoria</i><br>oiseau + Déf + Abs (Ø) | <i>beha-tzen</i><br>Rappeler + Inacc | <i>d-ü-t</i><br>3Abs + avoir + 1 <sup>er</sup> g |

Le sens est le même pour les deux phrases : « J'écoute l'oiseau. »

En (2 a), le verbe *beha* « écouter » est construit avec l'auxiliaire *izan* « être » et deux déterminants nominaux aux cas absolutif et datif. En (1 b), il est construit avec l'auxiliaire \**edün* « avoir » et deux déterminants nominaux aux cas ergatif et absolutif.

4. Liste des abréviations utilisées : Abs = cas absolutif de forme vide (Ø), Dat = cas datif, Déf = défini, Erg = cas ergatif, Inacc = inaccompli, Inst = cas instrumental, Plur = pluriel, 1Abs = indice verbal de 1<sup>re</sup> personne d'absolutif, 3Abs = indice verbal de 3<sup>e</sup> personne d'absolutif, 1<sup>er</sup>g = indice verbal de 1<sup>re</sup> personne d'ergatif. Les exemples sont donnés dans l'orthographe établie par l'Académie de la langue basque.

*La perspective diachronique :  
la variation et l'évolution dans le temps*

Pour mieux comprendre la situation actuelle, nous avons mené une étude diachronique, afin d'observer dans le temps ce phénomène de variation.

Voici, à titre d'exemple, un cas de fluctuation de construction chez Martin Maister dans *Jesü Kristen imitazionea* « L'imitation de Jésus-Christ » datant de 1751. Page 99, le verbe *goza* « savourer, apprécier » est construit avec l'auxiliaire *izan* « être », un indice de personne d'absolutif (Abs) *z-* dans *zira* « vous êtes », et un déterminant nominal *alegrantziaz* « l'allégresse » au cas instrumental (Inst). *Alegrantziaz ardüra gozatüren zira* : « ... vous savourerez souvent l'allégresse. »

(3 a)	<i>alegrantzia-z</i>	<i>ardüra</i>	<i>goza-türen</i>	<i>z-ira</i>
	(alégresse + Déf) + souvent		savourer +	2Abs + être
	Inst		futur	

À la même page, ce même verbe est construit avec l'auxiliaire *\*edün* « avoir », un déterminant nominal au cas ergatif (Erg) *nork* « qui » et l'autre *gäüzak* « les choses » au cas absolutif (Abs). *Nork gäüzak (...) gozatzen beitätü...* « Qui puisqu'il savoure les choses... » :

(3 b)	<i>nor-k</i>	<i>gäüza_k</i>	<i>goza-tzen</i>	<i>beit-ü-tü</i>
	qui + Erg	chose + Plur +	savourer +	(car + il) +
		Abs (Ø)	Inacc	avoir + les

Cet auteur utilise donc à la même page un même verbe sans diathèse marquée, c'est la voix active dans les deux cas, avec deux auxiliaires et un marquage casuel différents. À l'heure actuelle, la construction avec l'auxiliaire *izan* « être » est très exceptionnelle avec le verbe *goza*, même si elle ne choque pas. On peut dire que le changement de construction est effectif.

D'une façon générale pour les verbes que nous étudions, les auteurs du xvii<sup>e</sup> siècle au début du xx<sup>e</sup> examinés utilisent la construction avec l'auxiliaire verbal *izan* « être » plus fréquemment que nos informateurs et que les auteurs

contemporains. Pourtant, la construction avec l'auxiliaire *izan* n'est pas systématique, dans les siècles passés, comme le montre l'exemple (3 b) de Maister avec le verbe *goza*. C'est donc plutôt une tendance générale que nous mettons à jour par la comparaison à l'usage actuel plus important de l'auxiliaire \**edün* « avoir », mais avec des variantes selon le verbe considéré et selon l'auteur.

### *La perspective synchronique dynamique*

La perspective synchronique dynamique s'impose dans le cadre de cette recherche puisqu'on a également actuellement des variations en fonction du locuteur et du verbe considérés, avec des tendances assez nettes selon la tranche d'âge des informateurs. « Il peut donc être indiqué d'opposer, à l'étude diachronique visant délibérément à comparer différents états successifs du même objet d'étude, une synchronie dynamique où l'attention se concentre, certes, sur un seul et même état, mais sans qu'on renonce jamais à y relever des variations et à y évaluer le caractère progressif ou récessif. »<sup>5</sup> Ainsi, à l'heure actuelle d'une façon générale les constructions avec l'auxiliaire *izan* « être » sont plutôt utilisées par les locuteurs âgés, comme en (1 a) et (2 a). Les constructions avec l'auxiliaire \**edün* « avoir » le sont plus facilement par les plus jeunes, comme en (1 b) et (2 b). Il ne s'agit que de tendances puisqu'un locuteur âgé peut avec tel verbe utiliser plutôt la construction avec l'auxiliaire \**edün* « avoir », alors qu'un plus jeune pourra utiliser plutôt la construction avec l'auxiliaire *izan* « être ».

### À PROPOS DE LA VALEUR D'UN CORPUS

Nous voulons observer à travers les siècles la construction des verbes qui actuellement présentent une variation du type

5. André Martinet, 1975, « Diachronie et synchronie dynamique », *Évolution des langues et reconstruction*, p. 9.

ci-dessus, pour savoir s'ils étaient déjà sujets à variation, lesquels, depuis quand, dans quelle proportion. Et ensuite étudier ces verbes de façon plus systématique en synchronie, à l'heure actuelle, afin de statuer sur la nature de cette évolution postulée. Il a fallu pour cela construire un corpus à partir de textes divers, écrits et oraux, qui pouvaient contenir des occurrences des verbes à tester. Il permet de travailler sur des données écrites réelles, attestées. Avant de présenter la façon dont le corpus a été établi et d'en discuter la valeur, il est nécessaire de faire quelques remarques d'ordre plus théorique.

La valeur d'un corpus n'est pas déterminée par sa forme, par la quantité de données qu'il contient, mais bien par sa capacité à permettre une interprétation claire, non ambiguë<sup>6</sup>. Le but de la constitution d'un corpus est de pouvoir reconstituer *a posteriori* des régularités de la langue qui ne sont perçues jusque-là qu'intuitivement. Dans notre cas, les régularités seraient des doubles constructions avec un sens identique pour un même verbe au cours de l'histoire de la langue et actuellement. La comparaison de divers états de langue à divers moments de l'histoire de la langue pourrait nous permettre ensuite de statuer sur une éventuelle évolution dans ce domaine.

Voici selon Bénédicte Pincemin quels sont les « critères de bonne constitution d'un corpus » : « *Conditions de signifiante* : Un corpus est constitué en vue d'une étude déterminée (*pertinence*), portant sur un objet particulier, une réalité telle qu'elle est perçue sous un angle de vue [...] (*cohérence*).

*Conditions d'acceptabilité* : Le corpus doit apporter une représentation fidèle (*représentativité*) sans être parasité par des contraintes externes (*régularité*). Il doit avoir une ampleur et un niveau de détail adaptés au degré de finesse et à la richesse attendue en résultat de l'analyse (*complétude*).

*Conditions d'exploitabilité* : Les textes qui forment le corpus doivent être commensurables (*homogénéité*). Le corpus doit

6. Ces remarques d'ordre épistémologique sont inspirées par le texte suivant : Bénédicte Pincemin, 1999, Construire et utiliser un corpus : le point de vue d'une sémantique textuelle interprétative, Atelier *Corpus et TAL : pour une réflexion méthodologique*, VI<sup>e</sup> Conférence annuelle sur le Traitement automatique des langues naturelles (TALN 99), Cargèse (Corse, France), 12-17 juillet 1999, Anne Condamines, Marie-Paule Péry-Woodley et Cécile Fabre (éds.), p. 26-36.

apporter suffisamment d'éléments pour pouvoir repérer des comportements significatifs (au sens statistique du terme) (*volume*). »<sup>7</sup>

C'est à partir de ces critères que nous allons analyser la valeur du corpus construit pour cette recherche, son adéquation en vue de la mise à jour de l'évolution supposée. Nous commençons par le corpus en diachronie, c'est-à-dire les formes attestées et retenues sur plusieurs siècles. Ce corpus n'est pas du même type que celui que nous avons pu constituer pour analyser le phénomène en synchronie.

#### LE CORPUS EN DIACHRONIE – OBSERVER UN PHÉNOMÈNE SUR PLUSIEURS SIÈCLES

##### *Description de la production écrite en souletin au cours des siècles*

En diachronie par principe, on ne peut s'appuyer que sur des données uniquement écrites. Et sur ce point, le basque souletin pose une difficulté de taille au chercheur. C'est la rareté des documents, leur petit nombre du moins. C'est au XVI<sup>e</sup> siècle, en 1545, que le premier livre connu en basque est publié, *Linguae Vasconum Primitiae* de Bernard Etxepare, en dialecte bas-navarrais. Le premier livre connu en souletin date lui du XVII<sup>e</sup> siècle. Athanase Belapeyre publie en 1696 *Catechima laburra eta Jesus-Christ goure ginco jaunaren eçagutcia, salvatu içateco* « Court catéchisme et la connaissance de notre seigneur Jésus pour être sauvé »<sup>8</sup>. Un ouvrage plus ancien que nous avons quand même étudié n'est pas tenu pour souletin, même si son auteur l'est. La langue qu'il utilise est mixte avec des apports d'autres dialectes, comme il le dit lui-même dans l'adresse au lecteur. Il utilise d'ailleurs les constructions avec l'auxiliaire \**edün* « avoir » plus que les autres auteurs étudiés.

7. *Ibid.*, p. 2 de la version Internet : [icar.univlyon2.fr/membres/bpincemin/biblio/pincemin\\_taln99.pdf](http://icar.univlyon2.fr/membres/bpincemin/biblio/pincemin_taln99.pdf).

8. Un autre texte en souletin publié vingt ans plus tôt pourrait lui être attribué : *Pronus singulis diebus Dominicis clarè & distinctè immediatè post Evangelium populo legendus* (1676).

Il s'agit de *Ontsa hiltzeko bidea* [Le chemin pour bien mourir] de Jean de Tartas publié en 1666. On a aussi 17 lettres manuscrites de correspondance datant de 1616 et 1617 d'un procureur du roi s'adressant en souletin à un maire en Navarre espagnole.

Les textes manuscrits de pastorales qui nous sont parvenus sont eux nombreux, dès le XVIII<sup>e</sup> siècle<sup>9</sup>. Il s'agit d'un genre littéraire particulier. Écrit en strophes de quatre vers avec ou sans rime, le texte de longueur variable contient plusieurs centaines de quatrains. Les vers sont courts et les constructions verbales que nous étudions ne se présentent pas toujours dans leur complétude : l'auxiliaire verbal est parfois omis, et les indices de personne manquent alors, les déterminants nominaux ne sont pas transcrits, et on ne peut que faire des suppositions sur le cas qui leur serait suffixé puisque seuls les indices référant au nominal à l'ergatif et à l'absolutif sont obligatoirement présents dans la forme verbale conjuguée.

Au total les textes du XVIII<sup>e</sup> siècle au milieu du XX<sup>e</sup> siècle, hormis ceux de pastorales, ne sont pas nombreux et il en va de même à l'heure actuelle. La publication en 2010 et 2011 de deux ouvrages datant du XVIII<sup>e</sup> siècle a été évidemment une bonne nouvelle. Ce sont comme pour tous ces textes anciens des réimpressions sous forme d'édition critique mise en orthographe actuelle. Il s'agit de *Jesü Kristen imitazionea* de Martin Maister (1751)<sup>10</sup> déjà cité et de *Egiategiren filosofo huskaldunaren ekheia* « La matière du philosophe basque Egiategi » de Jüsef Egiategi (1796 ou 1797)<sup>11</sup>.

Parmi cet ensemble de documents disponibles qui constituent l'« univers » de l'étude, peu sont numérisés, ce qui aurait facilité la recherche des formes à étudier. Pour constituer

9. La pastorale est un théâtre populaire en langue basque, psalmodié, codifié, et joué en plein air par un village différent chaque année. Le texte de pastorale le plus ancien dont on est sûr de la datation est de 1750.

10. Martin Maister, 2010 [1751], *Jesü Kristen imitazionea*, préparé par Jean-Louis Davant, Bilbo/Bilbao, Euskaltzaindia/Académie de la langue basque.

11. Jüsef Egiategi, 2011 [1796-1797], *Egiategiren filosofo huskaldunaren ekheia*, bigarren liburukia/deuxième volume, préparé par Txomin Peillen, Bilbo/Bilbao, Euskaltzaindia/Académie de la langue basque. Il ne s'agit pas d'une réédition d'un ouvrage existant, mais de la publication d'un manuscrit inédit. Txomin Peillen dans l'avant-propos fait part des difficultés qu'il a rencontrées. Le texte est souvent raturé, parfois le mot n'est pas lisible et certaines phrases ne sont pas compréhensibles.

notre propre corpus de constructions verbales, l'ensemble des occurrences repérées, nous avons retenu et analysé dix ouvrages. Deux sont du xvii<sup>e</sup> siècle, trois du xviii<sup>e</sup> dont les deux récemment publiés, deux du xix<sup>e</sup> et un du début du xx<sup>e</sup> siècle.

*Les critères de représentativité, de complétude, d'homogénéité et de volume*

Voyons maintenant quelle est la valeur du corpus constitué en nous basant sur les critères présentés plus haut. Il s'agit évidemment d'une évaluation basée sur notre compréhension de ces critères, et non celle que pourrait en faire leur auteur.

Les conditions de signifiante, à savoir la pertinence et la cohérence comme définies ci-dessus, semblent remplies. L'étude porte sur un objet particulier, déterminé, et elle vise à mettre à jour une réalité, des régularités linguistiques, à partir d'un certain point de vue. Nous ne nous y attardons pas. Le corpus est constitué d'occurrences de verbes conjugués et de leurs deux déterminants nominaux, noms, pronoms, formes verbales nominalisées.

La « représentativité » est-elle assurée ? Certes les formes à étudier sont relevées dans des documents authentiques reconnus par les spécialistes comme écrits en basque souletin. Mais que penser de la langue d'un siècle donné lorsqu'on ne connaît que deux, trois ou quatre ouvrages lui correspondant ? Un, deux ou même trois auteurs sont-ils représentatifs de l'état de la langue à un moment donné ? Ils permettent certes de couvrir un siècle, mais pour le souletin, nous n'avons aucun texte qui serait véritablement contemporain d'un autre. Il y a souvent plusieurs dizaines d'années entre chacun. Un auteur est d'abord représentatif si l'on peut dire de lui-même, sachant que ces textes de caractère religieux sont souvent des traductions. Voilà un autre facteur qui peut influencer sur la forme de la langue. Or, statuer sur l'évolution, la généralisation, dépend de la représentativité du corpus. Ce que l'on peut dire, c'est que le corpus constitué est représentatif des documents en notre

possession à l'heure actuelle. Ce sont ceux que les spécialistes ont pu recueillir jusqu'à présent. Il n'y en a pas d'autre connu. Et même si nous ne les avons pas tous étudiés, hormis les textes de pastorales, notre échantillon représente presque la totalité des textes connus.

La « régularité » semble assurée puisqu'il n'y a pas de contraintes externes, les formes verbales sont attestées dans un ensemble clos.

Concernant la « complétude », le corpus d'occurrences recueillies doit avoir « une ampleur et un niveau de détail adaptés au degré de finesse et à la richesse attendue en résultat de l'analyse ». Où l'élaboration de notre corpus pêche sûrement, c'est que le relevé des constructions verbales a été « manuel » au fil de la lecture des textes. Certaines occurrences nous ont sûrement échappé. Le corpus constitué n'est pas exhaustif, car il n'y a pas d'accès sous forme électronique à l'ensemble de ces textes anciens. De plus, certains autres verbes ont pu ne pas être repérés comme variant, c'est-à-dire pouvant être construits de deux façons différentes, alors qu'ils le sont. S'il y en a, ils ne font pas partie de la liste des 33 verbes que nous avons pu constituer. Cependant, ici à nouveau l'ensemble des occurrences recueillies représente une très grande partie des occurrences réelles de ces verbes dans les textes étudiés. On peut l'estimer à plus de 80 %.

Les conditions d'« exploitabilité » sont constituées par l'« homogénéité » et le « volume » (rappel). L'homogénéité a trait à la taille du corpus qui doit être « commensurable », ce qui est le cas. Son « volume » doit permettre de repérer des « éléments significatifs », c'est-à-dire permettre de dégager des régularités. La particularité des textes d'auteurs étudiés est que les 33 verbes susceptibles d'être construits actuellement de plusieurs façons ne sont pas tous employés par chacun des auteurs, parfois dans un ouvrage pas plus d'une dizaine de verbes n'est utilisée. Mais quand tel auteur l'utilise, les occurrences sont souvent nombreuses. Ceci permet de pouvoir établir quelle construction fait partie de son idiolecte (écrit) ou s'il y a fluctuation de constructions comme en (3 a) / (3 b). De ce point de vue, le volume du corpus construit est suffisant.

*L'interprétabilité du corpus*

Les critères d'élaboration du corpus en vue de la recherche présentée en début d'article sont à peu près respectés. Mais une question plus générale, plus fondamentale se pose : ce corpus est-il interprétable ? « En adoptant une perspective interprétative, nous voudrions principalement définir le corpus comme un regroupement de données (textes) correspondant à un choix explicitable et intelligible<sup>12</sup>. » À nouveau, il semble que l'ensemble des conditions ci-dessus énumérées étant réalisées, notre corpus soit interprétable dans le sens défini ici. Il est donc pour conclure représentatif, avec comme réserve que cette représentativité est liée à l'état des textes connus à ce jour, et il est interprétable. Il permet de comparer la situation actuelle et celle antérieure, sans pour autant permettre de dater précisément un état de langue antérieur. En revanche, son volume est insuffisant pour faire des comparaisons pertinentes entre états de langue successifs au cours du temps, une comparaison siècle à siècle par exemple.

*Un corpus fini mais non figé*

Une autre caractéristique de notre corpus en diachronie est que, même s'il est fini, qu'il n'est pas figé. Il évolue puisque ces deux dernières années ont été publiés deux textes souletins inconnus jusque-là du grand public. Ils n'existaient que sous forme de manuscrit pour l'un ou d'une publication pour l'autre dont il ne restait plus que quelques rares exemplaires dispersés. Ceci a permis d'enrichir le corpus en ajoutant d'autres occurrences des verbes qui font l'objet de la recherche.

De la même façon, notre corpus pour étudier la situation actuelle s'enrichit à partir de nouveaux textes écrits, un ou deux par an, que nous analysons pour repérer les occurrences des verbes étudiés.

12. Bénédicte Pincemin, *ibid.*, p. 2 de la version Internet.

## LE CORPUS EN SYNCHRONIE

Nous allons maintenant décrire et évaluer brièvement le corpus construit afin d'établir, en ce début de XXI<sup>e</sup> siècle, l'état de la langue en ce qui concerne les verbes étudiés. C'est une synchronie assez large puisque nous avons pris en compte des textes écrits et oraux des années 1970 jusqu'à maintenant, soit une période d'une quarantaine d'années.

*La constitution du corpus pour décrire la situation actuelle*

En synchronie, en plus de productions libres et de productions à la demande, il a été fait appel aussi au jugement des informateurs. Les sources pour recueillir des occurrences des verbes ont été diverses, elles constituent l'« univers » de la recherche.

Les documents écrits sont trois ouvrages imprimés (1996, 1997, 2004) et douze textes de pastorales de 1978 à 2010. Dans chacun des trois ouvrages, entre cent et deux-cents occurrences de ces verbes sujets à variation ont été relevées, avec pour un même verbe jusqu'à vingt-cinq occurrences. Les pastorales ont permis de relever entre trois et dix-sept occurrences pour chacune. Des textes divers écrits en souletin extraits d'une base informatisée de textes basques du XX<sup>e</sup> siècle ont permis de recueillir 17 occurrences de verbes à étudier.

Des enregistrements oraux libres de locuteurs plutôt âgés ont été analysés, une heure et quart d'enregistrement en tout. Ils n'ont fourni que six occurrences. D'autres textes oraux en souletin enregistrés pour la réalisation d'un atlas linguistique des parlers basques et retranscrits ont été étudiés. Ils n'ont fourni que trois occurrences. Ceux faits pour un atlas ethno-linguistique en ont aussi fourni trois.

Enfin, une large enquête écrite a été effectuée en 2004 auprès de 66 informateurs souletins d'âge divers, dont 17 enfants entre 9 et 10 ans élèves de l'école basque immersive, 11 jeunes de 12 à 15 ans élèves en classe bilingue de collège, 16 adultes de 25 à 64 ans, 15 âgés de 65 ans et plus, et 7 adultes qui ne résident pas en Soule. L'enquête consistait d'une

part en 23 phrases en français qui contenaient des correspondants des verbes à étudier, à traduire en basque. D'autre part vingt-cinq groupes de trois ou quatre phrases en basque dont il fallait évaluer le degré d'acceptabilité : non, oui, le mieux. Ceci a été complété par une deuxième enquête avec certains de ces informateurs : 25 phrases à traduire en basque avec des verbes non encore étudiés ou d'autres qui demandaient confirmation, et une liste de 20 verbes non conjugués donnés en basque avec la consigne de les employer dans une phrase. Ces enquêtes écrites ont été complétées par quelques demandes orales de précisions auprès de certains de ces informateurs.

#### *La valeur du corpus établi pour décrire la situation actuelle*

Reprenons la grille de critères d'évaluation d'un corpus proposée par Bénédicte Pincemin. Le corpus élaboré pour décrire la situation actuelle permet-il une interprétation adéquate ? Les occurrences recueillies ont-elles un ancrage suffisant dans le réel pour être le reflet des régularités que nous cherchons à mettre à jour ?

On passe sur le critère de pertinence puisque ce corpus est bien « constitué en vue d'une étude déterminée ». De même, la cohérence est assurée puisque le corpus porte « sur un objet particulier, une réalité telle qu'elle est perçue sous un angle de vue ».

Qu'en est-il du critère de « représentativité », de la « représentation fidèle » ? Les occurrences obtenues au moyen des deux enquêtes écrites n'ont pas la même valeur que les formes attestées relevées dans les documents authentiques oraux ou écrits qui ont pu être analysées. On peut considérer que les données obtenues à l'aide des différents exercices proposés sont sujettes à caution, même si l'informateur les a produites de son plein gré et qu'il a eu tout loisir pour les produire, puisqu'elles ne sont pas extraites de productions authentiques spontanées. Le critère de régularité implique lui l'absence de parasitage par des contraintes externes. Les phrases obtenues par la traduction de phrases

en français posent le problème bien connu de l'influence de la langue source.

Quant à la complétude au sens d'une « ampleur et un niveau de détail adaptés au degré de finesse et à la richesse attendue en résultat de l'analyse » (rappel), il faut reconnaître que l'ensemble de textes oraux analysés est insuffisant pour obtenir des occurrences assez nombreuses des verbes à étudier. C'est pour cela qu'il a été procédé à des enquêtes écrites en complément. Et donc de même, le critère de volume n'est pas respecté. Ici, ce n'est pas le peu de documents existants accessibles qui en est la cause comme dans le cas de l'étude en diachronie. C'est l'insuffisance du travail du chercheur isolé. Un travail d'équipe avec la technologie adéquate serait à mener sur un grand volume horaire de documents enregistrés. Il en existe d'assez nombreux à disposition, comme les centaines d'heures de souletin enregistrées par l'Académie de la langue basque pour réaliser l'atlas linguistique des parlers basques en cours de publication<sup>13</sup>. Mais il s'agit de locuteurs plutôt âgés, souvent décédés maintenant. Les occurrences de chacun des verbes seraient alors peut-être suffisantes pour repérer le « comportement » de chacun des verbes, même si pour réaliser une comparaison intergénérationnelle, il faudrait également faire des enregistrements de jeunes locuteurs puisqu'il n'en existe pas pour l'heure. On peut dire que le corpus constitué est hétérogène et, surtout, déséquilibré au profit de l'écrit.

L'approche synchronique dynamique suppose de prendre en compte le facteur intergénérationnel. Elle n'a été effectuée pour l'instant qu'au moyen des enquêtes écrites où chaque informateur a produit des phrases en toute conscience et donné son jugement sur certaines autres. La comparaison a permis de mettre à jour chez les jeunes informateurs la tendance à utiliser plus facilement les constructions avec l'auxiliaire \**ediin* « avoir » par rapport aux plus anciens qui utilisent plus fréquemment les constructions avec l'auxiliaire \**izan* « être ».

13. *Euskararen Herri Hizkeren Atlas*, vol. I (2008), vol. II (2008), vol. III (2010), Bilbao, Euskaltzaindia.

## EN CONCLUSION

Une recherche qui vise à mettre à jour des régularités à partir de données linguistiques réelles, attestées, pose la question de la nature du corpus élaboré à cette fin. La valeur, l'adéquation, de ce corpus tient à sa capacité à permettre une interprétation non ambiguë des données, afin de statuer sur les régularités postulées avant la construction du corpus.

Notre recherche a pour objet de mettre à jour la variation et l'évolution de l'indication des fonctions syntaxiques pour certains verbes transitifs en basque souletin. Son caractère à la fois synchronique et diachronique implique de constituer deux corpus différents. On a souligné la rareté des textes écrits connus à ce jour. Certains résultats fiables ont pu être obtenus, comme l'emploi systématique d'une seule des constructions possibles avec tel verbe par tel auteur ou au contraire la mise à jour de fluctuations.

Le corpus visant à statuer sur la situation actuelle a montré les limites du travail d'un chercheur seul exploitant des données non numérisées. La place de l'écrit y est importante. Certaines régularités dans l'emploi des verbes étudiés ont pu cependant être mises à jour. Une des limites de l'étude est celle de l'idiolecte puisque chaque locuteur fera peut-être un emploi particulier de cette trentaine de verbes pris dans leur ensemble. Il faudrait pouvoir statuer sur la façon de chacun de construire chacun de ces verbes. La recherche, elle, se restreint à tenter de dégager des régularités qui pourront être attribuées à telle ou telle génération de locuteurs.