David Trotter, 2007, Habeas corpus ad testificandum: l'Anglo-Norman Dictionary et son corpus. in P Kunstman & A Stein (eds), *Le Nouveau Corpus d'Amsterdam: Actes de l'atelier de Lauterbad, 23-26 février 2006.* 1 edn, vol. 34, Zeitschrift für französische Sprache und Literatur – Beihefte. Neue Folge, vol. 34, Franz Steiner, Stuttgart, pp. 153-157. <https://www.steiner-verlag.de/en/Le-Nouveau-Corpus-d-Amsterdam/9783515089975>

ANTD
ANGLO-NORMAN DICTIONARY

HABEAS CORPUS AD TESTIFICANDUM: L'ANGLO-NORMAN DICTIONARY ET SON CORPUS

David Trotter, Aberystwyth

Volltextbanken, Wortlisten und Einzeltextglossare sind keine Wörterbücher. Sie sind je nach Nutzung, nicht mal hilfreich bei der Redaktion eines Wörterbuchs [...] eine Wortliste ist keine wissenschaftliche Arbeit und eine gigantische Wortliste ist kein Gewinn. [...] Übermengen von ungeordnetem Material führen wie am Sternenhimmel zu einem Schwarzen Loch, energiegeladen, alles verschlingend, unproduktiv bis es explodiert ... (Möhren 2003: 40–41).

1. LES PRINCIPES ET LA MÉTHODOLOGIE DU CORPUS ANGLO-NORMAND

La citation de Frankwalt Möhren est un avertissement important: un corpus à lui seul est non seulement parfois d'une utilité limitée, mais peut aussi être un obstacle à la recherche scientifique dans la mesure où la masse des données fournies risque de noyer le chercheur sous des flots de citations sans grand intérêt. Pour éviter ces deux pièges (corpus inutilisable, voire inutile; corpus qui de par sa taille ralentit la recherche), il est important de réfléchir aux buts qu'il est censé atteindre. Dans le cas du corpus anglo-normand dont nous esquissons ici la description, il fait partie d'un projet de lexicographie (révision fondamentale de l'Anglo-Norman Dictionary, première édition: 1977–1992) et de constitution d'un corpus anglo-normand (www.anglo-norman.net). Ce qui a été envisagé, c'était donc un corpus qui soit à la fois un support au dictionnaire et une extension du dictionnaire. L'élément essentiel (et innovateur) est l'existence de ce lien étroit (conceptuel et technique) entre dictionnaire et corpus, procédé relativement rare (l'évolution de FRANTEXT et du TLF et des projets associés en fournit un autre exemple). Le corpus lui-même est un projet du programme «Resource Enhancement» de l'AHRC (Arts and Humanities Research Council) du Royaume-Uni, dont le but est censé être «l'enrichissement et l'élaboration de ressources scientifiques». L'entreprise vise à numériser et à mettre en ligne des textes anglo-normands qui sont soit inaccessibles (vieilles éditions, textes historiques, juridiques), soit des textes prometteurs du point de vue lexical; fait partie du corpus aussi – grâce au président de la sociéte, Ian Short – l'intégralité de la série de l'Anglo-Norman Text Society (textes typiquement anglo-normands, mélange de littérature et de textes non-littéraires, «Plain Texts» hors commerce). Ainsi, le corpus n'est pas une base de données mais une republication intégrale de textes (à la différence des BFM, et de Frantext), ce qui n'est pas sans poser des difficultés d'ordre légal. Le corpus n'est pas à la base du dictionnaire mais un complé-

ment au dictionnaire ; son but : un corpus de 10 millions de mots anglo-normands. En même temps, l'AND lui-même est en quelque sorte un corpus ou, plus précisément, une banque de citations (le « double cliquage » sur un mot dans une citation renvoie à un article ainsi qu'à toutes les autres citations où ce mot se trouve).

1.1. Quelques chiffres sur l'AND

Les chiffres « bruts » montrent dans quelle mesure l'AND est lui-même un corpus : les 120.000 citations correspondent à environ 1.000.000 de mots (à raison de 8 mots par citation) :

	AND1, G–Z	AND2, A–E	AND2, F	Total
Articles	10.400	10.600	1.223	22.223
Citations agn.	40.000	72.000	9.906	121.906
Gloses en anglais	30.000	36.000	4.681	70.681

1.2. Aspects techniques

Sans entrer dans le détail (celui-ci étant disponible sur le site du dictionnaire), on peut signaler un encodage XML, le balisage TEI ; les textes sont livrés à l'utilisateur en HTML (donc sous forme d'une page web). Le même encodage/balisage est utilisé pour l'AND, ce qui garantit la communication entre ces deux éléments du site (voir diagramme).

2. FONCTIONNEMENT AVEC L'AND EN LIGNE

Des liens actifs et la liaison systématique corpus ↔ dictionnaire font du site web un outil particulièrement efficace. Le « double-cliquage » (à paraître pour le corpus, mais qui existe déjà pour les citations dans l'AND) permet de passer de n'importe quel mot dans une citation de l'AND aux autres emplois de ce même mot. A signaler aussi l'existence d'un outil de consultation (« look-up tool ») permettant de faire une recherche sur un mot dans n'importe quel texte anglo-normand qui existe sur l'internet ou sous forme électronique, et que l'on aura sur son écran. L'utilisation de l'outil fournit l'article correspondant de l'AND avec toutes les autres citations qui contiennent le même mot. L'outil est disponible à l'adresse suivante : http://www3.anglo-norman.net/cgi-bin/getbm. Par exemple : à partir de Marie de France, *Éliduc* v. 15, texte qui figure intégralement sur le site de l'Université de Manitoba http://www.umanitoba.ca/faculties/arts/french_spanish_and_italian/m12.htm, pour le mot *meschine* : l'utilisation du « look-up » de l'AND produit l'article **meschine** (AND1) avec 29 autres citations où figure ce mot, partout dans le dictionnaire. Bien sûr, la révision de « M » aura à tenir compte de cela.

2.1. Avantages

Parmi les avantages que présente ce système, et en principe tout corpus, on notera les suivants, d'une perspective de lexicographe : une vérification quasi-exhaustive du contenu des textes devient possible (tout texte intégré au corpus étant infiniment consultable) ; il existe la possibilité d'une recherche très rapide d'un mot dont on soupçonne l'existence ; le corpus fait ressortir des combinaisons (locutions, syntagmes, collocations) autrement invisibles. Évidemment, ce sont là des avantages dont peut aussi profiter le non-lexicographe qui s'intéresse aux divers aspects de l'anglo-normand. L'existence de ce corpus facilitera aussi des études différentielles sur l'anglo-normand d'une part, l'ancien français/le moyen français d'autre part. Pour cela, il serait sans doute souhaitable (mais est-ce réalisable ?) que tous les corpus disponibles adoptent le même système/format, mais même en l'absence de cela, des recherches parallèles sur différents corpus sont parfaitement possibles. Petit à petit, on aurait ainsi une vue d'ensemble du lexique de part et d'autre de la Manche, ce qui permettra de juger de façon plus scientifique de la spécificité (ou non) du lexique insulaire.

2.2. Difficultés

Le « trou noir » de Frankwalt Möhren est une réalité inquiétante : la quantité de textes risque de submerger la rédaction. Ce n'est pas un problème pour un mot relativement rare, mais parmi les mots dont doit s'occuper un dictionnaire figurent aussi des lexèmes très courants (des milliers de citations pour *gent* ... et bien entendu, étudier par exemple l'article défini à partir d'un corpus est impensable à moins d'y consacrer un temps disproportionné dans la rédaction d'un dictionnaire). La rédaction est forcément impliquée dans le balisage du corpus, ce qui constitue ou peut constituer une tâche considérable et compliquée (surtout dans le cas de textes imprimés au XIXe siècle, avec une reproduction typographique des abréviations manuscrites, qu'il faut évidemment résoudre pour les besoins de la lisibilité, et pour qu'une recherche « normale » soit réalisable). Le corpus est une republication intégrale de textes, d'où le grand problème du droit d'auteur (sans la permission des maisons d'édition, il faut attendre 70 ans après la mort d'un éditeur scientifique ... selon la directive 2001/29/CE de l'UE). En même temps, par contre, le corpus devient aussi un moyen de faciliter l'accès aux textes en tant que tels, même pour l'utilisateur qui n'a pas le moindre intérêt proprement lexicographique/lexicologique, voire même linguistique. L'historien et le juriste trouveront ainsi leur bonheur parmi nos textes.

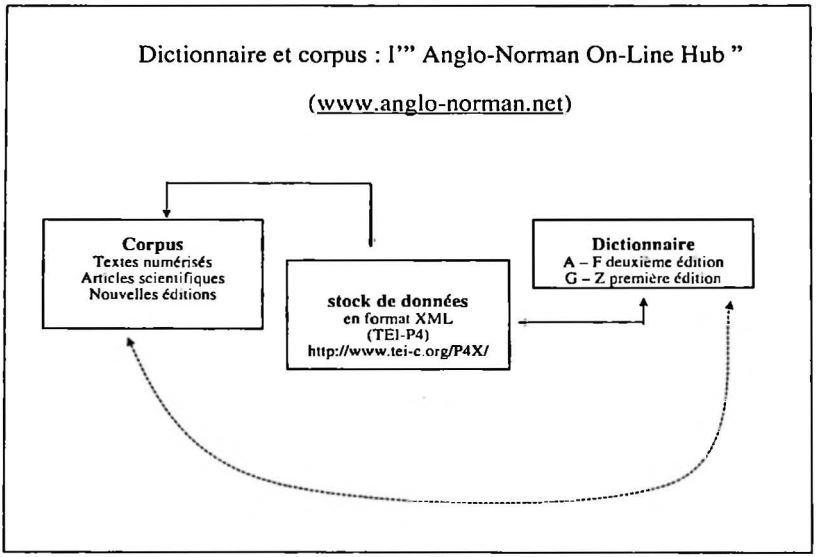
2.3.

Le diagramme suivant montre sous forme schématique le rapport entre le dictionnaire, et le corpus. Pour une vue d'ensemble plus développée, on se reportera à l'étude de Beddow (2007). L'élément essentiel se trouve – comme il sied – au centre, sous forme du stock de données en XML. C'est l'utilisation de ce format qui

permet l'échange de données entre les deux pôles du projet, car tout est conservé sous la même forme numérique. Ce diagramme est suivi d'une liste de textes déjà numérisés – que nous avons repris du site web au moment du colloque – ainsi que d'une liste de textes dont la numérisation est en cours (idem). Plus précisément, ceux-ci ont été numérisés mais attendent toujours la vérification et l'encodage définitif en XML.

3.

Évidemment, ce corpus sera, dans la mesure du possible, relié aux autres corpus déjà construits ou en voie de construction. En principe, le balisage XML-TEI devrait faciliter ce desideratum de la recherche internationale. Mais cela, c'est une autre étape ...



ANGLO-NORMAN SOURCE TEXTS

Britton. The French text carefully revised, [...] by Francis Morgan Nichols. 2 vol. Clarendon Press: Oxford, 1865

Femina
(Trinity College, Cambridge MS B 14.40)
edited with an Introduction and Notes by William Rothwell
The Anglo-Norman On-Line Hub 2005

Foedera
edited by Thomas Rymer London: Record Commissioners 1816–1869
Volume 1 (French-language documents AD 1256–1307); Volume 2 (AD 1307–1343); Volume 3 (AD 1344–1377); Volume 4 (AD 1377–1383)

Munimenta Gildhallae Londoniensis
edited by Henry Thomas Riley
London: Longman 1859–62
French Language items from Volume 1: Liber Albus

Guernes de Pont-Sainte-Maxence, *La Vie de Saint Thomas Becket*
edited by Emmanuel Walberg
Editions Honoré Champion: Paris 1936

The Oxford Psalter
(*Libri Psalmorum Versio Antiqua Gallica e cod. MS. in Bibl. Bodleiana asservato*)
edited by Francisque Michel
Oxford: The University Press 1860

Le livre de Seyntz Medicines
edited by E.J. Arnould
Oxford: Blackwell 1940
(Anglo-Norman Text Society No. 2)

Walter de Bibbesworth. *Le Tretiz*
edited by William Rothwell.
Anglo-Norman Text Society. Plain Texts Series 6. London 1990

Live links to the source, indicated by the icon, are in place where this edition is cited in W Rothwell's article Of Kings and Queens, or Nets and Frogs: Anglo-French Homonymics on this site. Unlike the live links between dictionary citations and sources mentioned below, this linkage is already freely available to all users.

Les Gius Partiz des Eschez. Two Anglo-Norman Chess Treatises
edited by Tony Hunt
Anglo-Norman Text Society. Plain Texts Series 3. London 1985

«SOUS PRESSE»

Rhymed Apocalypse (ANTS)
John of Gaunt's Register 1
John of Gaunt's Register 2
Hospitallers' Riwle (ANTS)
Lumere as Lais (ANTS from digital version)
Manieres de Langage (ANTS from digital version)
La Seinte Resureccion (ANTS)
Romance of Horn (ANTS)
Rotuli Scotiae
Secre de Secrez
Roman de toute Chevalerie (ANTS)
Amys e Amillyoun (ANTS plain texts)
St. Brendan
Livre de Catun (ANTS plain texts)
Chronique de Londres
Corset (ANTS)

La Destruction de Rome (ANTS plain texts)
18 A-N Fabliaux (ANTS plain texts)
Folie Tristan (ANTS plain texts)
Fouke le Fitz Waryn (ANTS)
Henry I Coronation Charter
A-N Letters & Petitions (ANTS)
Ripon Memorials (Surtees Soc.)
St. Fraunceys (ANTS from digital version)
Testamenta Eboracensia (Surtees Soc.)
Cambridge Psalter
Mirour de Seynte Esglise
St. Edmund (ANTS)
Petit Plet (ANTS)
St. Modwenna (ANTS)
Protheselaus (ANTS)
Simund de Freine: Vie St. George & Roman de Philosophie
La Petite Philosophie (ANTS)

RÉFÉRENCES

Beddow, M. (2007): «L'*Anglo-Norman Dictionary*: présentation technique» – Trotter, David (ed.): Actes du XXIVe Congrès International de Linguistique et de Philologie Romanes, Tübingen: Niemeyer.

Möhren, Frankwalt (2003): «Bilanz und Perspektiven» – Städtler, Thomas (ed.): Wissenschaftliche Lexikographie im deutschsprachigem Raum, Heidelberg: Universitätsverlag Winter, 33–47.

Rothwell, A.J. / Trotter, D.A. (2007): «L'*Anglo-Norman Dictionary*: une présentation» – Trotter, David (ed.): Actes du XXIVe Congrès International de Linguistique et de Philologie Romanes, Tübingen: Niemeyer.