

Le sens des mots chez Tony Blair (people et Europe)

Edward Arnold

French Dept/Centre for European Studies, Arts Building, Rm 4106
University of Dublin, Trinity College, Dublin 2, IRLANDE
ejarnold@tcd.ie

Abstract

Presentation of corpus of speeches by Tony Blair, British Prime Minister from 1997 to 2007. Description of initial processing of texts: correction of spelling, standardisation of written forms, mark-up of texts. This corpus reveals the words that were most frequently used by Blair when he was Prime Minister. It also shows the single or multiple specific meanings that Tony Blair gave to these words. For example, the words “People” or “Europe” had a very specific meaning for the former Prime Minister.

Résumé

Présentation du corpus des discours de Tony Blair, Premier ministre anglais de 1997 à 2007. Description des traitements préalables des textes : correction orthographique, standardisation des graphies, balisage des textes. Ce corpus permet de retrouver les mots les plus fréquemment employés par le Premier ministre. Il permet surtout de connaître le ou les sens spécifiques que T. Blair a donné à ces mots les plus usuels. On donne en exemple les mots « people » et « Europe » que le Premier ministre utilisait dans un sens bien particulier.

Mots-clés : Blair, people, Europe, Britain.

1. Introduction

Depuis la création du Labor Party en 1906, aucun travailliste n’est resté Premier ministre aussi longtemps que T. Blair (mai 1997- juin 2007). Bien que son règne s’achève tout juste, il est d’ores et déjà certain que ses dix ans de pouvoir ont marqué la politique et la société anglaise.

Avec, l’expression « gouverner c’est parler » a pris une nouvelle réalité du fait de l’importance prise par le « spin », caractérisé par un poids grandissant des conseillers en communication qui se chargent de l’image des partis et des hommes politiques et surtout du fait d’une présence constante du Premier ministre dans les médias.

T. Blair est beaucoup apparu aux informations télévisées et dans diverses émissions ; il a régulièrement sorti des articles dans la presse et accordé des entrevues aux journaux ; il a tenu des conférences de presse, dont certaines improvisées sur le « pas de la porte », sans compter les communiqués que le 10 Downing street a publié à chaque fois qu’un événement marquant survenait quelque part dans le monde. Une bonne partie de ces textes ont été placés en ligne sur le site du Premier ministre et ils offrent un matériau intéressant pour les politologues et les linguistes.

Depuis 2003, nous déchargeons régulièrement ces textes dans le but de constituer un large corpus du discours politique anglais, voire une base de données textuelles aisément

consultable (pour une première présentation : Arnold 2005). Cette communication tire un bilan provisoire de cette expérience et donne un exemple des informations que l'on peut tirer d'un tel corpus.

2. Le corpus et les traitements préalables

Sous réserve d'ultimes corrections et sans les séances de questions au Parlement (« Question time » qui sont en cours de traitement), le corpus « Tony Blair » comporte près de 800 textes, soit plus de 1,5 millions de « mots » (tokens pour « plus petit constituant de la surface d'un texte ou d'un corpus ») dont 22 100 « mots différents », ou vocables (en anglais « types » pour : « plus petit constituant du vocabulaire d'un texte ou d'un corpus »). Le tableau 1 résume les principales caractéristiques de ce corpus et signale que, à partir de 2001, le Premier ministre a eu tendance à multiplier les interventions publiques ou à en placer plus en ligne sur son site (nous n'avons pas encore éclairci ce point).

Années	Nombre de textes mis en ligne	Nombre de mots par an	Nombre de mots par mois	Nombre de mots par textes
1997 (8 mois)	23	29 747	3 718	1 293
1998	25	45 252	3 771	1 810
1999	31	55 401	4 617	1 787
2000	22	58 761	4 897	2 671
2001	71	119 515	9 960	1 683
2002	39	113 684	9 474	2 915
2003	71	161 886	13 491	2 280
2004	96	211 271	17 606	2 201
2005	102	156 763	13 064	1 537
2006	206	414 817	34 568	2 014
2007 (6mois)	102	155 494	25 916	1 524
Total	788	1 522 591	12 480	1 932

Tableau 1. Le corpus Tony Blair

Ces textes sont archivés dans leur version originale. En effet, ces documents peuvent disparaître à tout moment car le site est régulièrement réorganisé et il est sujet aux changements politiques... De plus, il est indispensable de pouvoir retourner, à tout moment, à la source, car les interventions nécessaires pour intégrer un texte dans un corpus ne sont pas bénignes.

La première étape consiste à :

- revoir l'orthographe du texte. Certes il y a peu de fautes de frappe dans les textes mis en ligne sur le site du 10 Downing street ou sur celui du Parlement, mais ce contrôle est indispensable pour éviter d'entrer dans le corpus des fantômes de mots et pour garantir au chercheur que toutes les occurrences du mot qu'il recherche lui sont bien fournies...

- les abréviations et les graphies non conventionnelles sont standardisées. En effet, avant de constituer un corpus, il faut répondre à quelques questions : UK et United Kingdom sont-ils deux (ou trois) « types » différents ? US et United States ? EU et European Union ? 1st et first ? etc. Conserver telles quelles ces variantes évite un travail préalable fastidieux qui présente de plus le risque d'en oublier et d'introduire des fautes (comme toujours lorsqu'on intervient manuellement sur un texte électronique...) Mais alors on augmente artificiellement le nombre de « mots différents » dans la base, ce qui rendra beaucoup plus difficile les recherches ultérieures. Etant donné le très grand nombre de sigles, d'abréviations, de nombres et d'unités de mesure dans le discours politique contemporain, les statistiques réalisées sur ce matériau brut n'auront guère de crédibilité.

- enfin, l'orthographe des « noms propres » est standardisée, spécialement celle des toponymes et des patronymes afin que les mêmes personnes, les mêmes lieux n'apparaissent pas sous deux ou trois noms différents...

Naturellement, ces interventions de pure forme ne concernent pas le texte du Premier ministre mais la transcription effectuée par son secrétariat. De plus, l'existence de l'archive mentionnée au début de ce paragraphe, permet de retourner à l'original si l'une ou l'autre de ces opérations est contestée.

La seconde étape consiste en un balisage du texte (sur le balisage des corpus : Burnard et Sperberg-McQueen 2002). Dans le cas présent, la mise en page n'est pas conservée car elle est propre au site internet et non pas au Premier ministre. Les balises servent d'abord à isoler les propos du Premier ministre de tout le reste : résumés, intertitres - que les services du premier ministre ajoutent habituellement - questions des journalistes, interventions des autres participants aux « conférences de presse conjointes », nombreuses mentions marginales (« rires », « interruptions », « inaudible »...) ajoutées par les sténographes...

Il est essentiel que ces passages ne soient pas présentés à l'utilisateur sur le même plan que les propos de T. Blair et il est évident qu'une analyse statistique du discours du Premier ministre doit les laisser de côté ! Mais il ne faut pas pour autant les effacer car ils sont parfois indispensables pour la bonne compréhension des propos (rappelons que nous voulons constituer une véritable base de textes pérenne et non un simple corpus à usage statistique unique).

Enfin, les « références » sont placées au début du texte (auteur, titre éventuel, lieu, circonstances et date du discours). Ces dernières informations sont essentielles pour permettre de découper le corpus en sous-parties homogènes du point de vue du genre (les discours du trône, les réponses aux questions des parlementaires, les discours aux congrès travaillistes...) de chronologie ou du thème traité.

La troisième étape consiste à découper automatiquement le texte en « mots » (tokens), en suivant la norme « OCP » (Hockey et Martin 1988) puis à constituer des index des « mots différents » (types) qui fourniront des portes d'entrée pour la « fouille » des textes et pour les calculs statistiques.

Ces opérations préliminaires présentent de nombreux intérêts. Par exemple, on peut établir avec certitude le vocabulaire de ce corpus ou d'un sous-corpus spécifique. A titre d'exemple, le tableau 2 ci-dessous donne la liste des substantifs et des adjectifs les plus fréquents.

Rang	Substantifs	Fréquence	Adjectifs	Fréquence
1	people	9 246	important	2 749
2	world	3 396	european	1 681
3	government	3 093	international	1 594
4	time	2 880	public	1 576
5	Europe	2 483	good	1 519
6	country	2 458	sure	1 483
7	year	2 417	clear	1 453
8	work	2 414	British	1 397
9	Irak	2 139	different	1 346
10	Britain	2 052	last	1 287

Tableau 2. Les substantifs et les adjectifs les plus fréquents chez T. Blair

Cette liste présente quelques surprises et soulève des questions. Par exemple, le Premier ministre aurait-il parlé plus souvent de l'Irak que de la Grande Bretagne ? Non ! car il faut ajouter à (Great) « Britain », de nombreux « our country » et 884 « United Kingdom » (généralement abrégé en « UK »), sans compter un grand nombre de « us » qui signifient « nous, les Britanniques ». Il n'empêche que la présence de l'Irak, si haut dans la liste, souligne assez combien la participation britannique à l'intervention militaire dans ce pays fut la grande affaire des onze ans de gouvernement Blair.

Quel sens T. Blair donne-t-il à ces mots les plus usuels ?

La statistique peut aider à donner une réponse à cette question. Pour illustrer cette proposition, nous allons utiliser deux de ces substantifs les plus fréquents : « people » et « Europe ». En effet, dans les propos des présidents français (Cotteret et Moreau, 1969, Labbé 1990, Mayaffre 2004) ou dans ceux des Premiers ministres (Labbé et Monière 2003), le « peuple » ne vient pas en tête de liste et ne figure d'ailleurs pas dans les mots les plus fréquents, les présidents disent : les « Français(es) ». En revanche, depuis G. Pompidou, l'Europe fait bien partie des substantifs les plus usuels des présidents français, mais sa fréquence relative est, en moyenne, inférieure à celle que l'on observe chez T. Blair, alors même que les Britanniques sont plutôt considérés comme « euroseptiques ».

Avant de lire ce qui suit, le lecteur est invité à consigner ses propres intuitions, concernant le sens de ces deux mots. Il pourra ainsi les confronter avec les résultats des traitements et juger de l'intérêt de la démarche.

3. Les univers lexicaux

D'après le postulat « distributionnaliste », dans un texte ou un corpus donné, un mot tire son sens de son « entourage », c'est-à-dire des mots auxquels il est associé. Si l'on accepte cette idée, il faut examiner les mots entourant chacune des apparitions de « people » ou de « Europe ». Les concordanciers, qui fournissent ces contextes avant et arrière, se révèlent difficiles à utiliser quand ils comportent plusieurs milliers de lignes. De plus, ces listes permettent de retrouver les mots associés mais ne répondent pas à deux questions évidentes :

- les associations relevées sont-elles « trop » fréquentes, « pas assez fréquentes » ou « normales » ? En effet, quand deux mots sont utilisés fréquemment, une distribution « au hasard » amènera nécessairement ces deux mots à se côtoyer un certain nombre de fois. Si l'on ne connaît pas ce « certain nombre de fois », comment juger de la singularité des associations effectivement rencontrées ? A quelles conditions, l'on pourra affirmer que deux mots s'attirent comme s'ils étaient aimantés ?

- quelles sont les associations « oubliées », celles qui n'apparaissent pas alors que, étant donné la fréquence des mots concernés, on pourrait s'attendre également à les voir associés un certain nombre de fois ?

La seconde question est particulièrement intéressante. En effet, postuler que certains mots s'attirent mutuellement conduit à admettre – d'abord d'un simple point de vue d'équilibre physique au sein d'une population fixée - qu'il y en a d'autres qui se repoussent, ou encore, que, à côté des associations positives, il y a des associations négatives.

Ces postulats ont inspiré la statistique textuelle et notamment la recherche des « cooccurrences » (Lafon 1981, Salem 1987, Tournier 1985). Cependant, ce que nous apprennent les longues listes de cooccurrents ne diffère pas fondamentalement des concordanciers : les substantifs sont généralement précédés d'un déterminant, les verbes d'un pronom personnel, etc.

Le calcul des univers lexicaux vise à neutraliser cet inconvénient et à établir l'ensemble des relations d'attraction ou de répulsion qu'un mot entretient avec les autres dans le corpus considéré (Hubert et Labbé 1995 ; Labbé et Labbé 2005). Pour repérer ces relations, on relève toutes les phrases contenant le vocable considéré. Le vocabulaire de ce sous-ensemble est comparé à celui du corpus entier. Lorsque, pour un mot donné, la fréquence relative dans cet univers dépasse significativement celle observée dans l'ensemble du corpus, l'association est dite « positive » ; à l'inverse, si elle est significativement inférieure, l'association est négative (antonymie).

Prenons un exemple pour illustrer la manière dont est fait ce calcul.

Les phrases contenant le mot « people » couvrent 8,97% du total des propos de T. Blair.

Dans ces phrases on rencontre 185 fois l'adjectif « British » (la plupart du temps dans l'expression « British people »).

Comme il y a en tout 1 397 occurrences de « British » dans le corpus, on s'attendrait donc à trouver dans les phrases contenant « people » :

$1\,397 * 0,0897 \approx 125$ occurrences de « British ».

Autrement dit, si l'utilisation des deux vocables avait été « indépendante » (comme deux variables statistiques sans lien entre elles), ils devraient survenir ensemble dans les mêmes phrases environ 125 fois.

Le test statistique présenté par Labbé et Labbé (1994) indique qu'il y a moins de 1 chance sur 10.000 pour que cet excédent (185 associations observées au lieu des 125 attendues) soit dû au hasard. L'expression « British people » est donc caractéristique des discours de T. Blair.

A l'inverse « European Union » apparaît seulement 12 fois dans les mêmes phrases en compagnie de « people » alors que, du fait de leurs fréquences mutuelles dans l'ensemble du corpus, on attend 46 associations entre ces deux vocables. Le même test statistique indique qu'un tel écart négatif (déficit) a également moins de une chance sur 10 000 de se produire

par hasard. Il est donc pratiquement certain que « people » et « European Union » exercent une sorte d'exclusion mutuelle, c'est-à-dire que les deux choses ne sont pas associées dans l'esprit du Premier ministre ou pour le dire plus simplement que les affaires européennes ne sont pas vraiment du ressort de « people »...

L'« univers lexical » d'un mot est constitué de l'ensemble des relations d'attraction et de répulsion qu'il entretient avec tous les autres mots du vocabulaire d'un auteur, relations mises en évidence grâce au calcul présenté ci-dessus. A titre d'exemple, l'annexe 1 donne l'univers du mot « Europe » dans les discours de T. Blair (celui de « people » est trop long pour être reproduit dans le cadre de cette communication).

Cependant, ces listes de mots peuvent sembler abstraites car le contexte de l'association (les mots autour) est souvent difficile à reconstituer. On propose donc un retour au texte en utilisant des phrases entières comme les dictionnaires de langue placent des citations après la définition d'un mot pour illustrer les utilisations possibles de ce mot. Mais ici la sélection ne dépend pas du bon plaisir du lexicographe, elle est réalisée par l'ordinateur sur des bases objectives. Pour cela, le programme relit l'ensemble du corpus en attribuant un score à chaque phrase contenant le mot recherché. Il incrémente le score de la phrase d'une unité à chaque fois qu'il rencontre un vocable positivement associé avec mot-clef. A l'inverse, il réduit ce score d'une unité lorsqu'il s'agit d'un vocable négativement associé avec ce mot-clef. Puis il classe les phrases, où figure le mot recherché, en fonction du score absolu puis relatif (le score absolu divisé par la longueur de la phrase) (faute de place, nous ne donnons en annexe 1 que les premières phrases).

4. Le peuple britannique ?

Les phrases où apparaissent « people » contiennent en tout 136 596 mots, soit : 8.97% du corpus total. Dans ces phrases, 502 vocables apparaissent « trop » souvent - au sens du test statistique décrit ci-dessus et avec un seuil de 1% de risque d'erreur - et 696 n'apparaissent pas « assez » souvent. Parmi les associations positives, la plus caractéristique concerne l'adjectif « British ».

L'expression « British people » ne surprend pas dans la bouche d'un dirigeant britannique et un Français traduira spontanément cette expression par « peuple britannique ». Ce serait un contresens qui vient d'une confusion entre « people » (pluriel de « person ») et « people » (l'ensemble des citoyens qui constitue une communauté, et exprimé au singulier) ! Pour le comprendre, voici l'une des phrases les plus caractéristiques de cet univers :

« British people want the best from their public services and their public servants » (2 décembre 1999).

Sans compter « people », cette phrase contient 13 mots dont 9 sont caractéristiques de l'univers de « people » (British, want, best, their, public, services, their, public, servants).

Naturellement s'il s'agissait du « peuple britannique », on aurait « wants the best from **its** public services »... L'expression « les gens britanniques » n'existe pas en Français. Il faut donc traduire par quelque chose comme : « En Grande Bretagne, les **gens** attendent le meilleur de leurs services publics ».

Beaucoup d'autres signes montrent que, lorsque T. Blair utilise « people », il s'agit d'un pluriel. Par exemple, a hundred, a thousand people, etc. Les verbes caractéristiques de cet univers sont également au pluriel - et en premier lieu « people are » qui est employé plus de 1 300 fois - de même que les articles et des pronoms (their, them), et non pas au singulier

comme cela serait le cas s'il s'agissait du « peuple britannique ». En fait, quand T. Blair emploie « people » au sens de « peuple » il le met souvent au pluriel (« peoples of Europe », « peoples of the Middle East », etc.).

A la réflexion, cela n'est pas spécialement surprenant. D'une part, dans les textes anglais contemporains, « people » est très rarement employé pour « peuple ». Il désigne des personnes plus ou moins bien identifiées (des « gens »). D'autre part, l'expression « peuple britannique » ne serait-elle pas une incongruité dans la bouche d'un homme politique de ce pays constitué de plusieurs peuples (Angleterre, Ecosse, Pays de Galles) formant une seule « nation » ? En effet, quand T. Blair évoque l'entité politique souveraine formée par ses concitoyens (ce qu'il fait beaucoup plus rarement qu'un politicien français), il emploie le mot « nation » (nous en donnons un exemple très caractéristique à la fin du paragraphe suivant consacré à l'Europe). Les sensibilités politiques d'une partie de la population des six comtés septentrionaux de l'Irlande sont donc implicitement respectées avec l'emploi du mot « Grande Bretagne » puisque ce dernier exclut l'Irlande du Nord. Ainsi seul le toponyme « Royaume Uni » comprend la Grande Bretagne et les six comtés.

En même temps qu'il calcule l'univers lexical du mot, le logiciel repère les phrases les plus caractéristiques de cet univers. En voici quelques exemples qui illustrent bien les deux sens principaux de « people » dans le discours de T. Blair.

« The truth is for some people, particularly those that are hard to place, people who maybe have got multiple disadvantages or difficulties in finding work, the voluntary sector can often do better » (Discours devant la conférence du patronat britannique - 24 janvier 2007).

« What I am trying to say to people in the western world is, whatever you thought about getting rid of Saddam, there is only one side to be on, and that is the side of democracy and liberty. » (Conférence de presse avec le Premier Ministre iraquien, 19 sept 2004)

« The people of Northern Ireland deserve that. The people of the island of Ireland deserve that. It is for everybody to play their part and for the Irish government to play ours » (Conférence de presse avec le Premier Ministre irlandais, Bertie Ahern, 25 juin 2004).

« I mean the people here are from different backgrounds and parts of the world that have come together with a real commitment » (Discours aux « commissioners for Africa » – Londres - 24 février 2005)

L'examen des listes complètes et des phrases les plus caractéristiques montre que le domaine d'utilisation de « people » est assez restreint et que ce terme est exclu de la majorité des thèmes traités par le Premier ministre. En effet, dès qu'un mot atteint une fréquence relativement élevée (ce qui est le cas ici), on observe habituellement un nombre à peu près équivalent d'associations positives et négatives. Ici le déséquilibre est important : le vocable génère près de 40% d'associations négatives de plus (696 contre 502). En fait, les « gens » surviennent surtout dans les propos de T. Blair quand il s'agit des questions de société (violence, chômage, immigration, communautés religieuses...) Comme l'indique la liste des associations négatives, l'expression n'est guère utilisée quand il s'agit de la politique intérieure, des institutions, du gouvernement, de la politique internationale, de l'Europe – dans ce cas, on parle de la « nation » - de la monnaie, du budget, des finances, de l'école...

5. L'Europe... dans l'alliance avec les Etats-Unis

Après « people » et « government », « Europe » est le troisième univers lexical par la superficie du texte couverte (3,2% du total des propos de T. Blair lui sont consacrés) et par le nombre des associations avec d'autres mots (718). L'annexe 1 donne la liste des vocables associés avec Europe ainsi que les phrases les plus caractéristiques.

Comme avec « people », les relations d'attraction (314) sont moins nombreuses que les relations de répulsion mutuelles (404 associations négatives). En quelque sorte, la définition de l'Europe chez T. Blair est plutôt négative. D'abord, lorsqu'il parle de l'Europe, le Premier ministre utilise très peu les pronoms personnels (au premier rang desquels « you », « I », « we ») et également très peu de chiffres, de dates et de noms propres (en dehors des noms des pays participants à l'Union). Autrement dit, il s'agit d'un discours de principe peu personnalisé et peu ancré dans le temps et dans l'espace. De plus, sans probablement s'en rendre compte, l'orateur définit l'Europe par tous les thèmes dont elle est exclue : l'école, le gouvernement, les gens, la livre sterling, la finance, l'Irak, la police, l'éducation, la santé et les hôpitaux, l'Etat, la politique internationale, la famille, la loi, l'armée, les médias, les programmes, l'Irlande du Nord, Israël et les Palestiniens, l'Afrique, la sécurité, les équipements, le nucléaire, la justice, la religion... et, par-dessus tout : pas de « super Etat » !

Les associations positives délimitent un champ d'action limité aux questions commerciales - les compagnies, le plein-emploi, le libre échange (free trade) - la croissance économique ainsi qu'un contrôle de l'immigration...

Tout au long de ses onze ans de pouvoir, T. Blair a répété que, politiquement, la Grande Bretagne devait imposer son leadership en Europe pour faire triompher sa vision de l'UE - c'est-à-dire une « union of strong nation state » - un espace de commerce libre (free trade), une alliance privilégiée avec les Etats-Unis.

T. Blair a été très souvent questionné à ce sujet. Il a toujours refusé d'opposer l'Europe et les Etats-Unis d'Amérique. Comme tous ces prédécesseurs –de droite comme de gauche– depuis la fin de la seconde guerre mondiale, il a présenté les deux « alliances » comme complémentaires et non comme opposées. Cependant, la manière dont il a répondu aux questions posées à ce sujet montre assez clairement où va sa préférence. Ces réponses sont toujours présentées dans une gradation ascendante : l'Europe, l'Alliance atlantique, le monde. Au fond, l'Union européenne n'était pour lui qu'un moyen de rendre la Grande Bretagne plus forte dans l'Alliance atlantique et dans le monde. Contrairement à ce qu'il aimait à dire, T. Blair ne portait probablement pas sur l'Europe un jugement plus favorable que la plupart des « eurosceptiques ». Il l'envisageait comme un mal nécessaire que les Britanniques devaient accepter comme une condition de la prospérité et de la grandeur du Royaume Uni. Mais il a avoué au moins une fois que cela était un casse-tête politique pour les Britanniques :

« Yes, Europe can be a political headache for a proud sovereign nation like Britain » (27 septembre 2006).

6. Conclusions

Sans doute le lecteur pourra estimer qu'il savait déjà tout cela avant de nous lire, surtout s'il suit de près la politique britannique. Toutefois, s'il a bien consigné ses intuitions a priori et qu'il les confronte a posteriori avec ce qu'il vient de lire, nous ne doutons pas qu'il reconnaitra quelques décalages et quelques surprises, même sur ces sujets les mieux connus.

D'ailleurs, même si ce n'était pas le cas, sur des sujets pareils, il n'est pas inutile d'apporter des confirmations solides aux intuitions des observateurs.

Au-delà de « people » et de « Europe », ce bref exposé aura fait sentir combien la constitution de vastes corpus peut avoir de multiples intérêts pour les chercheurs en sciences humaines et sociales. Il s'agit bien sûr de constituer des archives électroniques dans lesquelles ils peuvent retrouver aisément les informations qu'ils recherchent soit sous la forme d'index ou de concordanciers, soit en leur permettant de consulter les textes originaux. Il s'agit surtout de leur offrir des outils statistiques nouveaux parmi lesquels les univers lexicaux et la recherche des phrases les plus caractéristiques sont certainement très prometteurs.

Une condition préalable mérite toutefois d'être rappelée. Avant d'être admis dans une base de données, un texte doit être soigneusement préparé. Il est indispensable que la source figure en tête du document avec un minimum d'indications concernant la date, le lieu et les circonstances dans lesquelles il a été émis. Il faut également opérer une correction orthographique et une standardisation des graphies multiples. Enfin, un balisage interne permettra de repérer les différents locuteurs. Bon nombre des « corpus » actuels sont malheureusement constitués sans ce minimum de rigueur.

Remerciements

L'auteur remercie Cyril et Dominique Labbé qui ont mis à sa disposition les logiciels nécessaires à cette étude et qui l'ont initié à l'art de constituer des corpus et d'utiliser la statistique textuelle. Dominique Labbé a bien voulu relire une première version de ce travail.

Références

- Arnold E. (2005). Le discours de Tony Blair (1997-2004). *Corpus*, 4, p. 55-77.
- Burnard L. et Sperberg-McQueen C. M. (éds) (2002). *Guidelines for Electronic Text Encoding and Interchange*. University of Virginia Press.
- Cotteret J.-M. et Moreau R. (1969). *Le vocabulaire du Général de Gaulle*. Paris, Presses de la Presses de la Fondation nationale des sciences politiques.
- Hockey S. et Martin J. (1988). *OCP Users' Manual*. Oxford. Oxford University Computing Service.
- Hubert P. et Labbé D. (1995). La structure du vocabulaire du général de Gaulle. In Bolasco Sergio et Al (éds). *IIIe Giornate internazionali di analisi statistica dei dati testuali*. Rome, CISU, II, p 165-176.
- Labbé C. et Labbé D. (1994). *Que mesure la spécificité du vocabulaire ?* Grenoble, CERAT. Repris dans *Lexicometrica*, 3, 2001.
- Labbé C. et Labbé D. (2005). How to measure the meanings of words ? Amour in Corneille's work. *Langage Resources Evaluation*, 2005, 39, p 335-351.
- Labbé D. (1990). *Le vocabulaire de François Mitterrand*. Paris, Presses de la Fondation nationale des sciences politiques.
- Labbé D. et Monière D. (2003). *Le vocabulaire gouvernemental. Canada, Québec, France (1945-2000)*. Paris, Champion.
- Lafon P. (1981). Analyse lexicométrique et recherche des cooccurrences. *Mots*, 3, octobre 1981, p. 95-147.
- Lebart L. et Salem A. (1994). *Statistique textuelle*. Paris, Dunod.
- Mayaffre D. (2004). *Paroles de président. Jacques Chirac (1995-2003) et le discours présidentiel sous la V^e République*. Paris, Champion.

Salem A. (1987). *Pratique des segments répétés. Essai de statistique textuelle*, Paris, Klincksieck.

Annexe

L'univers lexical de « Europe » dans les discours de T. Blair

Vocables significativement sur-employés dans l'univers (classement par indices décroissants)

lot, respect, companies, leadership, discussion, security, concerns, agree, obviously, issue, discussed, G8, employment, act, working, rates, stability, sure, my, financial, indeed, better, far, though, United Kingdom, most, commission, sensible, how, own, growth, paper, illegal, right, over, here, doubt, should, social, continent, particularly, challenge, asylum, parts, perspective, whole, prosperity, growing, overall, differences, close, would, why, like, recently, free, setting, event, too, because, even, challenges, stage, has, lowest, environment, works, recent, essential, time, development, referendum, on, operating, wants, must, highest, driving, sovereignty, increasingly, share, collective, key, WTO, members, lead, organised, view, where, vital, round, coming, bold, Dutch, table, June, industry, politically, convention, years, gives, modern, open, obvious, work, move, French, economies, positive, prosperous, set, moving, creating, make, reflect, approach, change, apart, Chirac, moment, also, century, inside, nation, Germany, trading, nations, engagement, concern, agreement, nearly, Romania, biotechnology, budget, depend, versus, Cyprus, push, areas, other, engaged, both, Spain, Latin, play, itself, war, big, competition, fundamental, believe, trade, north, governments, transformed, more, attitude, role, strongly, Italy, foreign, wider, division structural, disagreements, France, allies, priorities, largest, federal, shows, markets, council, political, criteria, argument, cold, informal, importance, for, practical, ours, with, Holland, rival, liberalisation, shape, favour, Turkey, strategic, positions, Japan, ally, economically, elsewhere, issues, western, agenda, Balkans, within, opening, biggest, continental, British, needs, Lisbon, euro, model, part, energy, important, agricultural, closely, rules, regulation, constructive, isolation, European Union, interest, single, closer, position, India, effective, globalisation, constitution, relations, currency, United States, summit, strength, member, cooperate, cooperation, allows, partners, reconnect, Poland, USA, relationship, German, institutional, accession, faces, alliances, centre, arguing, vision, jobs, China, NATO, fastest, immigration, membership, presidency, player, economy, superstate, world, between, effectively, policy, enlargement, Russia, enlarged, transatlantic, Britain, citizens, market, stronger, partner, commun, Across, leading, influence, America, eastern, Asia, alliance, competitive, treaty, economy

Vocables significativement sous-employés dans l'univers (classement par indices décroissants)

you, they, school(s), government, service, Iraq, Iraqi(s), people, their, Saddam, there, pound(s), he, community, local, families, police, public, year, HHS, life, them, education, children, violence, system, United Nations, know, are, by, parents, health, this, help, weapons, job, your, individual, let, provide, go, can't, have, made, society, his, young, house, terrorist, don't, no, staff, may, choice, me, pay, every, resolution, law, waiting, plan, treatment, tribute, human, going, everyone, solution, patients, given, secondary, I, international, end, point, state, situation, teachers, authorities, increased, what, we, money, or, forces, those, media, who, care, some, skills, been, were, person, back, fair, done, it, someone, these, results, well, street, Israël, again, students, responsibility, settlement, party, regime, (northern) Ireland, ground, million, read, sides, behaviour, progress, extra, specialist, teaching, information, city, now, infrastructure, nuclear, intelligence, criminal, terrorist, Palestinian(s), wait, trust, chemical, authority, military, week, published, final, percent, democracy, president, used, obligations, hours, hospitals, thank, private, doing, said, Africa, charge, food, choices, operation, start, built, religious, down, support, patient, justice, welcome, into, very, actually, Bush, specific,

groups, when, introduced, services, biological, mission, but, success, announced, problem, primary, victims, advice, hope, conflict, officer, different, cut, training, already, total, Scotland, action, nothing, terrorism, breach, contribution, process, opportunities, continue, attack, income, troops, civil, fear, cost, minimum, step, unemployment, poor, facilities, coalition, map, sport, fund, greater, thanks, themselves, taliban, quality, voluntary, struggle, duty, reduce, funding, legislation, cannot, give, chance, strategy, freedom, saying, unless, yesterday, getting, simple, mean, country, many, increase, basid, deal, England, against, something, politics, see, contrary, diversity, thing, true, first, talk, three, sector, loock, old, need, clear, communities, way, plans, statement, rights, secretary, return, problems, course, weeks, cent, mass, proper, proposals, kind, armed, either, nature, middle, accept, control, road, debt, response, bit, wrong, understand, exactly, forward, run, September, anyone, risk, kingdom, report, poverty, child, today, bring, any, through, region, universities, Iran, understanding, powers level, properly, labour, vast, science, values

Phrases les plus caractéristiques de l'univers de « Europe »

Classement par scores absolus

What's fascinating about Europe is these new countries coming in from central and eastern Europe are countries that totally share Britain's view of Europe not as a super state, but as a Europe of nations, Europe standing strong with the United States of America and the transatlantic alliance, Europe reforming itself economically (31 mai 2003, score absolu : 30).

Secondly, it has been a foreign policy essentially underpinned by two alliances, one with the United States of America and the other the strategic partnership of Britain inside the European Union, so for a long period of my time in politics, people were told to choose between a strong alliance with America and a strong position in Europe (17 avril 2007, score absolu : 28).

I believe passionately that Britain's place is at the centre of Europe, I think it would be an act of extraordinary foolishness for us as a country in the early 21st century to withdraw from the biggest political alliance in the world and the biggest economic market, right on our doorstep, where we do 60% of our trade, where millions of British jobs depend (15 juin 2004, score absolu : 28).

Classement par scores relatifs

Iraq has dramatically surfaced differences between Europe and America and Britain's role in both alliances (15 novembre 2004, score: 0.733).

The next big area of reform and change for us has been in our relations with Europe (10 janvier 1998, score: 0,706).

The whole of Europe is moving in Britain's direction today (17 octobre 2003, score: 0,7)

We are stronger with the United States because of our growing strength in Europe (30 décembre 2001, score: 0,692)

Both of us with the United States, both of us with Europe, both of us in the Commonwealth and both also in the Pacific and Asia occupy a special place (23 février 2001, score: 0,677).

We believe passionately that Europe must take the road of reform in its economy and renewal of its alliance with America (15 novembre 2004, score: 0,667).

In respect of Europe, Britain is part of the continent of Europe; Europe is the key strategic alliance right on our doorstep (26 juin 2003).

Second principle: strong in Europe and strong with the United States (10 novembre 1997, score: 0,636).