

Analyser un corpus illisible? Le logiciel Alceste confronté à des registres de condoléances

Gérôme Truc

Institut Marcel Mauss (EHESS/CNRS), Paris

Casa de Velázquez, Madrid

gerome.truc@cvz.es

Qu'il s'intéresse à la mort d'un chef d'État, d'un Pape, ou d'une icône pop, à un accident industriel, une catastrophe naturelle, ou encore un attentat, le chercheur en sciences sociales peut être confronté à un matériau dont il ne sait pas toujours bien quoi faire : des registres de condoléances¹. De plus en plus, en effet, l'ouverture de registres de condoléances – à l'instar de l'observation de minutes de silence – devient un réflexe des pouvoirs publics confrontés à des situations de deuil collectif, quelle qu'en soit la cause, afin de canaliser l'expression des émotions populaires. De ce fait, la pratique consistant à rédiger un message de condoléances dans de tels registres, tout comme la signature d'une pétition (Contamin, 2005), tend de nos jours à s'effectuer dans un cadre normatif relativement lâche et malléable, tolérant une pluralité d'engagements de la part de chacun et par conséquent une assez grande variété de types de messages, des plus informels aux plus personnalisés, des plus succincts aux plus diserts.

-
1. Ce texte a initialement fait l'objet d'une présentation orale lors de la Journée d'études sur la méthodologie Alceste, organisée par Max Reinert et Choeb Zafar à Carcassonne, le 21 août 2009 (cf. *Bulletin de Méthodologie Sociologique*, octobre 2009, n° 104, p. 39-46). Il a aussi bénéficié de ma participation à la Journée d'études « Sciences érudites et écritures contemporaines », organisée par Claire Bustarret, Béatrice Fraenkel et Yann Potin, à l'EHESS-Paris, le 20 janvier 2010. Je tiens à remercier les participants à ces deux journées d'études pour leurs remarques et critiques, et à adresser tout particulièrement ma reconnaissance aux organisateurs.

Pour reprendre les catégories de Simmel, les messages de condoléances sont aujourd'hui une « forme » qui accepte un grand nombre de « contenus » différents.

Savoir qu'un message est « de condoléances » renseigne dès lors moins sur son contenu que sur la posture de la personne l'ayant rédigé : celui-ci prend part à la douleur de celui à qui il adresse son message (c'est le sens étymologique de « condoléance » : *con-dolere*, souffrir avec), mais l'on ne peut préjuger des mots et des formules qu'il emploie pour le manifester. C'est pourquoi, comme l'a très bien souligné Pierre-Yves Baudot, les registres de condoléances constituent, pour un chercheur, un matériau à la fois particulièrement riche et difficile à appréhender : « Face à cet impressionnant amas de témoignages d'affliction, le chercheur se retrouve paradoxalement démuné. Qu'ils soient d'une affligeante banalité – longue litanie de « sincères condoléances » – ou davantage personnalisés – sincérité de l'émotion exprimée par l'originalité des mots employés –, ces messages constituent une source délicate à utiliser » (Baudot, 2007 : 71).

Tout d'abord, il est en général très difficile de croiser le contenu des messages avec des variables classiques (âge, sexe, profession). Ensuite, comme le note Béatrice Fraenkel (2002) à propos des messages rédigés à la suite des attentats du 11 septembre 2001, ces messages sont répétitifs et le plus souvent d'une assez grande platitude : la plupart des formules employées sont convenues et schématiques, ce qui rend d'emblée problématique toute analyse de contenu *stricto sensu*. Enfin et surtout, les registres de condoléances constituent généralement des corpus pléthoriques impossibles à lire linéairement, et plus encore à appréhender de façon exhaustive. Ce sont, par exemple, des milliers de messages éparpillés dans tout Manhattan à la suite du 11-septembre (Fraenkel, 2002), plus de 12 mètres de linéaire aux archives municipales de Londres après les attentats du 7 juillet 2005 (Truc, 2010), ou encore 340 registres archivés à la mort du Général De Gaulle (Hazareesingh, 2010). Y a-t-il, dès lors, quelque espoir de parvenir à analyser un corpus impossible à lire en intégralité ?

Après avoir passé en revue les stratégies mises en place pour contourner ces difficultés par les quelques chercheurs s'étant déjà confrontés à des messages de condoléances, nous en viendrons, dans un second temps, au cœur de notre argumentation : déterminer dans quelle mesure ces messages constituent un corpus particulièrement approprié à l'utilisation du logiciel d'analyse textuelle Alceste, développé par Max Reinert. Ce logiciel rencontre un succès croissant auprès de chercheurs en sciences sociales qui l'utilisent fréquemment comme un assistant dans l'analyse de retranscrip-

tions d'entretiens ou de réponses à des questions ouvertes. Pour autant, il n'est pas évident que ce soit avec ce type de corpus que les possibilités d'Alceste soient le mieux exploitées. Ainsi, nous présenterons pour finir un cas d'application concrète: les premiers résultats d'une analyse Alceste des messages rédigés à la suite des attentats du 11 mars 2004 à Madrid. Il s'agit de presque 60 000 messages électroniques enregistrés de juin 2004 à mars 2005, via un dispositif informatique mis en place dans les gares de Madrid frappées par les attentats². Mis bout à bout, ils constituent un corpus de presque 99 000 mots et 6 millions de caractères, dont on peut considérer qu'il joue avec les limites d'Alceste³.

1. Un défi méthodologique: comment analyser des registres de condoléances ?

La stratégie la plus fréquemment adoptée par les chercheurs confrontés à un volume pléthorique de messages de condoléances consiste, assez logiquement, à n'en retenir pour l'analyse qu'un échantillon limité. Toute la question est alors celle des principes qui président à cette sélection, et des éventuels biais qu'ils induisent dans les résultats de l'analyse. Une première façon de faire consiste à parcourir quelques registres pour n'y retenir que les messages qui paraissent les plus intéressants, les plus originaux, ou les plus pertinents. C'est par exemple de cette manière que semble procéder Sudhir Hazareesingh lorsqu'il dépouille les registres de condoléances ouverts à la suite du décès du Général De Gaulle (Hazareesingh, 2010). Mais cette méthode suppose, premièrement, d'avoir déterminé *a priori* les critères de pertinence ou d'originalité à l'aune desquels évaluer les messages et, deuxièmement, de laisser de côté tous les messages trop pauvres en contenu, trop schématiques ou répétitifs dans leur formulation – dont on a pourtant de bonnes raisons de penser, comme on l'a souligné en introduction, qu'ils représentent l'essentiel des messages contenus dans des registres de condoléances...

-
2. Ce dispositif resta en place jusqu'en juillet 2007. Au total, près de 120 000 messages furent enregistrés, notre corpus correspondant donc à la première moitié d'entre eux, enregistrés depuis la mise en place du dispositif dans les gares jusqu'au premier anniversaire des attentats. Tous ces messages font partie du fond d'archives ethnographiques organisé dans le cadre du projet de recherche « *Archivo del Duelo* » au *Centro Superior de Investigaciones Científicas* (CSIC) à Madrid, avec le soutien financier du Ministère espagnol de l'Éducation et de la Science (Projet HUM-2005-03490) et la collaboration de la *RENFE* (compagnie de chemins de fer espagnol). Pour de plus amples détails sur ce dispositif, voir Truc (2011).
 3. Ce corpus est trop volumineux pour être traité par la version commerciale du logiciel et, à titre de comparaison, le corpus d'étalonnage du logiciel comporte 26 millions de caractères.

Une seconde manière de procéder, sans doute moins arbitraire, consiste à composer un échantillon aléatoire de messages. C'est pour cette démarche qu'opte notamment Bethan Jones dans son analyse des registres ouverts à la mort de Lady Diana (Jones, 1999) : parmi l'ensemble des registres ouverts et collectés dans tout le Royaume-Uni, elle choisit d'abord de se concentrer sur ceux d'une ville moyenne du sud de l'Angleterre puis d'en extraire « au hasard » (mais sans préciser comment) très exactement 1808 messages, auxquels elle en ajoute 1 017 autres provenant d'Internet. Elle classe ensuite ces messages suivant différentes catégories thématiques : messages exprimant de la sympathie pour Diana, messages blâmant les responsables de sa mort, messages d'enfants, messages adressés à Dodi Al-Fayed, messages à dimension religieuse, etc. C'est, *mutatis mutandis*, une méthode similaire que les membres de l'Association des Familles du 11-septembre emploient pour inventorier les messages rédigés par les visiteurs du *Tribute WTC Visitor Center* ouverts à *Ground Zero* cinq ans après les attentats⁴. Or, ce qui frappe, lorsque l'on ouvre les boîtes où sont archivés ces messages, c'est qu'en définitive la majorité d'entre eux ne rentrent dans aucune des catégories, et restent assemblés sous l'étiquette « divers » ou « à classer ».

Chercher à ordonner thématiquement les messages de condoléances suivant des catégories *a priori* ou élaborées à partir du dépouillement des premiers d'entre eux peut ainsi conduire à une survalorisation des messages conformes aux attentes préalables et, inversement, à une certaine cécité à l'égard d'autres messages pouvant très bien, en définitive, constituer à majorité du corpus. En procédant de la sorte, le chercheur introduit donc un biais de sélection, qui découle de l'écart existant entre ce qui retient son attention dans ce qu'il lit et le reste des messages de condoléances qui lui paraissent au contraire creux, répétitifs, et dont il peine à dire de quoi ils parlent. Le risque, ici, est de faire passer les exceptions pour la règle, et de faire primer, dans l'analyse des messages de condoléances, ceux qui dénotent, alors qu'ils ne peuvent en aucun cas être tenus pour représentatifs de l'ensemble du corpus étudié. Autrement dit, on risque de tout simplement manquer l'essentiel.

Ceci n'est à vrai dire problématique que si l'on souhaite réaliser une analyse du contenu des messages de condoléances. Or, le chercheur peut

4. Pour de plus amples informations sur ce centre, voir le site : www.tributewtc.org. Pour un aperçu des messages en questions (aux archives intégrales desquels nous avons eu accès) et des catégories de classements, voir en particulier la page : www.tributewtc.org/exhibits/visitor_cards.html (consulté le 24 juin 2010).

tout à fait préférer s'intéresser, plutôt qu'au contenu du message, à l'acte d'écriture (Fraenkel, 2007) qui se situe en amont de celui-ci. Et de ce point de vue, comme le remarque Clara Lamireau après avoir travaillé sur un corpus de 1 572 pages d'intentions de prière, dont les caractéristiques sont assez proches de registres de condoléances, une analyse exhaustive de contenu est sans intérêt (Lamireau, 2009 : 106). Un autre angle approche doit dès lors être privilégié, suivant lequel l'analyse du corpus, bien que partielle, ne sera pas nécessairement partielle. Il s'agit notamment, comme le fait Pierre-Yves Baudot (2007) dans son analyse des registres de condoléances ouverts lors des décès des présidents de la République française, de se focaliser plutôt sur les modalités d'énonciation et de signatures des messages, prises comme indicateurs des modalités d'engagement des scripteurs dans l'acte d'écriture⁵.

Bien entendu, ces deux méthodes ne sont pas nécessairement exclusives, et c'est pourquoi, dans notre étude exploratoire des messages rédigés après les attentats du 11 mars 2004, nous avons tenté de combiner un relevé des modalités d'énonciation et de signature des messages avec une certaine forme d'analyse de contenu (Truc, 2006). Pour ce faire, nous avons notamment procédé à une exploration par « mots-clés » de l'ensemble du corpus. Cette méthode n'était possible que parce que celui-ci était entièrement informatisé – condition *sine qua non*, du reste, pour soumettre des messages de condoléances à un logiciel de statistiques textuelles tel qu'Alceste. C'est ainsi que, dès le départ, nous avons considéré les résultats de cette première exploration comme *une* grille de lecture de notre corpus parmi d'autres, dans l'attente d'une analyse plus systématique, réalisée à l'aide d'Alceste. Car, contrairement à ce que l'on peut parfois lire, Alceste ne peut être considéré comme un équivalent strict de logiciels lexicométriques comme, par exemple, Lexico ou Hyperbase car, de l'aveu même de son créateur, sa finalité principale n'est pas de produire des analyses de contenu (Reinert, 1990). Bien au contraire, Max Reinert, qui a conçu son logiciel comme une réponse aux limites de l'analyse de contenu, a voulu faire en sorte que celui-ci « réconcilie d'une certaine manière l'objet d'une analyse de contenu avec le sujet d'une analyse de discours » (Reinert, 2006). Ainsi, Alceste, de la même manière

5. Renoncer à une analyse exhaustive du contenu des messages et concentrer l'analyse sur les modalités d'engagement des scripteurs dans leur acte d'écriture va de soi dans le cas des corpus qui n'ont, pour ainsi dire, ni début ni fin. C'est le cas particulièrement de messages de condoléances qui ne sont pas contenus dans des registres mais déposés ou affichés aux quatre coins de la ville, comme ce fut le cas à New York après le 11 septembre (Fraenkel, 2002) ou à Madrid en mars 2004, avant que ce ne soit mis en place le dispositif informatique de collecte des messages (Sánchez Carretero, 2006).

qu'un logiciel comme *Prospéro*, prête également attention aux modalités énonciatives et discursives des textes qui lui sont soumis. Il met sur un même plan le lexique (dans le langage du logiciel : les « mots-pleins ») et les indices d'une posture énonciative, les modalisateurs, les marqueurs syntaxiques (les « mots-outils »), qui permettent de contextualiser l'analyse du contenu lexical et de donner ainsi une plus grande cohérence aux « mondes lexicaux » dégagés par Alceste. En cela, il nous a semblé qu'il pouvait être d'une aide appréciable pour quiconque souhaite appréhender des registres de condoléances dans leur exhaustivité sans réifier leur contenu, et sans non plus les dissocier des actes d'écritures dont ils sont les produits, ou plus exactement : les traces.

2. Les registres de condoléances : un corpus idéal pour le logiciel Alceste ?

Ce que vise Alceste, selon Max Reinert, c'est « non pas décrire un objet que représenterait le texte du discours, mais analyser les *traces* de l'activité discursive qui le fonde » (Reinert, 2007 : 191). On semble dès lors tenir un outil particulièrement adapté à une recherche qui considère les messages de condoléances comme les *traces* d'actes d'écriture, dont la prise en compte dans l'analyse est tout à fait primordiale. Rédiger un message de condoléances n'est en effet pas chose aisée : c'est, au sens que Boltanski et Thévenot (1991) donnent à ce terme, une *épreuve*. Comme nous l'avons montré ailleurs à partir d'une étude ethnographique portant sur les usages du dispositif permettant d'enregistrer des messages de condoléances dans la gare d'Atocha à la suite du 11 mars 2004 (Truc, 2011), se joue d'abord ici une « épreuve de l'émotion », consistant en « une extériorisation de l'intériorité » (Boltanski et Godet, 1995 : 39). Il faut parvenir à trouver les mots justes pour exprimer ce que l'on ressent. Mais pour que les mots soient *justes*, ils doivent non seulement correspondre au ressenti intérieur du scripteur, mais aussi respecter les contraintes propres à la situation (Baudot, 2007 : 73)⁶. Autrement dit, celui qui se met en posture d'écrire un message tient compte du fait que, par exemple, il ne doit pas écrire quelque chose de trop long, éviter de faire des fautes d'orthographe ou des ratures, éviter aussi d'utiliser un vocabulaire trop relâché, etc. Il doit également, et surtout, trouver le moyen de symétriser sa relation avec celui à qui il s'adresse et, dans ce but, déterminer à quel titre il rédige son message : en son nom propre, au nom de sa famille, en délégation d'une association ou d'une communauté dont il serait le représentant, etc.

6. C'est, du reste, le lot de toutes les situations « d'expression obligatoire des sentiments » (Mauss, 1969).

Signe de cette difficulté, il n'est pas rare de tomber sur des messages de condoléances où le scripteur se résigne à écrire « Je ne sais pas quoi écrire », ou « Il n'y a pas de mots pour dire ce que je ressens... ». Comme le souligne Béatrice Fraenkel, c'est face à cet écueil que se révèle le pouvoir de la formule conventionnelle, toute faite, qui « donne forme à la difficulté de dire » (Fraenkel, 2002 : 61). Ainsi s'explique la litanie des messages-types que l'on retrouve dans tout registre de condoléances, quel que soit le lieu, la date, ou la langue : « Sincères condoléances », « Nos pensées sont avec vous », « Pour toujours dans nos cœurs », « Siempre estaréis en nuestro corazón », « God Bless », « Rest in Peace », « Our deepest sympathies », « In our thoughts », « You will never be forgotten », « No os olvidaremos », etc. De là découle la nette impression de répétitivité qui saisit toute personne entreprenant de lire linéairement un registre de condoléances. Toutes ces formules réitérées sans fin peuvent, en un certain sens, être considérées comme autant de « ratages » à exprimer de manière plus juste, plus personnelle et moins conventionnelle, le partage de la douleur.

À cet égard encore, la concordance avec les principes présidant au fonctionnement d'Alceste ne manque pas de frapper, puisque celui-ci vise « une mise à plat de ce cycle des répétitions, de ce cycle des ratages à énoncer ce qui ne peut être dit ; c'est-à-dire, une prise de conscience que quelque chose insiste et se répète malgré tous ces ratages » (Reinert, 2000 : 2). Selon Max Reinert, en effet, la signification d'un texte n'est jamais donnée *a priori*, mais ne s'élabore qu'au travers de répétitions qui traversent le texte et contribuent à stabiliser l'activité discursive : la répétition au sein d'un texte des mêmes mots, des mêmes formules « évoque une tension, un retour de quelque chose que l'auteur ne peut jamais arriver à formuler complètement puisqu'il doit le reprendre sans cesse », et qui constitue en définitive l'objet réel de son discours (Reinert, 2000 : 5). C'est pourquoi le repérage des répétitions et des co-occurrences joue un rôle tout à fait central dans la méthode mise en œuvre par Alceste (Reinert, 2003)⁷. Ce principe en tête, on comprend mieux que des corpus chargés émotionnellement, où ce qui cherche à se dire ne fait qu'affleurer, où les répétitions et les ratages abondent, dont la trame ne semble faite que d'approximations successives, puissent aboutir à des résultats intéressants lorsqu'ils sont soumis à Alceste. Ainsi, les registres de condoléances ne constituent pas seulement un corpus particulièrement approprié au logiciel Alceste mais, par sa logique

7. Peut-être convient-il d'ailleurs ici de rappeler de quoi Alceste est l'acronyme : *Analyse des Lexèmes Cooccurents dans un Ensemble de Segments de Texte*.

même, Alceste permet en outre de vérifier l'affirmation selon laquelle ces registres seraient répétitifs et plus encore, le cas échéant, de déterminer ce qui, très précisément, s'y répète.

Soumettre des messages de condoléances à Alceste implique néanmoins de renoncer à les considérer isolément les uns des autres. Il convient, au contraire, de les aborder dans leur globalité comme ne constituant qu'un seul et unique texte. Car, bien entendu, c'est entre eux que les messages sont répétitifs et c'est à l'échelle du corpus entier qu'une impression de répétitivité se dégage et demande à être vérifiée. Corrélativement, il faut donc faire l'hypothèse qu'à ce grand texte unique correspond un auteur collectif, un sujet pluriel. C'est précisément ce à quoi nous invite Béatrice Fraenkel dans son étude des messages affichés dans Manhattan à la suite du 11 Septembre, dont elle juge qu'ils sont l'œuvre d'un « sujet monumental » (Fraenkel, 2002 : 51). Cette approche, du reste, est peu coûteuse méthodologiquement – puisque, comme nous l'avons souligné d'entrée, il est rare de disposer d'informations exploitables sur les caractéristiques sociales de chaque auteur d'un message – et fondée sociologiquement – pourvu que l'on veuille bien considérer que l'ouverture de registres de condoléances répond à des situations « d'effervescence collective » où les membres d'une même communauté, partageant le même état d'esprit sous le coup du deuil collectif, renforcent leur cohésion en estompant leurs individualités (Durkheim, 1979 ; Collins, 2004).

Ceci étant dit, soumettre des messages de condoléances à Alceste ne va pas sans poser certains problèmes qu'il nous faut aussi considérer. Tout d'abord, cela suppose de disposer d'une version numérisée de ceux-ci. Les messages rédigés à la suite du 11 mars 2004 sur lesquels nous avons travaillé constituaient d'emblée un corpus numérique. Mais c'est un cas de figure pour l'instant encore très rare, pour ne pas dire unique en son genre. En dépit de l'essor des NTIC et de la démocratisation de l'accès à l'informatique, rien ne permet véritablement d'affirmer que les registres numériques de condoléances se développeront dans un futur proche⁸. Tout au contraire, on a plutôt de bonnes raisons de penser que les registres manuscrits perdureront, étant donné la prévalence de l'acte d'écriture dans la pratique consistant à signer un registre de condoléances. La trace

8. Il convient toutefois de noter la récente création d'un site Internet espagnol, conçu sur le modèle de Facebook, qui permet de créer une page personnelle pour chaque défunt, anonyme ou célèbre, sur laquelle quiconque peut publier un message électronique de condoléances (et/ou allumer une « bougie » virtuelle, tout en s'informant sur la tenue des obsèques, etc.): www.rememori.com

manuscrite joue en effet un rôle primordial dans la personnalisation du message (notamment parce qu'elle atteste de l'émotion ressentie par le scripteur, qui se laisse deviner par la graphie). Plus les messages sont simples et répétitifs et plus il importe d'attester que c'est bien « une main différente de l'autre qui a écrit » (Fraenkel, 2002 : 63), ce que ne permet pas un clavier informatique⁹. Or, la reconnaissance automatique de caractères n'étant pas possible à partir d'écritures manuscrites, ne disposer que de registres de condoléances manuscrites suppose nécessairement, si l'on veut les soumettre à une analyse informatisée, de réaliser soi-même la saisie informatique du corpus et, par conséquent, d'échantillonner d'une manière ou d'une autre le corpus initial – et l'on retrouve alors les problèmes évoqués précédemment. Mais s'il est peu probable que des registres numériques se substituent entièrement à des registres manuscrits de condoléances dans un futur proche, cela n'empêche pas que, de plus en plus souvent, des sites ou forums Internet recueillent de nombreux messages électroniques de condoléances. Rien n'interdit alors de combiner une analyse « classique » de registres de condoléances manuscrites avec une analyse informatique de messages électroniques, l'une et l'autre pouvant se compléter utilement.

Reste toutefois un important problème à régler : les messages électroniques susceptibles d'être soumis à une analyse Alceste comportent généralement nombre de coquilles et fautes d'orthographe risquant de perturber le logiciel. Dans quelle mesure peut-on alors entreprendre de « nettoyer » le corpus ? À cette question, il nous semble qu'il faut répondre : en aucun cas. Rien ne permet en effet au chercheur de distinguer *a posteriori* une faute d'orthographe d'une faute de frappe, ni de savoir si une faute relève d'un manque de compétence (orthographique ou technique) du scripteur plutôt que d'une perturbation émotionnelle de ce dernier lors de l'acte d'écriture. Autrement dit, il est envisageable de considérer qu'un message comportant plus de fautes qu'un autre a été rédigé par un scripteur plus ému, la coquille pouvant à ce titre être considérée comme un équivalent électronique de la rature manuscrite. Laisser les messages en l'état, avec toutes leurs imperfections, ne s'avère en outre, à l'essai, pas si problématique que cela. En effet, tant que ces fautes

9. Le problème est ici celui de l'absence à ce jour d'un véritable équivalent numérique, en termes de singularisation et d'identification de la personne, à la signature manuscrite (Fraenkel et Pontille, 2006). Pour pallier cette difficulté, le dispositif mis en place à Madrid après les attentats du 11 mars 2004 permettait à chaque scripteur de scanner sa main et d'ajouter à son message cette empreinte numérique, jouant le rôle d'un substitut de la signature manuscrite.

sont marginales (ce qui est par définition le cas d'une simple coquille), elles n'empêchent pas Alceste de procéder à la lemmatisation du lexique, et de rattacher par conséquent à une même forme réduite aussi bien les formes complètes correctement orthographiées que celles qui ne le sont pas. Ainsi, entre autres exemples issus de notre corpus (en espagnol), on peut relever que « mundoen » est bien rattaché à la forme réduite « mundo+ », ou que « terrorimos », « terrorita » ou « terroritas » le sont aussi à la forme réduite « terror+ ». Il en va globalement de même pour les messages faisant usage de mots abrégés ou écrits en « langage SMS » où, par exemple, la forme « qu » est remplacée par un k (« aki » pour « aqui », « akella » pour « aquella ») : si en raison de ces modifications orthographiques Alceste ne peut plus faire le lien entre, par exemple, « quiero » et « kiero », cela ne l'empêche pas pour autant de rattacher à la forme réduite « kier+ » les formes complètes du verbe conjugué « kiera », « kiere », « kieren » « kiero », exactement comme il le fait avec la forme correcte du radical « quier+ », et de prendre ensuite en compte ces formes dans son analyse. Ces erreurs restant, à l'échelle des milliers des messages, marginales, elles n'ont qu'un faible poids dans les calculs effectués par Alceste : par une sorte de loi des grands nombres, plus un mot est répété dans le corpus, plus ces variations orthographiques sont vouées à rester négligeables. L'essentiel est que, étant reconnues et lemmatisées, elles ne semblent donc pas contribuer à accroître la part du corpus non-analysée par le logiciel.

Enfin, une dernière difficulté peut être posée par des messages rédigés en plusieurs langues et qui composent ainsi un texte plurilingue, alors qu'Alceste, bien qu'il puisse analyser des textes en différentes langues, n'est capable de ne considérer qu'une seule langue à la fois. Notre expérience en la matière témoigne du fait que, une fois encore, ce problème n'en est pas un. Tout d'abord, le principe même de l'algorithme présidant au fonctionnement d'Alceste, fondé sur une recherche des co-occurrences, aboutit automatiquement à la constitution d'une classe spécifique regroupant la plupart des messages ou segments de messages rédigés dans des langues minoritaires par rapport au reste du corpus. Ainsi, dans le corpus que nous avons analysé, à côté de plusieurs classes en espagnol, Alceste en a constitué une regroupant des messages en anglais, français, portugais, basque, catalan, etc. Une seconde analyse sur ce sous-corpus plurilingue a abouti de nouveau à une ségrégation des messages par langues majoritaires, et à la production d'un résidu, pouvant encore être affiné jusqu'à obtenir des sous-corpus n'ayant plus la taille suffisante pour être analysés. Surtout, il faut souligner que dans cette classe plurilingue ne se retrouvent

en définitive que des mots étrangers n'ayant aucune racine commune avec un mot espagnol. Dans le cas contraire, au bénéfice d'une racine commune, des mots catalans, français ou anglais par exemple peuvent très bien être rattachés à la même forme réduite qu'un mot espagnol. Dans notre corpus, c'est par exemple le cas de la forme « fanatism+ » qui renvoie aussi bien à « fanatismo » (en espagnol) qu'à « fanatisme » (en français) ou « fanaticism » (en anglais).

3. Une application :

les « mondes lexicaux » des messages de condoléances rédigés après les attentats du 11 mars 2004 à Madrid

Afin d'illustrer ce à quoi peut aboutir l'analyse exhaustive d'un ensemble de messages de condoléances au moyen du logiciel Alceste, nous allons maintenant présenter les premiers résultats d'un travail que nous avons mené sur les messages rédigés à la suite des attentats du 11 mars 2004 à Madrid. Étant donné son volume, ce corpus a été soumis à plusieurs analyses successives ayant permis de le diviser en sous-corpus, jusqu'à obtenir des résultats à la fois robustes et interprétables. Ainsi, 68,5 % du corpus initial composé de 58 732 messages, sans aucune retouche ni correction, a d'abord été classé par le logiciel en trois mondes lexicaux A, B et C (correspondant respectivement à 42, 47 et 11 % du corpus analysé). Puis les analyses suivantes, menées sur ces sous-corpus A, B et C issus de cette première analyse, ont permis à nouveau de classer 67 % du sous-corpus A, 77 % du sous-corpus B et 83 % du corpus C. Le sous-corpus C, qui peut d'emblée être tenu pour résiduel, correspond à l'ensemble des messages en différentes langues étrangères dont le logiciel repère immédiatement qu'ils se distinguent de la masse de ceux rédigés en espagnol. Comme nous l'avons expliqué plus haut, les analyses suivantes de ce sous-corpus conduisent à un affinement par langue. Les sous-corpus A et B, eux, sont de prime abord plus difficiles à caractériser. Mais, fait significatif, les secondes analyses menées sur chacun d'eux aboutissent à des résultats remarquablement équivalents et cette fois-ci très clairement caractérisables¹⁰. Mis à part une nouvelle classe résiduelle

10. Dans la caractérisation de chaque monde lexical qui suit, nous ne retenons que les mots dont le Khi2 d'association à une classe est égal ou supérieur à 80 dans au moins l'un des deux sous-corpus, A ou B. La plupart de ces mots ayant été retenus par Alceste comme représentatifs de la même classe dans les deux sous-corpus, nous donnons pour chacun d'eux entre parenthèses d'abord son Khi2 d'association à la classe dans le sous-corpus A puis celui dans le sous-corpus B (ou sinon seulement dans l'un des deux). Puis nous citons aussi, à chaque fois, un extrait des UCE (Unités de contexte élémentaires) les plus significatives de chaque classe dans le sous-corpus

produite par l'analyse du sous-corpus A, Alceste dégage au sein des deux sous-corpus, à la fois dans le même ordre et dans des termes extrêmement proches – ce qui nous paraît un bon indicateur de robustesse des résultats –, trois « mondes lexicaux » (chacun correspondant à environ un tiers des deux sous-corpus réunis). Il s'agit bien de mondes lexicaux, et pas seulement de champs lexicaux, car les éléments syntaxiques et les indices de la posture énonciative comptent autant que le simple lexique dans leur caractérisation.

Le premier de ces trois mondes (correspondant à 26 % du sous-corpus A et 43 % du B) pourrait être décrit comme celui des exhortations pacifistes, si l'on en juge d'abord par les mots qui lui sont propres : « monde » (140 et 177), « humain » (124 et 102), « violence » (143 et 109), « paix » (152), « haine » (156 et 75), « religion » (94 et 86), « guerre » (93 et 78)¹¹. En ce qui concerne les modalités énonciatives, on note une moindre présence que dans le reste du corpus des pronoms personnels, des adjectifs et des adverbes, tandis qu'on trouve bien plus de substantifs rédigés en majuscules et de formules négatives. Les messages sont assez impersonnels et composés de phrases courtes (parfois juste de simples slogans), qui disent un refus des attentats et du terrorisme, mais sans que leur destinataire ne soit toujours bien identifié : « nous continuerons à lutter pour la non-violence. Paix et amour pour tous. Plus de guerres, qui finissent par semer la terreur dans tous les pays du monde et faire souffrir aux gens des horreurs comme celles du 11-M ou du 11-S, et de beaucoup d'autres encore ». « MAINS BLANCHES POUR LA PAIX. Nous voulons un monde libre, sans terrorisme. Seuls des monstres aveuglés par leurs croyances et sans sentiments sont capables de commettre des attentats pareils. Luttons pour un monde libre, pour la paix mondiale, contre la barbarie de la mort qui frappe aveuglément, et contre toutes les manières existantes de tuer ».

Le second monde lexical (32 % du sous-corpus A et 24 % du B) correspond, lui, à des évocations de la mort et des expressions du deuil qui lui est associé : « train » (109 et 116), « chaque » (136 et 128), « larmes » (45 et 133), « pleurer » (172), « jour » (72 et 121), « ciel » (36 et 105), « étoiles » (84), « Atocha » (92)¹². Les modalités énonciatives de ces mes-

A, puis B. Bien que l'analyse ait porté sur un corpus en espagnol, tous les termes et les messages cités ici ont été directement traduits en français.

11. Alceste procédant une lemmatisation, chacun de ces mots correspond à une forme réduite, et renvoie donc aussi aux différentes déclinaisons du terme, par exemple, pour « monde » : mondial, mondialement, mondiaux, etc.
12. Atocha est le nom de la gare de Madrid où explosèrent les trois premières bombes et vers laquelle convergeaient tous les autres trains frappés par les attentats le 11 mars 2004 au matin.

sages sont presque à l'exact opposé des précédentes : plus de marqueurs de la personne et moins de verbes modaux. Ces messages se caractérisent également par une plus grande présence d'adjectifs, d'adverbes, de couleurs et de marqueurs d'intensité. Il s'agit, enfin, de messages à la fois moins contextualisés que d'autres (moindre présence de prépositions que dans le reste du corpus) et mieux structurés (plus de conjonctions de coordination) : les scripteurs ne paraissent pas, ici, en proie à un quelconque désarroi linguistique. Ils trouvent au contraire les mots qui conviennent à ce qu'ils veulent exprimer, et rédigent des messages assez longs et élaborés, pouvant avoir une certaine qualité littéraire. « Comme toujours, des trains qui passent, chargés à ras bord de personnes. Comme dans un film, une secousse sur la caténaire, trois explosions, et tout d'un coup, la stupeur, la confusion, la peur ». C'est ici que figurent généralement les messages directement adressés par un proche à une victime : « Cela fait trois mois que je pleure. Tous les jours. Je ne peux plus dormir dans le train. Je continuerai à pleurer, chaque jour, avec tes parents, ton frère et ta fiancée, jusqu'à ce qu'il ne me reste plus de larmes ». On peut donc supposer que ces messages sont ceux dont la rédaction a pris le plus de temps, et qu'ils sont les traces d'un engagement fort dans l'acte d'écriture.

Enfin, le troisième et dernier monde dégagé par Alceste (30 % du sous-corpus A et 32 % du sous-corpus B) est constitué, lui, d'adresses compassionnelles clairement adressées aux victimes, des élans du cœur dont la formulation est généralement moins bien articulée que celle des messages précédents. Il y est question de « famille » (560 et 691), « soutien » (271 et 299), « condoléances » (260 et 184), « victime » (205 et 203), « affectés » (68 et 223), « amis » (59 et 118), « courage » (152), « aimé » (175 et 111), « sincère » (113 et 114), « abrazo » (123 et 113) (un terme très courant en espagnol que l'on pourrait traduire par « je vous embrasse »). Les auxiliaires être et avoir, de même que les quantificateurs, y sont significativement plus présents qu'ailleurs, ce qui semble indiquer que ces messages expriment la façon dont les scripteurs ont vécu les attentats et l'émotion qu'ils ont ressentie (ils sont par conséquent rédigés au passé). On y trouve donc plus encore que précédemment des phrases formulées à la première personne (du singulier ou du pluriel). La présence significative dans ces messages des noms de pays constitue également une autre marque d'un « je » assumé, puisque faire mention de son pays (et/ou de sa ville) d'origine est une manière, pour les scripteurs, de singulariser le message et d'attester son authenticité (Truc, 2006). Les messages de ce dernier monde lexical sont donc des assertions performatives plutôt courtes, comme les exhortations du premier monde lexical, mais cette fois-ci formulée à la première

personne, et explicitement adressées aux victimes (dont les scripteurs ne sont généralement pas ici des proches) depuis un lieu précis: « Nous vous envoyons tout notre soutien et toutes nos forces pour continuer à aller de l'avant. Nous vous embrassons très fort. Depuis l'Euskadi. Solidarité avec les victimes. / Depuis Tenerife, nous tenons à présenter nos plus sincères condoléances à toutes les familles des victimes de ce si brutal attentat. Nous vous garderons tous en mémoire ». « Je viens des Canaries et, parmi les victimes, il n'y a personne qui me soit proche, mais j'ai senti et je continue à sentir que tous et toutes sont pour moi comme de la famille ou des amis. Ma douleur est très grande. Mon admiration va à tous ceux qui les ont aidés, et mon affection et mes forces à toutes les familles des victimes. À toutes les personnes qui ont aidé, merci, et à aux familles, mon affection ».

Comme on le voit, il n'y a finalement que dans ce troisième et dernier monde lexical qu'il est explicitement fait mention de « condoléances ». Il apparaît donc que seule une minorité des messages rédigés à la suite des attentats du 11 mars 2004 que nous avons analysés sont effectivement des messages qui adressent aux victimes des « condoléances ». Ce constat relève à la fois de l'analyse de contenu (la présence effective du champ lexical des « condoléances » dans les messages) et d'une étude des modalités d'énonciation des messages du corpus: tous ne sont pas, en effet, rédigés et assumés à la première personne, par un signataire singulier s'adressant à un destinataire précis (qu'il s'agisse d'une victime en particulier, ou de leurs proches et familles en général). Aux autres mondes lexicaux dégagés par l'analyse Alceste correspondent aussi des messages impersonnels (c'est le cas en particulier des exhortations pacifistes), dont les destinataires peuvent être plus flous. Ce résultat indique clairement que, derrière chaque message, il y a un engagement spécifique dans l'acte d'écriture consistant à rédiger un message de condoléances à la suite d'un attentat. L'enquête demande dès lors à être complétée par une étude ethnographique *in situ* de ces engagements aux modalités multiples dans l'acte d'écriture (Truc, 2011).

Conclusion

Les registres de condoléances sont un matériau difficile à exploiter pour le chercheur. Une présomption de vacuité et de répétitivité de leur contenu, ajoutée à un volume généralement pléthorique qui rend vaine toute entreprise de lecture exhaustive, semble limiter toute possibilité d'analyse de contenu, qu'elle soit thématique ou lexicométrique. Mais il est envisageable que ces faiblesses du matériau deviennent en définitive ses forces, pourvu qu'on le soumette à un outil adapté. Il nous a semblé,

comme nous avons voulu le montrer dans cet article, que c'est le cas du logiciel Alceste. Celui-ci en effet, en ne réifiant pas le contenu du texte qui lui est soumis, mais en le considérant au contraire comme la trace d'une activité discursive préalable, est particulièrement adapté pour appréhender dans leur globalité les formes textuelles auxquelles aboutissent des engagements différenciés dans l'acte d'écriture particulier que constitue la rédaction d'un message de condoléances. Il l'est d'autant plus que, cette stabilisation des formes textuelles, le logiciel la repère par une attention portée aux répétitions. Ainsi, soumettre un corpus constitué de messages de condoléances à Alceste permet de vérifier si oui ou non leurs contenus sont effectivement répétitifs (plus ils le seront, et plus il est probable que l'analyse produite par le logiciel couvrira une large partie du corpus) et, plus important encore, de déterminer *ce qui* s'y répète exactement.

En ce qui concerne le corpus que nous avons soumis au logiciel, composé de messages de condoléances directement enregistrés au format électronique à la suite des attentats du 11 mars 2004 à Madrid, il est apparu que l'on pouvait y distinguer précisément trois types de formulations répétées, qui renvoient chacune à des modalités différentes d'énonciation et d'engagement dans l'acte d'écriture : les exhortations pacifistes, les expressions du deuil et les adresses compassionnelles. Or, seule cette dernière catégorie correspond en définitive à l'idée que l'on se fait communément d'un message de condoléances. Ce résultat, au demeurant, ne fait que confirmer l'observation par laquelle nous ouvrons cet article, à savoir que la rédaction d'un message dans un registre de condoléances est un acte d'écriture au cadre normatif aujourd'hui relativement relâché, tolérant des types d'accomplissement variés qui interdisent de considérer un registre de condoléances comme un matériau textuel uniforme et univoque. Nous espérons simplement avoir montré dans cet article que, face à cette difficulté, le recours au logiciel Alceste permet, plutôt que de renoncer à toute analyse du contenu, de ne pas dissocier celle-ci d'une attention aux situations d'engagement dans l'acte d'écriture dont les messages de condoléances sont les traces.

Références bibliographiques

- BAUDOT P.-Y. (2007), « Épitaphes oubliées. Les registres de condoléances à la mort d'un Président de la République », *Mots*, n° 84, p. 71-84.
- BOLTANSKI L. et GODET M.-N. (1995), « Messages d'amour sur le "Téléphone du Dimanche" », *Politix* n° 31, p. 30-76.

- BOLTANSKI L. et THÉVENOT L. (1991), *De la Justification, les économies de la grandeur*, Paris, Gallimard.
- COLLINS R. (2004), « Rituals of Solidarity and Security in the Wake of Terrorist Attack », *Sociological Theory*, vol. 22, n° 1, p. 53-87.
- CONTAMIN G. (2005), « La perspective du cadrage au révélateur du pétitionnement: pour une conception pluraliste des cadres d'interprétation », Communication au 8^e congrès de l'AFSP, Lyon: http://sites.univ-lyon2.fr/congres-afsp/imprimersans.php3?id_article=130 (consulté le 24 juin 2010).
- DURKHEIM É. ([1912] 1979), *Les formes élémentaires de la vie religieuse*, Paris, PUF.
- FRAENKEL B. (2002), *Les écrits de Septembre: New York 2001*, Paris, Textuel.
- (2007), « Actes d'écriture: quand écrire c'est faire », *Langage & Société* n° 121-122, p. 101-112.
- FRAENKEL B. et PONTILLE D. (2006), « La signature au temps de l'électronique », *Politix* n° 74, p. 103-121.
- HAZAREESINGH S. (2010), « Mort et Transfiguration: la renaissance du mythe gaullien en novembre 1970 », *Parlement(s)* n° 13, p. 24-36.
- JONES B. (1999), « Books of Condolence », in Walter T. (ed.), *The Mourning for Diana*. Oxford & New York, Berg (p. 203-214).
- LAMIREAU C. (2009), *La poursuite du divin. Écritures votives des lieux de culte catholiques parisiens*. Thèse de doctorat, EHESS, Paris.
- MAUSS M. ([1921] 1969), « L'expression obligatoire des sentiments » dans *Essais de sociologie*, Paris, Seuil, p. 81-88.
- REINERT M. (1990), « ALCESTE, une méthodologie d'analyse des données textuelles et une application: Aurélia de Gérard de Nerval », *Bulletin de Méthodologie Sociologique* n° 26, p. 24-54.
- (2000), « La tresse du sens et la méthode Alceste. Application aux *Rêveries du promeneur solitaire* », Communication aux 5^e Journées Internationales d'Analyse Statistique des Données Textuelles.
- (2003), « Le rôle de la répétition dans la représentation du sens et son approche statistique dans la méthode Alceste », *Semiotica* n° 147, p 389-420.

- (2006), « La Méthode informatisée d'analyse de discours "Alceste". Application aux Rêveries du promeneur solitaire », *Encyclopédie de l'Astrolabe*, Ottawa : www.uottawa.ca/academic/arts/astrolabe/articles/art0049/Alceste.htm (consulté le 24 juin 2010).
- (2007), « Postures énonciatives et mondes lexicaux stabilisés en analyse statistique de discours », *Langage & Société* n° 121-122, p. 189-202.
- SÁNCHEZ CARRETERO C. (2006), « Trains of Workers, Trains of Death: Some Reflections after the March 11 Attacks in Madrid », in Santino J. (ed.) *Spontaneous Shrines and the Public Memorialization of Death*, New York, Palgrave (p. 333-347).
- (ed.) (2010), *El Archivo del Duelo. Análisis de la respuesta ciudadana ante los atentados del 11 de marzo en Madrid*. Madrid : CSIC.
- TRUC G. (2006), « Le cosmopolitisme sous le coup de l'émotion. Une lecture sociologique des messages de solidarité en réaction aux attentats du 11 mars 2004 à Madrid », *Hermès* n° 46, p. 189-199. Seconde édition, revue et actualisée dans Rasse P. (dir.) (2010) *La mondialisation de la communication*, Paris, CNRS Éditions (p. 113-138).
- (2010), « Le cosmopolitisme européen à l'épreuve du terrorisme. Une analyse des réactions européennes aux attentats du 11 mars 2004 à Madrid et du 7 juillet 2005 à Londres », dans Rouyer M. et Bousquet E. (dir.) *Regards sur le cosmopolitisme européen*, Bruxelles, Peter Lang (à paraître).
- (2011), « Espacio de Palabras y rituales de solidaridad en Atocha », in Sánchez Carretero C. (ed.) *El Archivo del Duelo. Análisis de la respuesta ciudadana ante los atentados del 11 de marzo en Madrid*, Madrid, CSIC (à paraître).

