



HAL
open science

Un outil d'observation du cheminement linguistique des événements médiatiques

Delphine Battistelli, Charles Teissède

► **To cite this version:**

Delphine Battistelli, Charles Teissède. Un outil d'observation du cheminement linguistique des événements médiatiques . Les cahiers de praxématique, 2014, L'événement : du syntagme au discours, 63. halshs-01548444

HAL Id: halshs-01548444

<https://shs.hal.science/halshs-01548444v1>

Submitted on 27 Jun 2017

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

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

Un outil d'observation du cheminement linguistique des événements médiatiques

A tool to examine the linguistic path of media events

Delphine Battistelli et Charles Teissède



Édition électronique

URL : <http://praxematique.revues.org/2397>
ISSN : 2111-5044

Éditeur

Presses universitaires de la Méditerranée

Édition imprimée

Date de publication : 28 décembre 2014
ISSN : 0765-4944

Ce document vous est offert par SCD
Université de Paris Ouest Nanterre la
Défense



Référence électronique

Delphine Battistelli et Charles Teissède, « Un outil d'observation du cheminement linguistique des événements médiatiques », *Cahiers de praxématique* [En ligne], 63 | 2014, document 4, mis en ligne le 10 avril 2015, consulté le 27 juin 2017. URL : <http://praxematique.revues.org/2397>

Ce document a été généré automatiquement le 27 juin 2017.

Tous droits réservés

Un outil d'observation du cheminement linguistique des événements médiatiques

A tool to examine the linguistic path of media events

Delphine Battistelli et Charles Teissède

1. Introduction

- L'étude que nous présentons s'intéresse au cheminement linguistique de désignations d'événements dans un corpus volumineux de dépêches journalistiques. Nous privilégions un point de vue qui est celui des sciences du langage en nous intéressant aux formes linguistiques (marqueurs lexicaux ou grammaticaux, types d'intégrations syntagmatiques) et proposons une approche « outillée » pour l'observation de l'évolution de ces formes au cours du temps. Cette étude prend place dans une réflexion plus générale sur la notion de saillance d'un événement et les critères quantitatifs comme qualitatifs qui permettent de l'appréhender en corpus (dans une autre terminologie, on parlera parfois d'évènement discursif pour désigner un événement saillant en corpus). Après avoir présenté le cadre méthodologique et expérimental qui est le nôtre (section 2), nous présentons la stratégie générale d'automatisation que nous avons mise en place pour faciliter la constitution et l'analyse de corpus (section 3) et illustrons l'utilisation de l'outil que nous avons développé à partir de deux études de cas issus de notre corpus de dépêches journalistiques (section 4). Nous terminons par un bilan et sur les perspectives ouvertes par nos travaux (section 5).

2. Cadre d'analyse et de développement expérimental

2.1 La notion de saillance : d'une perspective appliquée à une perspective théorique

- 2 La problématique d'observation des désignations d'événements dans de larges corpus que nos travaux abordent est en partie née dans le cadre du projet ANR ChronoLines.¹ Ce projet a pour finalité applicative de développer des composants logiciels interactifs pour faciliter le travail des journalistes à l'AFP amenés à produire et diffuser très régulièrement des objets textuels appelés « chronologies événementielles » (CE). Une CE rassemble les événements majeurs (saillants) associés à une thématique donnée qui fait l'objet d'une médiatisation au moment de sa diffusion (par exemple l'affaire du Mediator, Fukushima, la vache folle, la chute de Moubarak, etc.). Une des problématiques soulevées dans ce cadre concerne la manière dont on peut, à partir d'un ensemble de dépêches AFP, fournir des critères pour mesurer la saillance de différents faits relatifs à une thématique, afin de sélectionner et extraire les événements qui figureront sur la CE générée, qui sera ensuite validée ou non par le journaliste de l'AFP. Notre corpus de départ regroupe l'ensemble des dépêches de l'AFP en français des années 2011 à 2013, soit un peu plus d'un million de dépêches. Ce corpus est par nature redondant. Les descriptions des faits/événements sont en effet régulièrement mises à jour à mesure de leur progression ou du complément d'information qu'en ont les journalistes. À l'aide de méthodes d'analyse automatique, il devient alors possible de faire émerger dans ce type de corpus des événements particulièrement « saillants », c'est-à-dire des événements qui, mentionnés à de nombreuses reprises sous des formes plus ou moins proches, sont susceptibles de correspondre à des « moments discursifs », puis à des « événements discursifs » (Moirand, 2006 et 2007).
- 3 C'est dans ce cadre que nous avons été amenés à préciser les critères qualitatifs d'ordre linguistique participant de l'appréciation de la saillance de certains faits décrits dans les textes (devenant alors « événements discursifs »). La méthodologie de construction automatique d'une CE que nous avons mise en place est décrite dans Battistelli et al. (2013). Un événement « saillant » est associé à une phrase extraite automatiquement du corpus. Partant d'un ensemble de phrases ainsi extraites, nous proposons, dans une deuxième étape -- qui est celle que nous présentons plus particulièrement ici -- de nous intéresser à la manière dont, linguistiquement, un même événement est désigné et comment cette désignation évolue au fil du temps. Nous montrons ainsi que ces désignations peuvent aboutir alors en particulier à des nominalisations au sein de structures adverbiales temporelles (du type *depuis la chute de Hosni Moubarak; cinq mois après la démission de Moubarak; peu avant la démission de Moubarak le 11 février 2011 sous la pression de la rue*). Ce processus aboutit aussi parfois à la constitution de véritables chrononymes (comme *l'après Moubarak*) (Bacot et al., 2008). Les composants logiciels que nous développons visent à faciliter l'observation de l'avènement de ces désignations.
- 4 Notre approche se situe à la croisée de travaux sur les événements en analyse du discours du type de ceux décrits dans Veniard (2009) ou Krieg-Planque (2009) et de travaux en traitement automatique des langues tels que ceux décrits dans Chen & Ji (2009) ou encore Bejan et Harabagiu (2010) qui décrivent des systèmes pour retrouver automatiquement

dans un texte ou dans un corpus les différentes désignations (dites « coréférentielles ») d'un même événement.

- 5 Le logiciel que nous avons développé, *Chronoliner*, permet de faciliter l'analyse du cheminement linguistique des événements médiatiques. En effet, dans les traitements qu'il effectue pour produire une CE, il regroupe des phrases dans lesquelles il est fait référence à chacun des événements saillants extraits automatiquement pour une thématique donnée. A partir des corpus ainsi obtenus, il devient possible d'observer la distribution dans le temps des mentions d'un même événement (en parle-t-on longtemps ? est-ce qu'un pic dans la distribution signale un nouvel événement ? etc.). Par ailleurs, nous cherchons à voir si le type d'intégration syntagmatique d'un événement participe de l'appréciation de la « saillance » de celui-ci. Une question que nous posons en particulier est ainsi la suivante : un événement « saillant » est-il toujours – ou du moins souvent – nominalisé et désigné à un moment ou à un autre au sein d'une structure adverbiale ?
- 6 Cette approche fait écho à un point de vue sociologique. L'étude de l'évolution des désignations événementielles revient à observer la dynamique par laquelle un événement nouveau est progressivement appréhendé en langue comme en discours, dans un mouvement qui consiste à déterminer cet événement, en particulier à travers le processus de sa nomination ou désignation qui a fait l'objet de nombreuses études (voir Calabrese Steimberg (2006) et 2009, Veniard (2009), Samouth (2013)), ou encore, comme très souvent dans la presse, en le mettant en regard avec d'autres événements plus anciens, permettant de l'éclairer et avec lesquels il entretient des formes d'affinité ou d'analogie (apparition dans une série, rapport de cause à effet, nature similaire, etc.). Une fois échu et ressaisi dans et à travers le discours, cet événement peut alors lui aussi par la suite entrer « en interaction avec les événements bruts et non traités » (Dewey, 1925, cité par Quéré, 2013) et fonctionner comme « agent de l'histoire qui se fait » (Quéré, 2013). C'est alors qu'on le trouve fréquemment désigné en position adverbiale – comme nous le verrons dans l'étude de cas A présentée en section 4 –, servant à la fois de repère temporel et d'élément permettant de faire émerger des séries ou des cycles, contribuant à former une histoire. C'est du reste un mouvement de même nature qui conduit l'AFP à diffuser régulièrement des chronologies dont l'objectif est d'éclairer un événement nouveau à la lumière d'événements plus anciens en mesure de lui donner sens. À chaque événement nouveau d'importance, les journalistes reconsidèrent le passé que cet événement (re)construit et dans lequel il s'inscrit : « un présent différent fait surgir un passé différent » – ce qu'illustre l'étude de cas B présentée en section 4 où un événement passé n'est mentionné dans le corpus que sept jours après s'être produit ; l'événement « crée un passé parce que se pose la question de ce qui l'a provoqué et conditionné ; il crée un futur parce qu'on s'intéresse à son issue et à ses conséquences, c'est-à-dire à sa signification » (Quéré, 2013).

2.2 Approche en discours et en langue de la saillance d'un événement

- 7 Nous entendons par cheminement linguistique le mouvement qui s'opère dans le temps en termes de variations d'intégrations syntagmatiques d'un même événement au sein d'une phrase. Il s'agit donc d'une approche que l'on peut dire « en langue » de la notion de saillance. Un type de syntagme nous intéresse en particulier, à savoir le syntagme

adverbial de localisation temporelle. Notre hypothèse de travail consiste en effet à poser que la saillance d'un événement est appréhendée en langue par des procédés d'ordre morpho-syntaxique, l'adverbialisation temporelle constituant une forme particulièrement intéressante à considérer en tant qu'elle signale qu'un événement en vient à fonctionner comme un repère temporel permettant de mettre en lumière d'autres événements.

- 8 L'intégration syntagmatique d'un événement évolue dans le temps et semble parfois suivre une trajectoire dont le schéma prototypique serait le suivant : d'abord désigné à l'aide d'un syntagme verbal, un événement « saillant » a tendance à être nominalisé puis par la suite désigné à un moment ou à un autre au sein d'une structure adverbiale, servant alors de repère pour éclairer d'autres événements. La question que nous posons pourrait se résumer ainsi : le type d'intégration syntagmatique observée d'un événement constitue-t-il un critère, qualitatif, de la plus ou moins grande « saillance » de celui-ci ? Corrobore-t-il, voire renforce-t-il, la saillance d'un événement telle qu'elle est mesurée par des critères quantitatifs ?
- 9 Dans la perspective développée en analyse du discours, la saillance d'un fait évaluée en corpus renvoie aux notions de « moment discursif » et « d'événement discursif » (Moirand, 2006 et 2007). Un moment discursif correspond non seulement à une abondante production médiatique, mais aussi au fait qu'il en reste également quelques traces à plus ou moins long terme dans les discours produits ultérieurement à propos d'autres événements, formant des « trajets mémoriels » (Moirand, 2001). Un « événement discursif » se traduit souvent dans ce cadre par des actes de nominalisation qui peuvent revêtir diverses formes, comme les chrononymes (la Belle Époque, l'Ancien Régime) ou encore les héméronymes — ces dates qui portent à elles seules la sémantique référentielle d'événements (le 11 septembre par exemple).
- 10 L'approche en discours revêt donc une dimension quantitative (nombre de fois où un événement est mentionné et durée pendant laquelle on en parle), mais aussi qualitative puisque l'on s'intéresse tout particulièrement à l'acte de nominalisation, qui sur le plan syntaxique, produit souvent des formes du type : [article + nom + complément du nom] ou [article + nom + adjectif]. On rejoint alors une analyse « en langue » de la notion d'événement. Si le processus de nominalisation des événements a fait l'objet de nombreux travaux, en revanche le processus d'adverbialisation, qui fait intervenir ou non ces nominalisations, a beaucoup moins été étudié. Nous avons montré dans de précédents travaux (Teissèdre (2012), suite à Battistelli *et al.* (2008)), que les adverbiaux dont le noyau est de type événementiel ou calendaire (*i.e.* faisant mention explicite d'un nom calendaire) constitue une classe sémantiquement homogène : la classe des adverbiaux de localisation temporelle (ALT). La notion d'adverbe temporel par opposition présente une hétérogénéité maintes fois soulignée, aussi bien sur le plan morphologique que syntaxique ou sémantique. Les ALT semblent eux pouvoir être décrits d'une façon uniforme sous la forme d'une succession d'opérateurs sémantiques : opérations de régionalisation (*depuis, avant, lors de*), opérations de focalisation (*début, fin, milieu*) et de déplacement (*trois jours avant*). Cette succession d'opérations enchâssées a pour opérande une base qui peut être soit calendaire (*depuis la fin du mois de mars*), soit événementielle (*peu avant les élections*). On fait l'hypothèse ici que le processus d'adverbialisation d'un événement renforce sa saillance, car il revient à le considérer comme quelque chose à partir de quoi on peut se repérer dans le temps.

- 11 Nous présentons dans la section suivante la stratégie d'automatisation mise en place pour constituer des corpus d'observation des différentes formes d'intégration syntagmatique d'un événement. Le verrou qu'il s'agit de lever dans ce cadre est le suivant : comment sait-on que l'on parle d'un même événement à partir du moment où, précisément, les formes varient ? Notre proposition repose sur une stratégie d'automatisation utilisant des critères quantitatifs (mesures de similarité lexicale).

3. Saillance des événements : stratégie d'analyse automatisée mise en place

3.1 Architecture générale

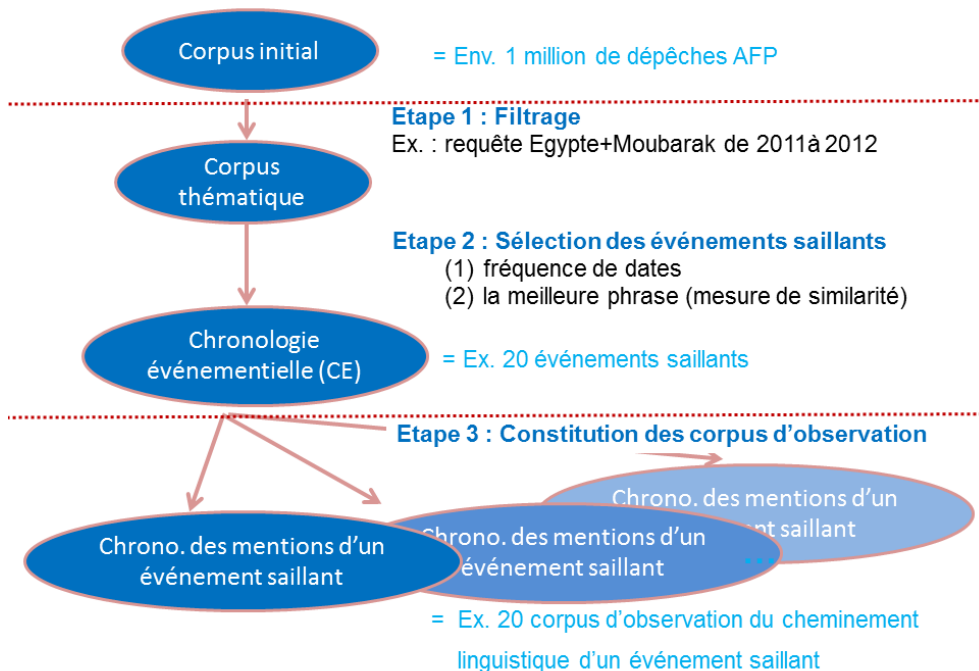


Figure 1. Architecture générale du système Chronoliner

- 12 La figure 1 décrit l'architecture générale du logiciel que nous avons développé, *Chronoliner*. À partir du corpus initial, une étape de filtrage à l'aide mots-clés et d'intervalles de dates (étape 1) permet de constituer un corpus thématique sur une période spécifiée par l'utilisateur (par exemple, Egypte+Moubarak de 2011 à 2012). Deux étapes principales sont ensuite à distinguer. L'étape 2 permet de constituer automatiquement une frise chronologique des événements saillants - autrement appelée CE - associée à la thématique. Un événement saillant est présenté sur une CE sous la forme d'une phrase issue d'une dépêche de l'AFP (cf. Fig 1). Le nombre d'événements figurant sur une CE est déterminé par l'utilisateur. Il peut choisir de visualiser les 10 ou 20 événements les plus saillants par exemple. La figure 2 montre un extrait de la CE associée à l'affaire Leonarda produite par notre système tandis que la figure 3 montre l'interface visuelle associée.

2013-10-09 :

"Ils réclament le retour de la collégienne rom LeonardaDibrani, 15 ans, expulsée le 9 octobre du Doubs vers le Kosovo après avoir été remise à la police lors d'une

sortie scolaire, et celui de KhatchikKachatryan, 19 ans, lycéen à Paris expulsé le 12 octobre vers l'Arménie.”

Titre de la dépêche du 5 novembre 2013 : Affaire Leonarda: la mobilisation des lycéens s'essouffle

2013-10-14 :

“C'est un message lundi sur un blog du Réseau Education sans Frontières (RESF), hébergé par Mediapart, qui a lancé la médiatisation autour du cas de la collégienne Leonarda, dont l'expulsion vers le Kosovo après plus de quatre ans en France a déclenché une tempête politique.”

Titre de la dépêche du 17 octobre 2013 : RESF à la baguette pour médiatiser l'expulsion de Leonarda

2013-10-17 :

“Plusieurs milliers de lycéens parisiens ont manifesté jeudi pour réclamer le retour en France des élèves étrangers expulsés, après le renvoi au Kosovo de la collégienne LeonardaDibrani et à la veille de la diffusion des conclusions d'une enquête sur les circonstances de son interpellation.”

Titre de la dépêche du 17 octobre 2013 : L'affaire Leonarda fait descendre les lycéens dans la rue

Figure 2 : Extrait de la CE générée par Chronoliner sur l'affaire Leonarda



Figure 3 : Interface montrant un événement saillant de la CE sur l'affaire Leonarda

- 13 La méthodologie complète associée à cette première étape a été décrite dans Battistelli et al. (2013). Notons simplement que le processus de sélection des événements saillants tire parti d'une caractéristique forte des corpus de l'AFP : la redondance des informations. La plupart des dépêches sur un sujet donné ne diffèrent des précédentes que par une mise à jour de quelques informations (le décompte de manifestants, par exemple). Notre système s'appuie entre autres sur des mesures de similarité entre phrases (mesure de Jaccard) qui permettent ensuite de filtrer et d'ordonner les différentes phrases entre elles et d'extraire celle qui présente le plus de similarité avec l'ensemble des autres phrases d'un même ensemble. La phrase ainsi sélectionnée apparaît dans la CE associée à une thématique donnée à l'étape 2 : elle est considérée par le système comme celle décrivant le mieux un événement saillant. Les autres phrases sont conservées pour former un corpus d'observation des différentes mentions de ce même événement à l'étape 3.
- 14 Lors de l'étape 2, le logiciel permet en effet de s'intéresser aux différentes phrases dans lesquelles il est fait référence à chacun des événements d'une CE. Autrement dit, il permet

de réunir un corpus qui traite du même objet discursif circonscrit dans le temps et d'observer à travers ce corpus de phrases le cheminement linguistique d'un événement au fil du temps. Nous allons détailler dans la section 4 cette étape à partir de deux études de cas.

- 15 Au final, *Chronoliner* permet donc d'intégrer des critères quantitatifs (nombre de fois où un événement est mentionné) et qualitatifs (types d'intégration syntagmatique d'un événement) dans l'étude de la saillance d'un événement. Rappelons que l'hypothèse initiale que nous cherchons à vérifier peut s'énoncer de la manière suivante : l'appréciation du type d'intégration syntagmatique d'un événement participe-t-elle de l'appréciation de la « saillance » de celui-ci ?

4. Deux études de cas

- 16 Afin de discuter et d'évaluer la pertinence de ces critères et donc de notre hypothèse initiale, nous avons utilisé le système *Chronoliner* pour observer les corpus associés aux événements saillants de plusieurs CE. Nous choisissons de montrer ici les résultats d'analyse de deux corpus d'observation associés à deux événements médiatiques différents. L'étude de cas A concerne un des événements saillants (la chute de Hosni Moubarak) associés à la CE sur la thématique de l'Égypte entre début 2011 et fin 2012. L'étude de cas B concerne un des événements saillants (l'expulsion de Leonarda) associés à la CE relative à l'affaire Leonarda survenue fin 2013. Nous allons voir que le cas A illustre le fait que la saillance quantitative d'un événement se retrouve d'un point de vue qualitatif puisque de nombreux exemples de nominalisations et d'adverbialisations référant à cet événement sont présents dans le corpus d'observation. *A contrario*, le cas B illustre le fait qu'une saillance d'ordre quantitative n'est pas nécessairement corroborée ensuite par une intégration syntagmatique de type adverbialisation temporelle, et qu'un autre phénomène notable peut cependant être observé : le fait de recourir à des nominalisations du type « crise/affaire/ ... de... » qui ne désignent plus un événement en particulier, mais plutôt un ensemble d'événements et cristallisent souvent la présence de controverses dans les opinions.

4.1 Étude de cas A : La chute de Moubarak

- 17 Le corpus à partir duquel est produite la CE sur l'Égypte depuis les manifestations ayant entraîné la chute de Hosni Moubarak début 2011 couvre la période allant de janvier 2011 à décembre 2012. Il contient près de 3500 dépêches de l'AFP. La CE produite par *Chronoliner* à l'étape 1 contient les vingt événements considérés comme étant les plus saillants.
- 18 L'analyse des différents corpus produits à l'étape 2 et associés à ces événements montre que la répartition dans le temps des différentes mentions d'un événement saillant peut être très variable : certains événements sont annoncés avant leur réalisation effective (la date du verdict du procès de Moubarak était ainsi fixée plus de six mois à l'avance), d'autres continuent d'être évoqués après coup, sur une période de temps plus ou moins étendue. Sans surprise, les événements qui font date (en particulier les événements « historiques » comme la chute de Hosni Moubarak ou l'élection du président Mohammed Morsi) sont ceux qui sont mentionnés le plus fréquemment, sur de vastes plages de temps ; ce sont aussi les désignations de ces événements qui évoluent le plus au fil du temps. Le

corpus des références à la chute de Moubarak contient 554 phrases. C'est ce corpus que nous avons choisi d'observer plus particulièrement.

- 19 Deux modes de désignations majeurs semblent émerger de ce corpus : un mode de désignation à travers un syntagme nominal (env. 60 %, cf. exemple 1²) et un mode de désignation à travers un syntagme verbal (env. 20 %, cf. exemples 2 et 3), qui l'un comme l'autre sont susceptibles de s'intégrer dans des structures adverbiales (env. 48 %, cf. exemples 4 et 5), ce qui signifie que l'événement désigné occupe alors une fonction particulière : il sert de référence pour éclairer et mettre en perspective de nouveaux événements.

(1) La chute du président Hosni Moubarak provoque un sentiment de fierté et de dignité retrouvée chez les Egyptiens, qui évoquent l'époque où le charismatique Gamal Abdel Nasser avait fait du pays un phare du monde arabe. (dépêche datant du 13 février 2011)

(2) Hosni Moubarak a été chassé du pouvoir le 11 février, après trois décennies à la tête du pays, par une révolte populaire au cours de laquelle 846 personnes ont été tuées, selon un bilan officiel. (dépêche datant du 2011-08-20)

(3) L'ancien président égyptien Hosni Moubarak, qui a démissionné le 11 février sous la pression de la rue, doit présenter ses excuses à la nation et plaider en faveur d'une amnistie, affirme mardi le journal indépendant al-Chorouk. (dépêche datant du 17 mai 2011)

(4) Les Egyptiens sont appelés à descendre en masse dans la rue vendredi, cinq mois après la chute de l'ancien président Hosni Moubarak le 11 février, sur fond de mécontentement croissant face à la manière dont l'armée, qui dirige le pays, gère la transition promise vers la démocratie. (dépêche datant du 7 juillet 2011)

(5) Depuis que le président égyptien Hosni Moubarak a cédé le pouvoir sous la pression de la rue le 11 février, les nouvelles autorités du pays ont desserré l'étouffement sur le passage de Rafah, qui n'était auparavant ouvert que de manière exceptionnelle, quelques jours par mois, pour des raisons humanitaires. (dépêche datant du 29 avril 2011)

- 20 Dans une proportion moindre, de l'ordre de 10 %, on trouve des constructions adjectivales, parfois en apposition (cf. ex. 6) :

(6) Chassé du pouvoir par la rue le 11 février, Hosni Moubarak, 82 ans, a été hospitalisé mardi après un interrogatoire dans le cadre d'une enquête sur la répression des manifestations de janvier et février, qui a fait quelque 800 morts selon des sources officielles. (dépêche datant du 2011-04-17)

- 21 La désignation de cet événement (que l'on peut qualifier d'événement historique) observe une trajectoire particulière. Dans un premier temps les structures verbales telles que celle figurant dans l'exemple (1) semblent prédominer. Ces structures représentent la moitié des références sur le premier mois qui suit la chute de Hosni Moubarak, alors qu'elles ne représentent plus que 16 % des références sur les six mois suivants. A l'inverse les intégrations syntagmatiques de type nominales, se rencontrant notamment au sein d'adverbiaux de localisation temporelle comme dans les exemples (2) et (3), se font de plus en plus nombreuses avec le temps. Les références nominales représentent 30 % des références le premier mois, puis 71 % les six mois suivants. Les références adverbiales représentent 15 % le premier mois, puis 60 % ensuite.

(1) Le président égyptien Hosni Moubarak, 82 ans, a démissionné vendredi, chassé par la rue, et a remis les pouvoirs à l'armée, provoquant une explosion de joie dans le pays sur lequel il régnait depuis trois décennies. (dépêche du 11 février 2011)

(2) Les Egyptiens sont appelés à descendre en masse dans la rue vendredi, cinq mois après la chute de l'ancien président Hosni Moubarak le 11 février, sur fond de

mécontentement croissant face à la manière dont l'armée, qui dirige le pays, gère la transition promise vers la démocratie. (dépêche du 7 juillet 2011)

(3) Le drame a relancé la contestation contre l'armée, à la tête du pays depuis la démission de Hosni Moubarak le 11 février 2011 sous la pression de la rue. (dépêche du 13 février 2012)

- 22 Il est intéressant de constater que le cheminement linguistique d'un autre événement saillant sélectionné dans la même CE n'est pas du tout du même ordre. C'est le cas de l'ouverture du procès de Hosni Moubarak, annoncée pour le 3 août 2011. Dans le corpus d'observation de cet événement, la proportion des intégrations syntagmatiques de type verbal est importante (de l'ordre de 66 % de références sur un total de 72 mentions) et constante au fil du temps, alors que la nominalisation de l'événement reste faible (23 %). L'événement cesse même complètement d'être mentionné dès le lendemain de la publication du verdict, le 2 juin 2012, où Hosni Moubarak est condamné à la prison à vie. C'est que désormais l'ouverture de son procès importe moins que les conséquences politiques de sa conclusion.

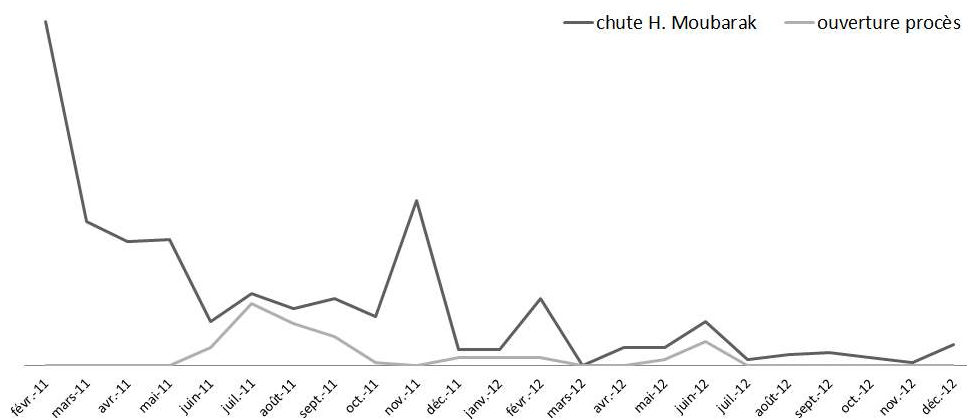


Figure 4 : répartition dans le temps des références à deux événements

- 23 Dans le tableau de répartition dans le temps des références à ces deux événements, la chute de Moubarak et l'ouverture de son procès (Fig. 4), on observe des pics soudains dans une distribution globalement descendante. Ces pics signalent de nouveaux événements saillants que les anciens événements viennent mettre en perspective : en novembre 2011, par exemple, le pic correspond aux premières élections législatives et l'on rencontre par exemple le syntagme adverbial « depuis la chute de Hosni Moubarak » ; février 2012 est la date anniversaire de la chute de Moubarak ; en juin 2012, l'élection présidentielle est remportée par Mohamed Morsi et le verdict du procès de Moubarak est annoncé ; en décembre 2012, la nouvelle Constitution est adoptée et des troubles avec l'opposition fragilisent le pouvoir en place.

4.2 Etude de cas B : L'affaire Leonarda

- 24 Le corpus constitué automatiquement sur le thème de l'affaire Leonarda regroupe 132 dépêches publiées entre début septembre et fin décembre 2013. Ce corpus est beaucoup plus petit que le précédent et étalé sur une période beaucoup plus courte. La saillance, telle qu'elle se mesure sur un plan quantitatif (la fréquence relative du nombre de références à un événement par rapport à un ensemble), est pour ce type de corpus

beaucoup moins facile à déterminer, le volume des données étant faible : la redondance est donc plus difficile à exploiter.

- 25 Un des événements saillants calculés par le système à l'étape 1 est l'expulsion de Leonarda le 9 octobre 2013 (après son interpellation lors d'une sortie scolaire). Le corpus d'observation des références à cet événement, obtenu à l'étape 2, regroupe 26 phrases. On constate que cet événement, bien qu'ayant eu lieu le 9 octobre, n'est mentionné pour la première fois dans le corpus que le 16 octobre. Il est mentionné une dernière fois le 18 novembre. Il s'agit donc d'un corpus très circonscrit dans le temps, qui correspond à un vrai moment discursif dans la presse. Du point de vue du critère quantitatif des dates les plus fréquemment mentionnées dans le corpus, cet événement pourtant central n'occupe que la quatrième place (derrière les dates du 17, 18 et 19 octobre).
- 26 Les constructions adjectivales (11 au total) sont les plus fréquentes (exemple 1). On dénombre sept constructions verbales (exemple 2), cinq nominalisations (exemples 3 et 4) et trois adverbiaux (exemples 3 et 4) :
- (1) Manuel Valls est vivement critiqué, notamment à gauche, pour les conditions d'interpellation de Leonarda, expulsée le 9 octobre vers le Kosovo avec sa famille après avoir été prise en charge par la police en pleine sortie scolaire. (19 octobre 2013)
 - (2) LeonardaDibrani, collégienne du Doubs, a été expulsée le 9 octobre vers le Kosovo. (18 octobre 2013)
 - (3) L'annonce que la famille de la collégienne rom LeonardaDibrani avait été agressée dimanche, alors qu'elle se promenait dans Mitrovica, la ville du nord du Kosovo où elle réside depuis son expulsion de France le 9 octobre, a ajouté à l'émotion suscitée par l'affaire. (20 octobre 2013)
 - (4) Le jour de l'expulsion, le 9 octobre, la police était allée chercher leur fille de 15 ans, Leonarda, à la descente d'un bus scolaire alors qu'elle participait à une sortie avec sa classe, créant une tempête politique. (28 octobre 2013)
- 27 S'il y a peu de syntagmes adverbiaux intégrant cet événement dans le corpus produit automatiquement par *Chronoliner* à l'étape 2, il n'en reste pas moins que la saillance de l'événement est plus marquée que celle d'autres événements de la CE pour lesquels ce mode d'intégration syntagmatique n'apparaît jamais, comme par exemple pour l'événement déclencheur de la médiatisation de l'affaire Leonarda (la publication d'un message le 14 octobre 2013 sur un blog du Réseau Éducation sans Frontières, hébergé par Mediapart) ou encore les manifestations lycéennes du 17 octobre 2013, événement mentionné uniquement sous la forme de syntagmes verbaux dans les onze phrases du corpus des mentions de cet événement.
- 28 Fait notable, le terme « affaire Leonarda » apparaît dans le titre d'une dépêche le 17 octobre, puis, le 20 octobre, dans le corps d'une dépêche. Ce n'est donc plus un événement particulier dont la désignation se retrouve encore longtemps après, mais la séquence d'événements désignée par « l'affaire Leonarda » (qui comprend l'interpellation et l'expulsion, la décision du gouvernement et du chef de l'état, les conséquences politiques, les réactions dans l'opinion, etc.). On assiste bien là à l'identification d'un véritable « événement discursif », au sens de (Moirand, 2006, 2007).

5. Bilan

- 29 Dans ce travail, nous cherchons à intégrer des critères d'ordre qualitatif pour appréhender la saillance des événements, en ne la fondant donc plus sur des critères

uniquement quantitatifs (comme c'est le cas, rappelons-le, dans les travaux actuels relevant du traitement automatique des langues). D'une certaine manière, nous tentons de donner corps à des analyses développées en analyse du discours qui croisent ces deux types de critères, au travers notamment des notions de « moment discursif » ou d'« événement discursif » (voir section 2.2 *supra*). Deux types de critères qualitatifs, non exclusifs, nous semblent tout particulièrement intéressants à retenir. Le premier confirme nous semble-t-il la justesse de notre hypothèse initiale, à savoir que les événements saillants sont plus souvent nominalisés et désignés à un moment ou à un autre au sein d'une structure adverbiale, servant alors de repères pour éclairer de nouveaux événements. C'est le cas dans les exemples vus en section 4.1, pour des événements que l'on peut qualifier d'historiques (chute de Hosni Moubarak, élection de Mohamed Morsi, verdict du procès de Hosni Moubarak). Le second critère qualitatif qui nous semble intéressant à retenir et qui a émergé de notre travail d'observation grâce à l'outil *Chronoliner* que nous avons développé concerne le fait qu'il y ait identification ou non au sein d'un corpus de nominalisations particulières (intégrées ou non à des structures adverbiales temporelles) du type « affaire (de) » comme dans l'exemple de la section 4.2. Ce n'est alors plus un événement mais un ensemble d'événements indissociables les uns des autres qui font saillance, formant un tout, avec un contenu référentiel finalement assez flou et qui dénote souvent l'existence de controverses au sujet de cet événement. Il est clair que l'identification dans un corpus de formes particulières de nominalisations tels que des chrononymes, des héméronymes ou des toponymes pourrait fournir un troisième critère d'ordre qualitatif.

- 30 Dans cette approche expérimentale outillée du cheminement linguistique des événements que nous adoptons, il nous semble en tout cas qu'apparaît clairement l'apport réciproque que l'ingénierie des langues (ou le traitement automatique des langues) et la linguistique peuvent entretenir : d'un côté l'analyse linguistique permet d'orienter l'évolution d'un outil d'extraction d'information (cf. section 2.1), en contrepartie elle se nourrit des corpus et des mesures quantitatives que l'ingénierie des langues permet de produire automatiquement.

BIBLIOGRAPHIE

- BACOT P., DOUZOU L. & HONORÉ J., 2008, « Chrononymes. La politisation du temps », *Mots. Les langages du politique* 87/2, 5-12.
- BATTISTELLI D., COUTO J., MINEL J.-L. & SCHWER S. R., 2008, « Représentation algébrique des expressions calendaires et vue calendaire d'un texte », *Actes de TALN'08/1*, 9-13.
- BATTISTELLI D., CHARNOIS T., MINEL J.-L. & TEISSÈDREC., 2013, « Detecting salient events in large corpora by a combination of NLP and data mining techniques », *Computación y Sistemas* 17/2, 229-237.
- BEJAN C. A. & HARABAGIU S., 2010, « Unsupervised event coreference resolution with rich linguistic features », *Proceedings of ACL 2010*, 1412-1422.

- CALABRESE STEIMBERG L., 2006, « La construction de la mémoire historico-médiatique à travers les désignations d'événements », *Travaux du Cercle belge de linguistique*, 1, 1-16.
- CALABRESE STEIMBERG, L., 2009, « Nommer un événement ou les marges du sens dans la désignation médiatique : l'exemple de la canicule », *Le sens en marge : Représentations linguistiques et observables discursifs*, 15-28.
- CHEN Z. & JI H., 2009, « Graph-based event coreference resolution », *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*, 54-57.
- DEWEY J., 1925, *Experience and Nature*. New York : Minton, Balch & Company.
- HAN B., GATES D. & LEVIN L., 2006, « From language to time: A temporal expression anchorer », *Proceedings of TIME'06*, 196-203.
- KRIEG-PLANQUE A., 2009, « À propos des noms propres d'événement », *Les carnets du Cediscor* 11, 77-90.
- MOIRAND S., 2001, « Du traitement différent de l'intertexte selon les genres convoqués dans les événements scientifiques à caractère politique », *Semen* 13, 97-117.
- MOIRAND S., 2006, « Responsabilité et énonciation dans la presse quotidienne : questionnements sur les observables et les catégories d'analyse », *Semen* 22.
- MOIRAND S., 2007, *Les discours de la presse quotidienne. Observer, analyser, comprendre*, Paris : PUF.
- QUÉRÉ L., 2013, « Les formes de l'événement », *Les facettes de l'événement : des formes aux signes, mediAzioni* 15, <http://mediazioni.sitlec.unibo.it> ISSN 1974-4382.
- SAMOUTH E., 2013, « Nommer un événement en cours : les journées d'avril 2002 au Venezuela », *Les facettes de l'événement : des formes aux signes, mediAzioni* 15, <http://mediazioni.sitlec.unibo.it>, ISSN 1974-4382.
- TEISSÈDRE C., 2012, *Analyse sémantique automatique des adverbiaux de localisation temporelle : application à la recherche d'information et à l'acquisition des connaissances*, thèse de doctorat, Université Paris-Ouest Nanterre La Défense.
- VENIARD M., 2009, « La dénomination propre à la guerre d'Afghanistan en discours : une interaction entre sens et référence », *Les Carnets du Cediscor* 11, 61-76.

NOTES

1. Le projet est décrit ici : <http://chronolines.fr/>. Les partenaires sont l'AFP, Exalead, LIMSI, MoDyCo et Xerox.
2. Les désignations d'événements sont soulignées manuellement.

RÉSUMÉS

L'étude que nous présentons s'intéresse à l'évolution au fil du temps de désignations d'événements dans un corpus de dépêches journalistiques, jusqu'à les rencontrer en particulier au sein d'adverbiaux de localisation temporelle, c'est-à-dire lorsque ces événements en viennent à fonctionner comme des repères temporels pour d'autres événements. Le logiciel que nous avons développé, *Chronoliner*, vise à faciliter ce processus d'observation du cheminement linguistique des événements médiatiques. L'étude de l'évolution des désignations événementielles revient à observer la dynamique par laquelle un événement nouveau est progressivement appréhendé en langue comme en discours, dans un mouvement qui consiste à déterminer cet événement, en particulier à travers le processus de sa nomination, ou encore, comme très souvent dans la presse, en le mettant en regard avec d'autres événements plus anciens, permettant de l'éclairer et avec lesquels il entretient des formes d'affinité ou d'analogie (apparition dans une série, rapport de cause à effet, nature similaire, etc.). Une fois échu et ressaisi dans et à travers le discours, cet événement peut alors lui aussi par la suite entrer « en interaction avec les événements bruts et non traités » (Dewey, 1925, cité par Quéré, 2013) et fonctionner comme « agent de l'histoire qui se fait » (Quéré, 2013). C'est alors qu'on le trouve fréquemment désigné en position adverbiale, servant à la fois de repère temporel et d'élément permettant de faire émerger des séries ou des cycles, contribuant à former une histoire.

The research study we present here focuses on the evolution over time of event descriptions in newswire corpora, in particular when those descriptions appear within temporal locating adverbials, i.e. when the events described become temporal markers for other events. The software we have developed, *Chronoliner*, aims to facilitate the observation of the linguistic evolution of media events. The study of how the designation of events evolves consists in observing the process by which a new event is progressively apprehended both in language and discourse, through its nominalization for instance, or, as often occurs in the media to put an event into perspective, by relating it to former events with which it has some affinity or analogy (appearance in a series, cause and effect relationship, similar nature, etc.). Once it has occurred and has been captured in discourse, this event can then enter "in interaction with untreated raw events" (Dewey, 1925, quoted by Quéré, 2013) and work as an "agent of ongoing history" (Quéré, 2013). It then frequently appears in an adverbial position, working both as a temporal reference and element in a series or a cycle, contributing to the production of history.

INDEX

Mots-clés : événement médiatique, génération automatique de chronologies, processus de désignation des événements

Keywords : automatic timeline generation, event designation process, media event

AUTEURS

DELPHINE BATTISTELLI

MoDyCo, UMR 7114 CNRS – Université Paris Ouest Nanterre

CHARLES TEISSÈDRE

STIH, EA 4509 Université Paris Sorbonne