

La rhétorique des dépêches de presse à travers les marques énonciatives du temps, du lieu et de la personne

Nadine Lucas

GREYC, CNRS, Université de Caen (France)
Nadine.Lucas@info.unicaen.fr

Les dépêches de presse sont devenues accessibles sur Internet et permettent de constituer des corpus électroniques facilitant l'étude du discours. Ce type de documents structurés présente un formatage typo-dispositionnel reflétant les règles du genre journalistique, à ce titre il se prête particulièrement bien à une approche multilingue, les lois du genre s'appliquant prioritairement. Par exemple, la date d'émission est toujours mentionnée et toujours différemment des dates relatives aux événements rapportés. Le corpus présenté est limité à quelques langues (anglais, suédois, français, japonais). Les sources sont principalement Reuters, AFP et AP qui présentent de fortes similitudes dans la manière de présenter l'information. Une évolution récente des dépêches liée au médium électronique est l'apparition de textes longs, qui enrichissent le corpus. Ces textes de plusieurs pages peuvent être comparés aux dépêches brèves de quelques lignes ou de quelques paragraphes. La longueur est distinguée de la complexité, laquelle est évaluée en fonction du modèle de discours utilisé, celui de l'énonciation selon Jakobson.

L'objectif du travail présenté consiste à dégager rapidement des strates informatives dans le corps de texte. L'étiquetage permet d'identifier le contenu informationnel (pour l'indexation), les relais d'informants (pour la constitution de bases de citations) et *in fine* le point de vue de l'auteur ou tonalité (pour la caractérisation de l'attitude des médias sur un thème particulier) ; ces applications suscitent une forte demande de la part des professionnels de l'information.

La segmentation automatique des dépêches s'appuie sur une mise en forme matérielle régulière, elle dégage le titre, le corps de texte et la clôture (mentionnant le copyright). Le corps de texte est segmenté en en-tête (mentionnant les paramètres de l'émission, le lieu la date et le nom de l'agence) et en texte proprement dit, puis celui-ci en paragraphes et en phrases.

Nous avons choisi de mettre à l'épreuve une théorie, celle de l'énonciation selon Jakobson, qui propose de partir de la situation de communication et de dégager successivement les relais d'information qui aboutissent à une situation décrivant les événements eux-mêmes. Nous nous proposons d'explicitier les principes généraux de Jakobson sur des exemples concrets, en tenant compte des trois axes de référence proposés pour ancrer les événements, le temps, le lieu et la personne. Une implémentation grossière du modèle a été réalisée sur des dépêches courtes. Nous mettons l'accent sur les dépêches complexes ou longues pour poursuivre cette exploration.

La structure d'une dépêche est vue comme une suite de situations imbriquées, avec, à un extrême, la situation de communication entre l'émetteur et le lecteur (l'émetteur pouvant être un site reprenant des dépêches d'agence, ou l'agence de presse elle-même). Cette situation est dite d'énonciation, elle est captée dans le périphrase. À l'autre extrême, la situation rapportée concerne des protagonistes, temps et lieux relatifs à l'énoncé, ces informations sont présentes dans le corps de texte. Il peut y avoir plusieurs strates intermédiaires entre ces deux extrêmes. La date d'émission des dépêches sert de point d'entrée à l'analyse et permet de situer le lieu d'émission sur l'axe du lieu et le nom de l'émetteur effectif sur l'axe de la personne. Selon la complexité du document, on rencontre ensuite une ou plusieurs situations intermédiaires

formant autant de strates informatives : par exemple, un cadre général peut être proposé, avec mention de lieu, de temps et des informants. Le rôle d'informant principal a une grande importance pour établir la hiérarchie des dépêches complexes.

L'ancrage des dépêches sur les trois axes possibles permet de distinguer des points de vue cohérents. Les dépêches en structure simple portent des indices convergents et en nombre réduit. Par exemple, une dépêche sans commentaire présentant un informant unique permet de superposer le propos de l'informant et celui du journaliste, le lieu du discours rapporté et le lieu d'émission de la dépêche. Les dépêches commentées, les dépêches expliquées proposent des ancrages spécifiés, avec de nouveaux référents sur l'axe de la personne (cas du commentaire) du temps ou du lieu (cas de l'explication). Ces procédés définissent des événements subordonnés à l'événement principal. Une nouvelle situation est définie par l'apparition d'un ensemble de mentions reliant temps lieu et protagonistes et ceci indique un nouveau point de vue. Les dépêches élaborées présentent plusieurs points de vue, généralement plusieurs informants et plusieurs événements. L'axe de référence dominant permet de reconnaître une ligne argumentative, et ainsi de tenir compte des prises de position du locuteur. On peut reconnaître les informants privilégiés relayant un point de vue particulier sur un événement en exploitant les caractérisations spatiales et/ou temporelles co-occurentes.

Dans l'optique d'un traitement informatique, s'appuyant explicitement sur les attendus du genre journalistique, quelques stratégies sont présentées. Les zones de texte à explorer sont ordonnées, en tenant compte de la mise en forme des documents. Les variables du modèle énonciatif sont spécifiées avec des contraintes plus ou moins fortes sur les mentions à rechercher (dates, noms propres). La signalisation automatique des mentions de temps, de lieu et de personne propres à l'énonciation (Te, Le, Pe) et des mentions propres à l'énoncé (Tu, Lu, Pu) a déjà été implémentée. Contrairement aux études fondées sur un classement sémantique *a priori* des entités nommées, cette caractérisation est descendante. Elle utilise à la fois la position relative des formes dans le texte et la co-occurrence de formes capitalisées ordonnées en couples ou triplets (typiquement l'en-tête du corps de texte contient par hypothèse {Te, Le, Pe} conjoints).

Une tâche plus difficile consiste à ordonner les mentions de l'énoncé pour structurer le corps de texte, spécialement dans les documents complexes, le but étant de trouver les principales strates informatives du document. Il peut en effet y avoir plusieurs informants et commentateurs relayant l'information entre l'agence de presse et les protagonistes de l'événement rapporté. La stratégie proposée repose sur la reconnaissance de couples ou triplets disjoints d'entités nommées et de dates. Les indices formels de structuration une fois relevés, l'évaluation en contexte s'appuie sur la mise en forme matérielle du texte. L'intérêt de la modélisation informatique est de permettre une recherche exploitant des formes graphiques discriminantes assez stables (chiffres et lettres, casse, formats de date, noms propres capitalisés). Un petit nombre d'indices lexicaux différentiels (par exemple, les mesures de temps et les indices de civilité, comme *M. Herr*, etc. qui accompagnent les noms propres de personnes et non les noms de lieux) sont utilisés pour étiqueter chacun des termes. Ceci permet le traitement d'un corpus multilingue avec des ressources lexicales très réduites ou acquises par apprentissage automatique.

Références bibliographiques

- ADAM J.-M., GRIZE J.-B. et ALI BOUCHA M. (2004). *Textes et discours : catégories pour l'analyse*. Dijon, Editions universitaires de Dijon, 272 p.
- ADAM J.-M. (1997). Unités rédactionnelles et genres discursifs : cadre général pour une approche de la presse écrite. *Pratiques*, (94).
- BASTIAN S. & HAMMER F. (2004). La citation journalistique : une étude contrastive. In López-Muñoz, J. M., Marnette, S. et Rosier, L. (eds), *Le discours rapporté dans tous ses états : Question de frontières*, Paris, L'Harmattan, pp. 519-529.
- BIBER D. (1989). A typology of English texts. *Linguistics* (27), pp. 3-43.
- BROWN G. & YULE G. (1983). *Discourse analysis*. Cambridge..., Cambridge University Press.
- CICUREL F. (1992). Pré-visibilité des discours journalistiques: à propos d'un événement-catastrophe. *Carnets du Cediscor*, (1), pp. 55-78.
- DAHLET V. (2003). *Ponctuation et énonciation*. Guyane, Ibis rouge.
- Van DIJK T. (1985). *Handbook of discourse analysis*. New York, Academic Press, 2 vol.
- GIGUET E., LUCAS N. & COUSIN G. (2000). Document structure identification as a means for relevant indexation. In Gelbukh, A. (ed) *Proceedings CicLing-2000 (Conference on Intelligent Text processing and Computational Linguistics)*, Mexico, Instituto polytechnico nacional, pp. 415-428.
- HABERT B., ILLOUZ G., FOLCH H., FLEURY S., HEIDEN S., LAFON P. & PRÉVOST, S. (2000) Prendre *Le Monde* en main : choix d'architecture. In *Actes RIAO 2000*, Paris.
- JAKOBSON R. (1960). Linguistics and Poetics. In Sebeok (ed), *Style in Language* New York, Wiley, pp. 350-77.
- JAKOBSON R. (1963). *Essais de linguistique générale*. Paris, Editions de Minuit.
- JAKOBSON R. (1971). Shifters, verbal categories and the Russian verb, (Harvard University, 1957) In *Selected writings II Word and Language*, The Hague, Paris, Mouton.
- JAKOBSON R. (1973). *Questions de poétique*. Paris, Seuil.
- KARLGREN J. & JÄRVINEN T. (2002). Foreground and background text in retrieval. In Karlgren J., Gambäck B. et Kanerva P. (eds) *Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, Menlo Park, Spring 2002, The American Association of Artificial Intelligence.
- KARLGREN J. (2000) Stylistic Experiments for Information Retrieval. PhD, University of Stockholm.
- LIDDY E. (1993). Development and implementation of a discourse model for newspaper texts. *Dagstuhl Seminar on Summarizing Text for Intelligent Communication*.
- LUCAS N. (2000). Le rôle de la citation dans la structuration des articles de presse. In Murakami-Giroux, S. & Séguy, C. (eds), *Actes du premier colloque d'études japonaises de l'Université Marc Bloch*, Strasbourg, Université Marc Bloch Strasbourg, pp. 215-244.
- LUCAS N. (2004). "Le temps, dans les dépêches de presse" *Le temps dans les systèmes complexes 11^e journées de Rochebrune systèmes complexes naturels et artificiels*, ENST 2004 S001. pp. 187- 197.
- NIEL A. (1979). *L'analyse structurale des textes: littérature, presse, publicité*. Paris, Jean-Pierre Delarge Editions universitaires.
- PASCUAL E. & PÉRY-WOODLEY M-P. (1997). Modèles de texte pour la définition. In *Actes des Premières Journées Scientifiques et Techniques du Réseau francophone de l'Ingénierie de la Langue de l'AUPELF-UREF*. AUPELF-UREF, pp. 137-146.
- PETÖFI J. S. (1988). *Text and Discourse Constitution*. Berlin/New York, De Gruyter.
- WAUGH L. (1995). Reported Speech in Journalistic Discourse: The Relation of Function and Text. *Text*, 15 (1), pp. 129-173.
- WEITZMAN M. (1986). Les caractéristiques numériques indépendantes de la longueur du texte. In Brunet, E. (ed.) *En hommage à Charles Muller. Méthodes quantitatives et informatiques dans l'étude des textes: computers in literary & linguistic research*, Paris Genève, Slatkine Champion.