

Construire un corpus à forte dimension socio-historique pour une analyse du discours : le cas de la presse de tranchées

Cyrielle Montrichard*

*Laboratoire ELLIADD, Université Bourgogne Franche-Comté
UFR SLHS, rue Mégevand, 25000 Besançon
cyrielle.montrichard@edu.univ-fcomte.fr
<http://elliadd.univ-fcomte.fr/fiches/montrichardcyrielle>

La Première Guerre mondiale est un conflit singulier de par sa durée et sa violence. Persuadés de s'engager dans une guerre courte, les soldats doivent s'accoutumer à un quotidien violent et meurtrier (Cazals et Loez, 2012). C'est dans ce contexte que naissent les premières feuilles de tranchées. La presse de tranchées regroupe les journaux parus entre 1914 et 1918¹ et rédigés par des combattants français sur les fronts occidentaux ou orientaux.

Notre objet d'étude porte une dimension socio-historique forte et a, par conséquent, déjà été l'objet de recherche d'historiens (Audoin-Rouzeau, 1986; Roynette, 2010) mais aussi de linguistes contemporains au conflit (Sainéan, 1973). Cependant, notre étude s'inscrit en analyse du discours et propose une approche originale de ces documents par une étude systématique et outillée de ces titres de presse constitués en corpus. L'historicité et l'hétérogénéité des journaux de tranchées représentent un défi dans la construction d'un corpus textuel cohérent.

Dès lors, cette contribution souhaite répondre à la question : quels choix avons-nous effectués pour nous assurer de la cohérence et de la pertinence du corpus d'étude destiné à une analyse du discours outillée ? Cette proposition vise à présenter la construction du corpus d'étude : des objectifs d'analyse formulés à la sélection des titres et à la normalisation des données.

Pour construire un corpus voué à une analyse de données textuelles, il est nécessaire, en amont, de penser les objectifs de l'analyse et de l'étude (Charaudeau, 2009; Mayaffre, 2002; Pincemin, 1999, 2007). Pour ce faire, il était essentiel de se confronter aux documents. Cette étape a notamment permis de mettre en évidence la diversité des titres issus de la presse de tranchées. La lecture des documents et leur description ont par ailleurs conduit à l'élaboration de pistes d'analyse portant notamment sur l'étude : en diachronie des titres (sur la période 1914-1918), des titres selon le lieu de production, des liens entre presse de tranchées et presse traditionnelle.

Après avoir élaboré des pistes d'analyse précises, il s'agit ensuite de construire un corpus d'étude permettant leur exploration. En effet, constituer un corpus offrant la possibilité d'investir les pistes d'analyse envisagées est une condition élémentaire pour l'étude. Outre cette exigence, la sélection des titres du corpus s'est aussi faite sous la tension de deux contraintes : les contraintes techniques dues au caractère ancien des documents (disponibilité, qualité de conservation) et les contraintes liées à l'"interprétabilité" des documents notamment la représentativité, la cohérence et le volume (Pincemin, 1999).

1. Le syntagme "Journal de tranchées" apparaît dans certains titres de tranchées (*L'Argonnaute*) et l'expression "presse de tranchées" est employée par la communauté scientifique (Roynette, 2010; Audoin-Rouzeau, 1986).

Construire un corpus à dimension socio-historique en analyse du discours

La seconde étape porte sur la transcription et la normalisation des données textuelles. Face à la matérialité discursive, nous constatons l'hétérogénéité de ces données, notamment dans la graphie. Il y a d'abord les variations graphiques à la norme qui sont involontaires : les fautes d'orthographe et d'accent - semblables à celles repérées par J.-C. Pellat (voir Steuckardt (2015), 67-77). Il y a ensuite les variations graphiques à la norme volontaires. Ces dernières semblent avoir deux fonctions : une fonction que l'on pourrait décrire comme ludique (fondée sur des jeux de mots) et une fonction réaliste par l'emploi d'emprunt (à l'allemand par exemple avec "Kamerad") et des effets d'oralité (par la troncation de certains termes).

Un nouveau choix nous est alors offert dans le traitement de ces formes. Nous pouvons opter pour une transcription "fidèle" qui permet d'interroger le niveau de langue des rédacteurs mais aussi d'observer les effets visés par les variations graphiques volontaires. Cependant, dans le cadre d'une étude outillée, les variations graphiques nombreuses perturberont le logiciel dans le repérage et l'étiquetage grammatical des formes et engendreront une dispersion et donc une comptabilisation erronée de celles-ci. Afin d'obtenir des résultats statistiques précis sans dénaturer la matérialité discursive originelle, nous avons fait le choix de construire au format de codage XML-TEI un corpus de données brutes et un corpus de données normalisées (Habert et al., 1997). Ces deux états de corpus offriront des parcours de lecture différents de la matérialité discursive pour une étude fine des discours des combattants improvisés journalistes.

Cette contribution a pour objectif de montrer en quoi la construction du corpus est fonction des perspectives d'étude envisagées d'une part, et de la matérialité discursive d'autre part. Nous démontrons ainsi que la construction et la normalisation du corpus sont loin d'être des étapes uniquement techniques : elles permettent un questionnement substantiel sur les documents voués à l'analyse et les perspectives de l'étude. Par ailleurs, nous montrons comment les choix effectués en amont de l'analyse ont un rôle central dans l'étude puisque ceux-ci construisent le socle sur lequel l'analyse s'établira.

Références

- Audoin-Rouzeau, S. (1986). *14-18, les combattants des tranch : avers leurs journaux*. A. Colin.
- Cazals, R. et A. Loez (2012). *14-18, vivre et mourir dans les tranch*. Tallandier.
- Charaudeau, P. (2009). Dis-moi quel est ton corpus, je te dirai quelle est ta problématique. *Corpus* 8, 37-66.
- Habert, B., A. Nazarenko, et A. Salem (1997). *Les linguistiques de corpus*. A. Colin.
- Mayaffre, D. (2002). Les corpus rexifs : entre architextualité hypertextualité *Corpus 1*.
- Pincemin, B. (1999). Construire et utiliser un corpus : le point de vue d'une sntique textuelle interprtive. *Corpus et TAL : pour une rexion modologique*.
- Pincemin, B. (2007). Introduction. *Corpus 6*.
- Roynette, O. (2010). *Les mots des tranch : l'invention d'une langue de guerre, 1914-1919*. A. Colin.
- Sain, L. (1973). *L'argot des tranch : d'après lettres des poilus et les journaux du front*. Slatkine Reprints. 1re ed. 1915.
- Steuckardt, A. (2015). *Entre village et tranch. L'iture des Poilus ordinaires*. Inclinaison.