

La subjectivité à travers les médias : étude comparée des médias participatifs et de la presse traditionnelle

Mai Ho-Dac

UE TAL, 10 oct. 2011

Plan

- 1 ARCproject
- 2 Constitution d'un corpus dédié
- 3 Méthodologie
- 4 Résultats
- 5 Conclusion

Plan

- 1 ARCproject
- 2 Constitution d'un corpus dédié
- 3 Méthodologie
- 4 Résultats
- 5 Conclusion

Cadre



projet d'Action de Recherche Concertée : « Transformation du rapport à l'information en communication multimédia ».

Objectif : étude de l'impact des nouvelles technologies (NTIC), et particulièrement de l'Internet, sur la communication informationnelle et ses structures.

Équipes : Facultés universitaires Notre-Dame de la Paix de Namur (FUNDP), Université catholique de Louvain (UCL)

6 angles d'étude <-> 6 thèses de doctorat :

- Les nouvelles structures de communication (Sociologie)
- Les pratiques de production de l'information (Ethnologie)
- La régulation des nouveaux médias (Droit)
- Usages linguistiques et presse en ligne (Linguistique)
- Les contenus informationnels (Communication)
- La réception individualisée (Psychologie)

Cadre



projet d'Action de Recherche Concertée : « Transformation du rapport à l'information en communication multimédia ».

Objectif : étude de l'impact des nouvelles technologies (NTIC), et particulièrement de l'Internet, sur la communication informationnelle et ses structures.

Équipes : Facultés universitaires Notre-Dame de la Paix de Namur (FUNDP), Université catholique de Louvain (UCL)

6 angles d'étude <-> 6 thèses de doctorat :

- Les nouvelles structures de communication (Sociologie)
- Les pratiques de production de l'information (Ethnologie)
- La régulation des nouveaux médias (Droit)
- Usages linguistiques et presse en ligne (Linguistique)
- Les contenus informationnels (Communication)
- La réception individualisée (Psychologie)

Usages linguistiques et presse en ligne (Linguistique)

- Comparer le degré de l'expression de la subjectivité et l'organisation discursive dans différents médias de masses écrits, francophones
- Catégorisation des médias en fonction du types de texte en fonction des pratiques communicationnelles et l'usage langagier

Questions de Recherche

- Impact d'Internet : lignes éditoriales différentes, structures de travail, statut professionnel des auteurs ?
- Variation dans la subjectivité : fréquence, constructions, diversité lexicale
- Variation dans l'organisation discursive : différents types de texte (Ho-Dac 2007) ?

Plan

- 1 ARCproject
- 2 Constitution d'un corpus dédié
- 3 Méthodologie
- 4 Résultats
- 5 Conclusion

Besoins

- Un large corpus de presse francophone (8 millions de mots)
- permettant de comparer différentes influences d'internet (avec/sans, etc.)
- dans le cadre d'un projet Belge (locuteurs francophones belges vs. francophones français)

Choix

3 médias sources

Le Soir : presse traditionnelle papier + version en ligne

Rue89 : site d'actualité de journalisme en réseau avec comité de rédaction "professionnel"

AgoraVox : presse citoyenne avec ligne éditoriale "démocratique"

1 période commune

2005-2009 (aligné sur la durée de vie du plus jeune média : Rue89)

Idée première : comparer avant et après l'apparition de la version en ligne (1996)

Presse traditionnelle et nouveaux médias

différences

- deadlines
- but du scoop
- professionnalisme des auteurs
- documentation 'objective' vs. expression d'opinion

parallèles

- actualisation du site tous les jours
- traitement de sujets communs (d'actualité)
- médias avec une 'instance de contrôle'
- des rubriques comparables

Deux caractéristiques contrastives interdépendantes

le degré de professionnalisme des auteurs : Le Soir fait appel des journalistes professionnels, Rue89 est un lieu de journalisme en réseau (journalistes ou spécialistes de domaine) et AgoraVox est un lieu d'expression de citoyens tout court.

la philosophie du média relayée par l'intention de l'auteur de publier via tel ou tel média : d'un côté Le Soir recherche le scoop et l'information objective (hors genre "opinion" comme les éditoriaux, les commentaires, les critiques); de l'autre, Rue89 et AgoraVox affichent une volonté de "faire de l'information différemment", sans ligne éditoriale particulière et avec une certaine liberté quant au ton, au temps et au format de publication.

Deux caractéristiques contrastives interdépendantes

Le Soir

Mettre la nouvelle en ligne le plus vite possible, à être le premier à publier l'information pour avoir l'exclusivité, ce qui peut l'amener à simplement copier-coller et adapter à la ligne éditoriale une dépêche d'agence de presse

Rue89 vs. AgoraVox

C'est au niveau de la présentation que tout semble se jouer, donnant à Rue89 [qui devrait satisfaire des lecteurs exigeants envers une information qu'ils veulent complète et objective] un aspect plus professionnel avec les articles rédigés par des journalistes formés et/ou d'expérience. [...] AgoraVox donne une fausse impression d'amateurisme et, surtout, on sent un esprit différent, une sorte de proximité, de regard plus subjectif sur le monde, les événements. (Dugué 2008)

Typologie a priori

Expression de la subjectivité





H1 : Davantage de subjectivité :
AgoraVox > *Rue89* > *Le Soir*

H2 : Une subjectivité, un ton plus libre :
AgoraVox = *Rue89* > *Le Soir*

Organisation du discours

H3 : Des textes plus structurés et descriptifs :
Le Soir > *Rue89* > *AgoraVox*

Taille des 3(/4) sous-corpus

	Mots	Articles	Mots/Articles
	2 187 333	3 879	564
 <small>Le média citoyen</small>	3 281 208	4 368	751
	2 744 270	5 873	467
	8 212 811	14 120	582
 <i>BIS</i>	<i>3 000 000</i>	<i>6 000</i>	<i>???</i>

Plan

- 1 ARCproject
- 2 Constitution d'un corpus dédié
- 3 Méthodologie
 - Une étude contrastive
 - Nettoyage et formatage
 - Étiquetage
 - Marquage de traits
 - Marquage de traits associés à l'organisation discursive
 - Marquage de traits associés à l'expression de la subjectivité
 - Analyses quantitatives
- 4 Résultats




Méthodologie

- 1 Rubriques comparables
- 2 Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique (gros travail de nettoyage manuel)
- 3 Etiquetage de catégories grammaticales et dépendances syntaxiques (Treetagger + Syntex)
- 4 Marquage automatique de traits liés à l'organisation discursive et d'éléments potentiellement subjectifs
- 5 Analyses quantitatives (distributions spécifiques) et qualitatives (annotation via Glozz, Analec)

Méthodologie

- 1 Rubriques comparables
- 2 Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique (gros travail de nettoyage manuel)
- 3 Etiquetage de catégories grammaticales et dépendances syntaxiques (Treetagger + Syntex)
- 4 Marquage automatique de traits liés à l'organisation discursive et d'éléments potentiellement subjectifs
- 5 Analyses quantitatives (distributions spécifiques) et qualitatives (annotation via Glozz, Analec)

Rubriques comparables

		
Actualité	Monde Politique Société	International Europe Politique Société
Culture	Culture Médias	Culture Médias Religion Étonnant Mode et tendances People

Méthodologie

- 1 Rubriques comparables
- 2 Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique (gros travail de nettoyage manuel)
- 3 Etiquetage de catégories grammaticales et dépendances syntaxiques (Treetagger + Syntex)
- 4 Marquage automatique de traits liés à l'organisation discursive et d'éléments potentiellement subjectifs
- 5 Analyses quantitatives (distributions spécifiques) et qualitatives (annotation via Glozz, Analec)

Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique



Scripts perl + nettoyage manuel

- extraction des articles des archives (par URL)
- conversion des différents codes source (codage des caractères, propreté) en code xml
- sauvegarde des méta-données (rubrique(s), auteur, titre d'origine, URL, date de publication...)
- effacement de dépêches, entretiens, poèmes, chants
- encodage selon le modèle TEIP5

Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique



```
<teiCorpus>
+<teiHeader>
```

Décrit le corpus : son nom, qui a fait quoi, dans quel cadre (projetARC), d'où proviennent les textes, etc.

```
<TEI>
+<teiHeader>
Décrit le texte ...
```

```
<text>
+<front>
Lignes de titre
+<body>
Corps du texte
```

```
</text>
</TEI>
+<TEI>
+<TEI>
```

```
...
</teiCorpus>
```



Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique

```
<div type="chapo"> chapeau de l'article (facultatif)
  <p> paragraphe généralement simple </p>
</div>
```

```
<div type="level_1"> section de niveau 1 (facultatif)
  <head> titre de section (facultatif) </head>
  <p> texte <bold> texte </bold> texte <ital> texte </ital> texte </p>
```

```
<div type="level_2">
  <head> titre </head>
  <list type="bulleted"> | <list type="ordered">
    <item> texte </item>
    <item> <p> </p> <p> </p> </item>
    <item> <cit> <quote> « citation » </quote> </cit> </item>
  </list>
  <figure/>, <table/>, <video/>, <audio/>, <div type="entretien"/> balises auto-
  fermantes qui signalent la présence d'un objet dans le fichier source, mais n'en donne
  aucun détail ni contenu
</div>
```

```
</div>
```



Méthodologie

- 1 Rubriques comparables
- 2 Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique (gros travail de nettoyage manuel)
- 3 Étiquetage de catégories grammaticales et dépendances syntaxiques (Treetagger + Syntex)
- 4 Marquage automatique de traits liés à l'organisation discursive et d'éléments potentiellement subjectifs
- 5 Analyses quantitatives (distributions spécifiques) et qualitatives (annotation via Glozz, Analec)

Treetagger + Syntax

```

</front>
-<body>
-<div type="chapo">
  -<p>
    -<phrase num="1">
      -<txt>
        Depuis 2000 ans, l'Église s'est toujours opposée à la science et son histoire est marquée d'épisodes connus, comme la condamnation de Galilée, pour réfuter, régulièrement, les avancées de la science.
      </txt>
    -<etiq>
      Prep|depuis|Depuis|1||NOMPREP;2|1511 NomXXDate|2000|2000|2|NOMPREP;1|1518 NomXXDate|an|ans|3|1523 Typo|,|4|1526 Det??|le|5|DET;6|1528 NomPr|église|Église|6|SUJ;8|DET;5|1530 Pro|se|s'|7|REF;8|1537 VCONJS|être|est|8||SUJ;6,REF;7,AUX;10|1539 Adv|toujours|toujours|9|ADV;10|1543 PpaFS|opposer|opposée|10|AUX;8|ADV;9,PREP;11|1552 Prep|à|à|11|PREP;10|NOMPREP;13|1560 Det??|le|la|12|DET;13|1562 NomFS|science|science|13|NOMPREP;11|DET;12|1565 CCoord|et|et|14|1573 DetMS|son|son|15|DET;16|1576 NomMS|histoire|histoire|16|SUJ;17|DET;15|1580 VCONJS|p|être|est|17||SUJ;16,AUX;18|1589 PpaFS|marquer|marquée|18|AUX;17|PREP;19|1593 Prep|de|d'|19|PREP;18|NOMPREP;20|1601 NomP?|épisode|épisodes|20|NOMPREP;19|ADJ;21|1603 PpaMP|connaître|connus|21|ADJ;20|1612 Typo|,|22|1618 CSub|comme|comme|23|1620 Det??|le|la|24|DET;25|1626 NomFS|condamnation|condamnation|25|DET;24,PREP;26|1629 Prep|de|de|26|PREP;25|NOMPREP;27|1642 NomPr|Galilée|Galilée|27|NOMPREP;26|1645 Typo|,|28|1652 Prep|pour|pour|29||NOMPREP;30|1654 VINF|réfuter|réfuter|30|NOMPREP;29|OBJ;35|1659 Typo|,|31||I ADV;33|1666 Adv|régulièrement|régulièrement|32|1668 Typo|,|33|I ADV;31|1681 Det??|le|les|34|DET;35|1683 NomP?|avancées|avancées|35|OBJ;30|DET;34,PREP;36|1687 Prep|de|de|36|PREP;35|NOMPREP;38|1696 Det??|le|la|37|DET;38|1699 NomFS|science|science|38|NOMPREP;36|DET;37|1702 Typo|,|39|1709
    </etiq>
  -</phrase>
  -<phrase num="2">
    -<txt>
      On croyait que l'Église avait enfin compris qu'on ne pouvait s'opposer éternellement aux évidences scientifiques, que d'ailleurs cela ne gênait pas son développement ni son rayonnement, et qu'il suffisait de considérer que les textes sacrés avaient un aspect " symbolique " pour pouvoir réconcilier science et religion.
    </txt>
  </phrase>

```


Méthodologie

- 1 Rubriques comparables
- 2 Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique (gros travail de nettoyage manuel)
- 3 Etiquetage de catégories grammaticales et dépendances syntaxiques (Treetagger + Syntex)
- 4 Marquage automatique de traits liés à l'organisation discursive et d'éléments potentiellement subjectifs
- 5 Analyses quantitatives (distributions spécifiques) et qualitatives (annotation via Glozz, Analec)

Organisation discursive via la position initiale

Décrire l'organisation discursive via la position initiale (la position thème), définie par l'ensemble des éléments apparaissant en position préverbale :

- éléments détachés + sujets grammaticaux
- associés à leur position textuelle : début de paragraphe (P1) ou intraparagraphique (P2)

Position initiale et Orientation

Circonstants

temporels En 2002, le musicien s'envole enfin pour l'Asie. (Le Soir)

spatiaux Aux Pays-Bas, en Allemagne, aux Etats-Unis, au Canada, en Italie, en Finlande... les vertus médicales du cannabis sont peu à peu reconnues [...] (Rue89)

notionnels Selon de nouvelles estimations américaines, l'Arctique renfermerait des réserves inexploitées, [...] (Le Soir)

Séquenceurs

Pour d'autres, il s'agit d'une maladie chronique...c'est le cas d'Eric Deshayes et de Dominique Grimaud. (AgoraVox)

Ensuite, connaissez-vous beaucoup de cas où il y a eu des anomalies, des dérapages [...] (Le Soir)

Position initiale et Connexion, continuité topicale

en position sujet

pronoms personnels, SN possessifs, Répétitions ...

Le texte a été présenté par le président de la conférence, [...] || devrait être signé ce mardi lors de la cérémonie [...] (Le Soir)

Appositions détachées

Refusant cette pente fatale, Descartes réinvente la preuve ontologique, [...] (AgoraVox)

Connecteurs

Mais Michel Rollier, le patron du groupe, ne dit pas autre chose. (Rue89)

Éléments potentiellement subjectifs

La subjectivité : difficulté de repérage

- Une expression subjective n'est pas forcément utilisée dans sa signification subjective
- Forte dépendance du contexte, une expression à priori subjective peut être employé dans un sens objectif et vice versa

Élément Potentiellement Subjectif (PSE)

'linguistic element that may be used to express subjectivity' (Wiebe et al. 2004 : 281f)

Éléments potentiellement subjectifs

Pronoms 1/2 Ces fameuses années 68-70, qui **nous** submergent aujourd'hui, ça commence à **m'**énervé. (Rue89)

On assiste ainsi une soirée organisée en l'honneur des Amis américains de Versailles dans la Galerie des Glaces du fameux château. (LeSoir)

Adverbes modaux Car il serait **évidemment** bien dangereux de se replier frileusement sur les égoïsmes nationaux d'antan. (Rue89)

Clivées **C'est l'amer constat que** l'on peut faire soit qu'on y habite ou qu'on y arrive pour la première dans cette ville qui jadis, présentait fière allure. (AgoraVox)

il-impersonnel Pour Alain Menand, **il est de toute façon hasardeux de** prétendre 'classer' les différentes licences : [...] (Rue89)

hapax

Marquage systématique

```

- <body>
- <div type="chapo">
- <p>
- <phrase num="1">
  <span type="init" form="SP" fct="circ" sem="tps" postxt="P1" head="2000">Depuis 2000 ans</span>
  <span type="sujet" form="NPdef" genre_nombre="??" postxt="P1" head="église">l'Église</span>
  <span type="GV" head="être opposer" REF="1">s'est toujours opposée</span>
  <span type="complement" rel_verb="PREP" form="SP" fct="circ" sem="spa" postxt="P1" head="science">à la science et son histoire
  <span type="adjoint" form="undo" fct="todo" postxt="P1" head="épisode">est marquée d'épisodes connus</span>
  ,
  <span type="adjoint" form="Sub" fct="todo" postxt="P1" head="">comme</span>
  <span type="adjoint" form="undo" fct="todo" postxt="P1" head="condamnation">la condamnation de Galilée</span>
- <span type="adjoint" form="SP" fct="todo" postxt="P1" head="avancée">
  pour réfuter, régulièrement, les avancées de la science
  </span>
</phrase>
- <phrase num="2">
  <span type="sujet" class="loc" form="Pro12" genre_nombre="?S" postxt="P2" head="">On</span>
  <span type="GV" head="croire">croyait</span>
- <span type="complement" rel_verb="OBJ" form="Sub" fct="todo" postxt="P2" head="évidence">
  que l'Église avait enfin compris qu'on ne pouvait s'opposer éternellement aux évidences scientifiques
  </span>
- <span type="adjoint" form="Sub" fct="todo" postxt="P2" head="religion">
  que d'ailleurs cela ne gênait pas son développement ni son rayonnement, et qu'il suffisait de considérer que les textes sacrés avaient un
  symbolique " pour pouvoir réconcilier science et religion
  </span>
</phrase>
</p>
</div>
- <p>
- <phrase num="1">
  <span type="connect" form="Adv" head="bien">

```

Méthodologie

- 1 Rubriques comparables
- 2 Extraction automatique, normalisation au format XML selon la TEI-P5 semi-automatique (gros travail de nettoyage manuel)
- 3 Etiquetage de catégories grammaticales et dépendances syntaxiques (Treetagger + Syntex)
- 4 Marquage automatique de traits liés à l'organisation discursive et d'éléments potentiellement subjectifs
- 5 **Analyses quantitatives (distributions spécifiques)** et qualitatives (annotation via Glozz, Analec)

le Log-Likelihood ratio (LL)

Objectif : comparer les fréquences dans les 3 médias des éléments marqués systématiquement afin de vérifier grossièrement notre typologie a priori et dégager premières pistes à creuser qualitativement

Définition : le LL permet de comparer des fréquences deux à deux sur des corpus de grande taille (Rayson & Garside 2000)

Procédure : comparaison pour chaque couple de médias (A vs. R, A vs. S et R vs. S) des fréquences de chacun des éléments marqués, en tenant compte à chaque fois de leur position textuelle **PLUS** comparaison des fréquences pour chaque couple P1 / P2 dans chaque média

Interprétation : seuil de significativité fixé à $p < 0.0001 \Rightarrow$ tout $LL > 15,13$ dans un média ou une position textuelle révèle une distribution significative

Plan

- 1 ARCproject
- 2 Constitution d'un corpus dédié
- 3 Méthodologie
- 4 Résultats**
- 5 Conclusion

Structure des documents

	Rue89	AGORA VOX	lesoir.be	3 médias
mots	2 187 333	3 281 208	2 744 270	8 212 811
intertitres	687	896	715	2 298
articles	3 879	4 368	5 873	14 120
mots/article	564	751	467	582
phrases/article	50	68	42	53
paragraphes/article	12	14	8	11
phrases/paragraphe	2,7	3,5	2,5	2,9

AgoraVox le plus en tout (sauf nombre d'articles) et les ratio les plus élevés

Rue89 le moins au niveau des quantités; mais d'un rang intermédiaire
quant au ratio mots et para./article et phrases/paragraphe

Le Soir le plus d'article + les articles les plus courts

Distribution des traits associés à l'organisation discursive

Initiale	Rue89	AGORA VOX <small>Le média engagé</small>	lesoir.be	3 médias
Organisation discursive	74 271	110 518	102 115	286 904
Circonstant	12 032	14 814	15 659	42 505
séquenceur	5 973	9 856	7 553	23 382
connecteur	10 712	17 205	13 403	41 320
apposition	14 872	20 733	25 552	61 157
redénomination	13 449	22 796	19 168	55 413
pronom ou SN poss.	17 233	25 114	20 780	63 127
continuité topicale	30 682	47 910	39 948	118 540

Distribution des EPS – éléments potentiellement subjectifs

Phrases contenant un	Rue89	AGORA VOX	lesoir.be	3 médias
EPS	195 395	298 636	249 830	743 861
Pronom 1re pers. + On	29 468	42 952	24 413	96 833
SN poss. 1re pers.	4 715	7 467	4 531	16 713
adv. modalisateurs	3 229	6 059	3 201	12 489
impersonnelles	1 219	2 814	1 103	5 136
clivées	4 567	6 256	4 762	15 585
Nb de phrases total	181 192	122 630	153 974	457 796
Nb d'hapax*	26 849	38 006	29 777	94 632
Nb de token	2 187 333	3 281 208	2 744 270	8 212 811

Organisation du discours via la position initiale

Phrases contenant un	Rue89/AGORA VOX	AGORA VOX/lesoir.be	Rue89/lesoir.be
Organisation discursive	117 (R)	880 (S)	263 (S)
circonstant	316 (R)	427 (S)	P2 : 39 (S)
séquenceur	P1 : 22 (A)	P1 : 33 (A)	
connecteur	P1 : 16 (A)	P1 : 38 (A)	
apposition	P2 : 72 (R)	1 716 (S)	832 (S)
redénomination	P2 : 91 (A)	P1 : 35 (S)	94 (S)
pronom ou SN poss.		P1 : 151 (A)	P1 : 16 (R)
Continuité topicale		647 (S)	499 (S)

(R,A,S) indique le corpus présentant un écart significativement positif, R = Rue89, A = AgoraVox, S = Le Soir.

P1 indique que la différence significative n'a été observée qu'en P1, idem pour P2.

Organisation du discours via la position initiale

Phrases contenant un	Rue89/AGORA VOX	AGORA VOX/lesoir.be	Rue89/lesoir.be
Organisation discursive	117 (R)	880 (S)	263 (S)
circonstant	316 (R)	427 (S)	P2 : 39 (S)
séquenceur	P1 : 22 (A)	P1 : 33 (A)	
connecteur	P1 : 16 (A)	P1 : 38 (A)	
apposition	P2 : 72 (R)	1 716 (S)	832 (S)
redénomination	P2 : 91 (A)	P1 : 35 (S)	94 (S)
pronom ou SN poss.		P1 : 151 (A)	P1 : 16 (R)
Continuité topicale		647 (S)	499 (S)

Journalisme pro vs. citoyen

Davantage d'indices de l'organisation discursive dans LeSoir et Rue89

Organisation du discours via la position initiale

Phrases contenant un	Rue89/AGORA VOX	AGORA VOX/lesoir.be	Rue89/lesoir.be
apposition	P2 : 72 (R)	1 716 (S)	832 (S)
redénomination	P2 : 91 (A)	P1 : 35 (S)	94 (S)
pronom ou SN poss.		P1 : 151 (A)	P1 : 16 (R)
Continuité topicale		647 (S)	499 (S)

le Soir, topical

Une organisation a priori construite autour de continuités topicales (redénominations en P1, beaucoup d'appositions – style très journalistique)

Éleveur d'un cheptel de septante moutons à Theux, dans la province de Liège, Dany Gavage s'interrogeait depuis une dizaine de jours sur les symptômes développés par son bélier

rôle du paragraphe

Un rôle différent selon les médias ?

Organisation du discours via la position initiale

Phrases contenant un	Rue89/AGORA VOX	AGORA VOX/lesoir.be	Rue89/lesoir.be
circonstant	316 (R)	427 (S)	P2 : 39 (S)
séquenceur	P1 : 22 (A)	P1 : 33 (A)	
connecteur	P1 : 16 (A)	P1 : 38 (A)	

Le Soir et Rue89

circonstants pas spécialement en P1
des textes plus descriptif ?

AgoraVox

connecteurs/séquenceurs en P1
une organisation davantage
argumentative ?

Expression de la subjectivité via les EPS

Phrases contenant un	Rue89/AGORA VOX	AGORA VOX/lesoir.be	Rue89/lesoir.be
EPS		3321 (A)	2830 (R)
Pronom 1re pers. + On	39 (R)	2 396 (A)	2 529 (R)
SN poss. 1re pers.		297 (A)	188 (R)
adv. modalisateurs	90 (A)	460 (A)	104 (R)
impersonnelles	152 (A)	499 (A)	69 (R)
clivées	31 (R)	24 (A)	97 (R)
Nombre d'hapax*	300 (R)	26 (S)	139 (R)

Le Soir

Aucun LL du côté de
la presse traditionnelle

Rue89

constructions clivées
on dit...

AgoraVox

constructions
impersonnelles

Plan

- 1 ARCproject
- 2 Constitution d'un corpus dédié
- 3 Méthodologie
- 4 Résultats
- 5 Conclusion

Bilan

Des différences de subjectivité validées



Plus d'éléments potentiellement subjectifs dans les médias alternatifs

Une organisation – un type de texte ? – différents



Plus d'éléments associés à l'organisation discursive (indices de continuité topicale et circonstants) chez les journalistes "professionnels"

Des paragraphes longs, des connecteurs/séquenceurs en initiale chez les internautes

Bilan

Une méthodologie qui tourne

Corpus contrastif, marquage automatique, analyse quantitative, résultats significatifs

Un simple défrichage

Des traits, indices, éléments pertinents pour donner une première idée de la validité d'un corpus contrastif de presse francophone.

De nombreuses pistes à creuser (rôle du paragraphe, continuité topicale, les intertitres ?)

Un besoin d'analyse qualitative

Quelle proportion d'éléments potentiellement subjectifs réellement subjectifs en contexte ?

Observer sur le terrain l'opposition "journalisme professionnel" vs. "journalisme internaute" au niveau de l'organisation discursive ?

- Dugué B. (2008). *Rue89 et AgoraVox sont alternatifs, mais sont-ils différents dans l'esprit ?*, AgoraVox 7 janvier 2008, URL <http://www.AgoraVox.fr/actualites/medias/article/Rue89-et-AgoraVox-sont-alternatifs-33894>.
- Ho-Dac L.-M. (2010). *An exploratory data-driven analysis for describing discourse organization*, in A. Sánchez & M. Almela (eds.) *A Mosaic of Corpus Linguistics. Selected Approaches*, Frankfurt/Berlin : Peter Lang, 79-100.
- Rayson P. & Garside R. (2000). *Comparing corpora using frequency profiling*, Proceedings of the workshop on Comparing Corpora, held in conjunction with ACL 2000. 1-8 October 2000, Hong Kong, 1-6.
- Wiebe J., Wilson T., Bruce R., Bell M. & Martin M. (2004). *Learning subjective language*, Computational Linguistics 30(3) : 277-308.