

# Analyse de comportements relatifs exceptionnels expliquée par des textes : les votes du parlement européen.

Charles de Lacombe\*, Antoine Morel\*\*, Adnene Belfodil\*, François Portet\*\*, Cyril Labbé\*\*, Sylvie Cazalens\*, Marc Plantevit\*\*\*, Philippe Lamarre\*

\*Université de Lyon, INSA Lyon, LIRIS UMR 5205  
prenom.nom@liris.cnrs.fr

\*\*Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France  
prenom.nom@imag.fr

\*\*\*Université de Lyon, Université Claude Bernard Lyon 1, LIRIS UMR 5205  
prenom.nom@liris.cnrs.fr

**Résumé.** L’algorithme de fouille DEBuNk permet d’identifier des groupes et des contextes montrant un comportement relatif exceptionnellement différent par rapport à celui généralement observé pour ces mêmes groupes. Pour rendre la compréhension et l’interprétation des résultats plus aisées, l’interface graphique inclut un module de génération de texte qui transcrit les motifs trouvés en langage naturel. Nous illustrons notre approche avec la plate-forme web ANCORE, sur les données de vote du parlement européen durant les deux derniers mandats.

## 1 Introduction

La plateforme web ANCORE<sup>1</sup> (Analyse de Comportements Relatifs Exceptionnels) permet de découvrir, dans le contexte des votes du parlement européen, des groupes d’individus et des contextes montrant des différences remarquables concernant leur entente, ou mésentente, usuelle. Par exemple, alors que, globalement, les votes des députés européens du Rassemblement National et du Front de Gauche témoignent d’un fort désaccord entre ces deux partis, on observe une convergence des votes pour la thématique “Relations extérieures avec l’UE”.

La plateforme (section 2) met en œuvre l’algorithme de fouille de motifs exceptionnels DEBuNk (Belfodil et al. (2017)). La visualisation des résultats intègre un module de génération automatique de textes facilitant la compréhension et l’interprétation des résultats. La démonstration développe un scénario de vérification (section 3) d’affirmation et montre comment apporter un éclairage sur les accords et les désaccords de groupes parlementaires, tels que reflétés par les votes.

## 2 La plateforme ANCORE

**Module DEBuNk (Belfodil et al. (2017)).** Il relève d’une tâche EMM (Exceptional Model Mining (Duivesteijn et al. (2016))). Il s’applique à des “données comportementales”  $\langle E, I, O, o \rangle$

1. Accessible via le lien : [contentcheck.liris.cnrs.fr](http://contentcheck.liris.cnrs.fr)

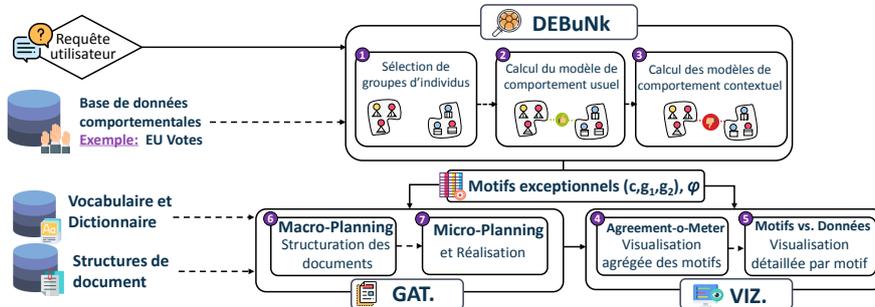


FIG. 1 – Aperçu général de la plateforme ANCORE

où un ensemble d'individus ( $I$ ) expriment des actions ( $O$ ) (par ex. votes, opinions, évaluations) ( $o : I \times E \rightarrow O$ ) sur des entités ( $E$ ). Par exemple, au Parlement Européen<sup>2</sup>, des parlementaires (individus) votent ('pour', 'contre', 'abstention') sur des scrutins (entités). Entités et individus sont décrits par des attributs qui peuvent être catégoriques (ex. commission), numériques (ex. date) ou hiérarchiques (ex. hiérarchie de thématiques). Une conjonction de restrictions sur les valeurs des attributs permet de définir, soit un groupe d'individus, soit un groupe d'entités aussi appelé *contexte*. Certains contextes, particulièrement clivants ou très consensuels, induisent des changements très importants par rapport aux comportements usuels. DEBuNk découvre de tels contextes ainsi que les couples de groupes pour lesquels de tels changements sont observables.

En considérant deux groupes d'individus  $g_1$  et  $g_2$  dans un contexte  $c \subseteq E$ , DEBuNk estime la similitude de leurs comportements en étudiant le nombre de scrutins sur lesquels les deux groupes ont voté de la même manière<sup>3</sup> :  $sim^c(g_1, g_2) = \frac{\sum_{e \in c} \mathbb{1}[M(g_1, e) = M(g_2, e)]}{|c|}$ . Comparer cette similarité avec celle obtenue en considérant tous les scrutins  $sim^*(g_1, g_2)$  permet d'évaluer l'impact du contexte  $c$  sur les groupes  $g_1$  et  $g_2$  (cf. 1, 2 et 3 de la Figure 1).

L'espace à explorer est si important qu'une évaluation par force brute se révèle impossible. DEBuNk met en œuvre plusieurs techniques (opérateurs de fermeture, heuristiques optimistes...) pour fournir une liste exhaustive des motifs pertinents avec leur qualité  $\langle p = (c, g_1, g_2), \varphi(p) \rangle$  tout en validant des contraintes données a priori par l'utilisateur : seuil minimal de qualité, taille minimale d'un groupe, taille minimale d'un contexte.

**Modules GAT et VIZ.** La plateforme ANCORE vise à pallier la difficulté d'interprétation des résultats issus de DEBuNk en fournissant une description claire, en anglais, exprimant de manière neutre la signification de chacun de résultats.

La difficulté la plus importante réside dans la description des groupes identifiés. Ceux-ci pouvant être formés à partir des propriétés individuelle des députés (nom, prénom, date de naissance, genre), des groupes parlementaires (ex : S&D), des partis politiques (ex : LR), des propriétés des pays (population, système politique, langues, GDP, nombres de sièges, appartenance à l'espace Schengen...), il en résulte un grand nombre de possibilités.

<sup>2</sup> Parltrack <http://parltrack.euwiki.org/> - données du Parlement Européen.  
<sup>3</sup> Avec  $M$  fonction calculant le vote majoritaire d'un groupe  $g$  sur un scrutin  $e$ ;  $\mathbb{1}(\cdot)$  fonction indicatrice (1 si la condition paramètre est vraie, 0 sinon); et,  $\bullet$  support de la description  $\bullet$ .

L'approche classique de GAT consiste à utiliser des modèles de phrases prédéfinis permettant de décrire chacune des associations possibles entre propriétés. Le nombre de possibilités étant ici très important, les modèles de phrases sont générés automatiquement en fonction des résultats de DEBuNk. Une propriété est décrite par un adjectif et le reste par un complément. Le paramètre mis en adjectif est choisi selon l'ordre de priorité suivant : genre, nationalité, parti européen, parti national. La date de naissance est toujours donnée en complément.

Les modèles de phrase prennent la forme d'arbres capturant la structure syntaxique de la phrase. Ces modèles sont complétés en fonction des résultats fournis et un réalisateur de surface SimpleNLG (Gatt et Reiter (2009)) permet de produire la forme finale de la phrase en appliquant les règles grammaticales : accords en nombre et en genre des verbes et des adjectifs.

En complément des informations incluses dans le motif et de la transcription textuelle, le module VIZ fournit d'autres éléments synthétiques comme par exemple l'*Agreement-O-Meter* qui décrit grâce à une jauge le degré d'accord usuel entre les deux groupes et l'intensité du changement observé. L'utilisateur peut aussi naviguer dans les données correspondant à un motif pour mieux l'analyser et le comprendre : scrutins correspondant au motif, code couleur indiquant le niveau d'accord/désaccord, lien vers leur descriptif officiel, votes individuels ...).

### 3 Scénario

L'article "Les groupes au Parlement européen, des alliances parfois surprenantes"<sup>4</sup>. En parlant du PPE, le principal groupe politique de la 8e législature, il affirme que la volonté de rassembler le plus de partis possible occasionne parfois "des alliances parfois surprenantes". Un exemple en particulier est donné, celui du parti Fidesz (Hongrie). Nous pouvons nous poser plusieurs questions pour étudier cette affirmation, dont :

- Le Fidesz est-il en conflit par rapport au reste du PPE ?
- Le Fidesz a-t-il des conflits avec certains partis spécifiques du PPE ?
- Existe-t-il d'autres conflits au sein du PPE ?

Nous étudions ici les deux premières questions. La troisième le sera lors de la démonstration.

**Le Fidesz par rapport au reste du PPE.** Grâce à DEBuNk, nous recherchons les sujets conflictuels entre le Fidesz et le reste du PPE. En filtrant les résultats pour n'afficher que les motifs qui concernent l'ensemble du PPE, le premier chiffre qui ressort est que le Fidesz vote en accord avec le reste du groupe pour 94% des scrutins. Les motifs de désaccord les plus importants trouvés concernent l'agriculture et le fonctionnement administratif de l'UE.

**Le Fidesz par rapport à d'autres partis du PPE.** Avec les mêmes paramètres de recherche, on s'intéresse cette fois aux motifs qui opposent le Fidesz à un autre parti. Celui qui présente le plus grand désaccord exceptionnel avec le Fidesz est le Partido Popular espagnol, avec une chute de 91% à 10% d'accord, pour deux contextes (figure 2) identifiés par les sujets : [2 *Internal market - 4.10 Social policy*] d'une part et [4.10.07 *The elderly*] d'autre part. Parmi les autres partis présentant de forts désaccords avec le Fidesz sur des sujets spécifiques, on trouve plusieurs partis, tous constitués de 3 députés ou moins.

Nous recherchons ensuite les motifs d'accord exceptionnel avec d'autres partis membres du PPE ; pour obtenir des résultats, nous devons baisser le seuil minimal de qualité, ce qui tend à montrer que le PPE a globalement une ligne cohérente. Cette requête fait ressortir deux partis :

---

4. [goo.gl/43MM3k](http://goo.gl/43MM3k), article paru sur le site de la RTBF le 23 octobre 2015

## ANCORE: Analyse de comportements relatifs exceptionnels expliquée par des textes

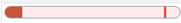
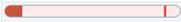
Patterns	Relative change	Group A	Group B	Context
 Normal agreement: 91% Context agreement: 10% Intensity: 81%	-89%	NATIONAL PARTY: Fidesz-Magyar Polgári Szövetség-Kereszténydemokrata Néppárt	NATIONAL PARTY: Partido Popular	PROCEDURE SUBJECT: - 2 Internal market, single market - 4.10 Social policy, social charter and protocol
 Normal agreement: 91% Context agreement: 10% Intensity: 81%	-89%	NATIONAL PARTY: Fidesz-Magyar Polgári Szövetség-Kereszténydemokrata Néppárt	NATIONAL PARTY: Partido Popular	PROCEDURE SUBJECT: - 4.10.07 The elderly
The first group gathers Fidesz-Magyar Polgári Szövetség-Kereszténydemokrata Néppárt MEPs. The second group gathers Partido Popular MEPs. In general terms the two groups are in agreement. But on 30 ballots we observe a strong disagreement between majorities of the two groups. This is witnessed by a decrease of similar majority vote from 91% to 10%. The conflictual subject is: - 4.10.07 The elderly The two groups bring together 12 and 19 individuals respectively.				

FIG. 2 – Motifs de désaccord entre le Fidesz et le Partido Popular - Le texte décrit le motif

le Partido da Terra (Portugal), et le Centre Démocrate Humaniste (Belgique, mentionné dans l'article), tous deux constitués d'un seul député et présentant respectivement un accord habituel de 75% et 76% avec le Fidesz. Les partis qui présentent des situations d'accord exceptionnel avec le Fidesz, et qui sont donc dans le cas général suffisamment en désaccord avec, sont les mêmes que précédemment. À noter qu'ils sont constitués de peu de membres.

Dans des travaux futurs, il serait intéressant d'explorer d'autres mesures de similitude de comportements pour découvrir des contextes exceptionnellement clivants/consensuels au sein d'un même parti/groupe politique, par exemple, en exploitant l'index de cohésion (Hix et al. (2003)) ou l'index de Krippendorff (Hayes et Krippendorff (2007)).

**Remerciements.** Ce travail a été partiellement financé par le projet ANR *ContentCheck*.

## Références

- Belfodil, A., S. Cazalens, P. Lamarre, et M. Plantevit (2017). Flash points : Discovering exceptional pairwise behaviors in vote or rating data. In *ECML/PKDD*.
- Duivesteijn, W., A. J. Feelders, et A. Knobbe (2016). Exceptional model mining. *Data Mining and Knowledge Discovery* 30(1), 47–98.
- Gatt, A. et E. Reiter (2009). Simplenlg : A realisation engine for practical applications. In *The 12th European Workshop on Natural Language Generation*, ENLG.
- Hayes, A. F. et K. Krippendorff (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures* 1(1), 77–89.
- Hix, S., A. Kreppel, et A. Noury (2003). The party system in the european parliament : Collusive or competitive? *JCMS : Journal of Common Market Studies*.

## Summary

We use DEBuNk , a data mining algorithm which enables identifying groups and contexts showing a relative behavior that is exceptional, compared to the usual relative behavior of these groups. For a better understanding and interpretation of the results, the visual interface includes a text generation module which transcribes the obtained patterns into texts. We illustrate our approach through the web platform ANCORE with voting data from the European Parliament.