

# HackaTAL 2019: participation de l'équipe Karaoké

---

Plein de gens <sup>1</sup>

<https://github.com/nicolasdugue/hackatal2019>

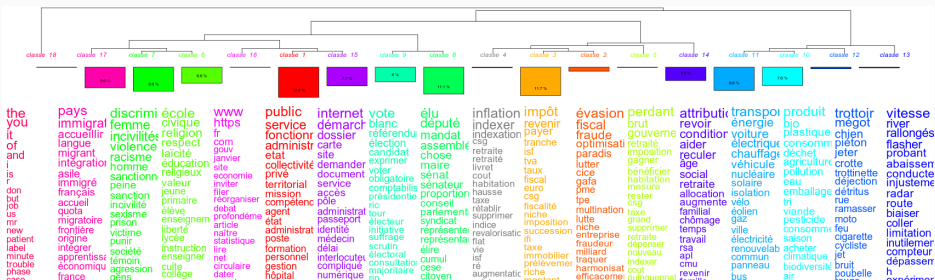
<sup>1</sup>Plein d'institutions

# Karaoké : analyse du grand débat national

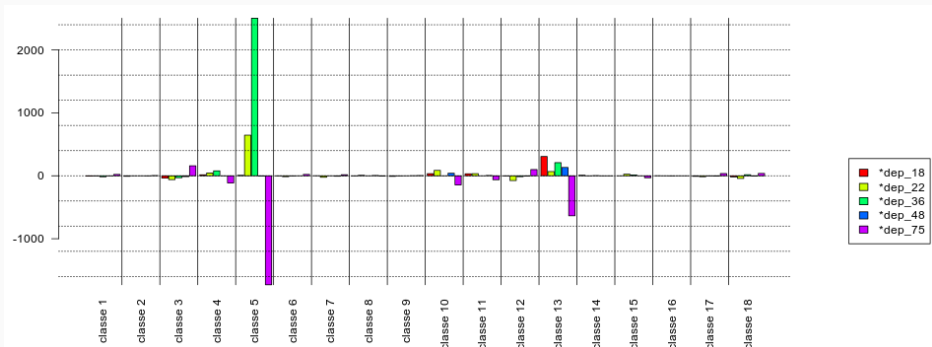
Plusieurs angles d'attaque :

- Catégorisation thématique des données ;
- Influence géographique/densité sur les thématiques ;
- Catégorisation basée syntagmes ;

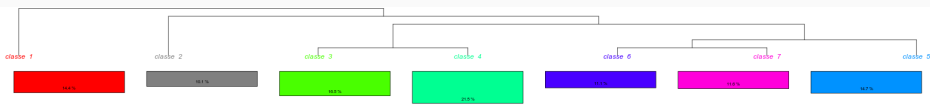
# Lucie – Catégorisation thématique



# Lucie – Catégorisation thématique : l'influence géographique



- extraire et classer propositions et centres d'intérêt
- dépendances syntaxiques avec UDPipe → extraire patrons syntaxiques :
  - V → NC (ex : *supprimer sénat*)
  - V → NC → ADJ (ex : *faire politique fiscal*)
  - V → NC → PREP → DET (ex : *traiter conducteur en citoyen*)
  - NC → ADJ (ex : *gilet jaune*)
  - NC → PREP → DET (ex : *communauté de commune*)
- analyse spécificité avec Iramuteq
- 7 clusters intéressants



traiter conducteur  
pénaliser conducteur  
acteur indispensable  
condition de circulation  
réseau secondaire  
détourner attention  
acteur de sécurité  
h sur réseau  
service de conduct  
mettre politique de  
vitesse réduire  
mettre politique rou  
faire politique fisca  
avis de français  
baisser limitation d

ancien président  
assemblée national  
président de républic  
nombre de député  
supprimer sénat  
suppression de sénat  
réduction de nombre  
cumul de mandat  
réduire nombre de dépu  
premier ministre  
conseil économique  
diminution de nombre  
haut fonctionnaire  
diminuer nombre de dépu  
conseil constitutionnel  
nombre de parlementaire  
train de vie  
cour de compte

carte gris  
carte de identité  
démarche administratif  
service de carte  
avoir interlocuteur  
demande de carte  
horaire de ouverture  
délai long  
sécurité social  
carte vital  
tout démarche  
simplifier démarche  
état civil  
temps de attente  
papier de identité  
demande de retraite  
acte de naissance  
avoir internet  
faire démarche  
caisse de retraite

avoir droit  
pôle emploi  
tout monde  
avoir an  
avoir besoin  
pôle emploi  
créer entreprise  
faire travail  
trouver emploi  
faire chose  
recherche de emploi  
demandeur de emploi  
prise en charge  
parcours de combattant  
avoir temps  
situation de handicap  
chercher emploi  
trouver travail

collectivité local  
service fiscau  
mission régalién  
aménagement de terri  
niveau national  
niveau local  
supprimer doublon  
collectivité territoriale  
administration central  
principe de subsidiarité  
niveau régional  
niveau de collectivité  
échelon administratif  
niveau de région  
développement économi  
administration fiscal  
déclaration de revenu  
état central

fonction public  
éducation national  
statut de fonctionnaire  
education national  
secteur privé  
emploi à vie  
fonction régalién  
statut de fonction  
ressource humain  
mettre fin  
droit privé  
chef de établissement  
grand corps  
gestion de ressource  
supprimer statut de fonctio  
fonction de état  
enseignement supérieur  
autre pays  
agent public  
agent de fonction

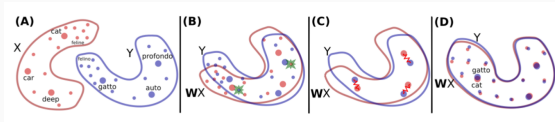
service public  
zone rural  
maison de service  
service de santé  
désert médical  
maison de santé  
milieu rural  
service de proximité  
territoire rural  
tout territoire  
petit commune  
maison médical  
forme de service  
transport public  
nouveau forme  
création de maison  
réseau routier  
fermeture de service  
accès à service

Deux corpus → Deux espaces de représentation (embeddings)



## Comparer les représentations du vocabulaire

Hypothèse : détecter l'utilisation spécifique du vocabulaire à l'occasion du grand débat



**Figure 1:** Word Translation Without Parallel Data : A Conneau, G Lample, MA Ranzato, L Denoyer, H Jégou



# Analyse sémantique : les mots dont la représentation a le plus changé d'après le procruste



A word cloud centered around the word 'gilet'. The words are arranged in a circular pattern. The word 'gilet' is the largest and most central. Other words include 'casseur' (top), 'jaune' (top-left), 'novembre' (top-right), 'rouge' (left), and 'mouvement' (bottom-right). The colors of the words are orange, yellow, and red.

Sur wikipedia : **Gilet**

veste, pantalon, pull, pare-balles, jupe, blouson, etc.

# Analyse sémantique : les mots dont la représentation a le plus changé d'après le procruste



Sur wikipedia : **Passé**

présent, ressurgir, souvenir, resurgi, revivre, futur, etc.

# Analyse sémantique : les mots dont la représentation a le plus changé d'après le procruste



Sur wikipedia : **Désert**

désertique, aride, oasis, saharien, chameau, dune, immensité, etc.

# Méthodologie

## Contribution en ligne au Grand Débat

05. Afin de baisser les impôts et réduire la dette, quelles dépenses publiques faut-il réduire en priorité ? (1 seule réponse possible)

- ☐ Les dépenses de l'Etat ☐ Les dépenses sociales  
☐ Les dépenses des collectivités territoriales ☐ Je ne sais pas

06. Parmi les dépenses de l'Etat et des collectivités territoriales, dans quels domaines faut-il faire avant tout des économies ? (plusieurs réponses possibles)

- ☐ L'éducation et la recherche ☐ La défense ☐ La sécurité  
☐ Les transports ☐ L'environnement ☐ La politique du logement  
☐ Autres (préciser) :

Coordination des  
politiques



Thierry

Protection animale



baldie

Reconversion automobile



JC

Ecologie



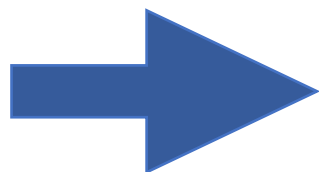
Hubert

Une écologie sans laissés  
pour compte



TDauphin

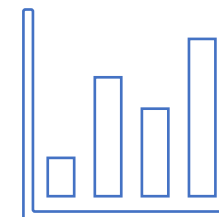
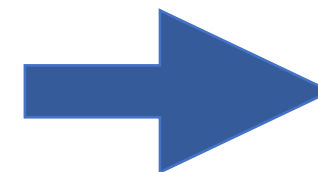
Croisement avec le  
code postal



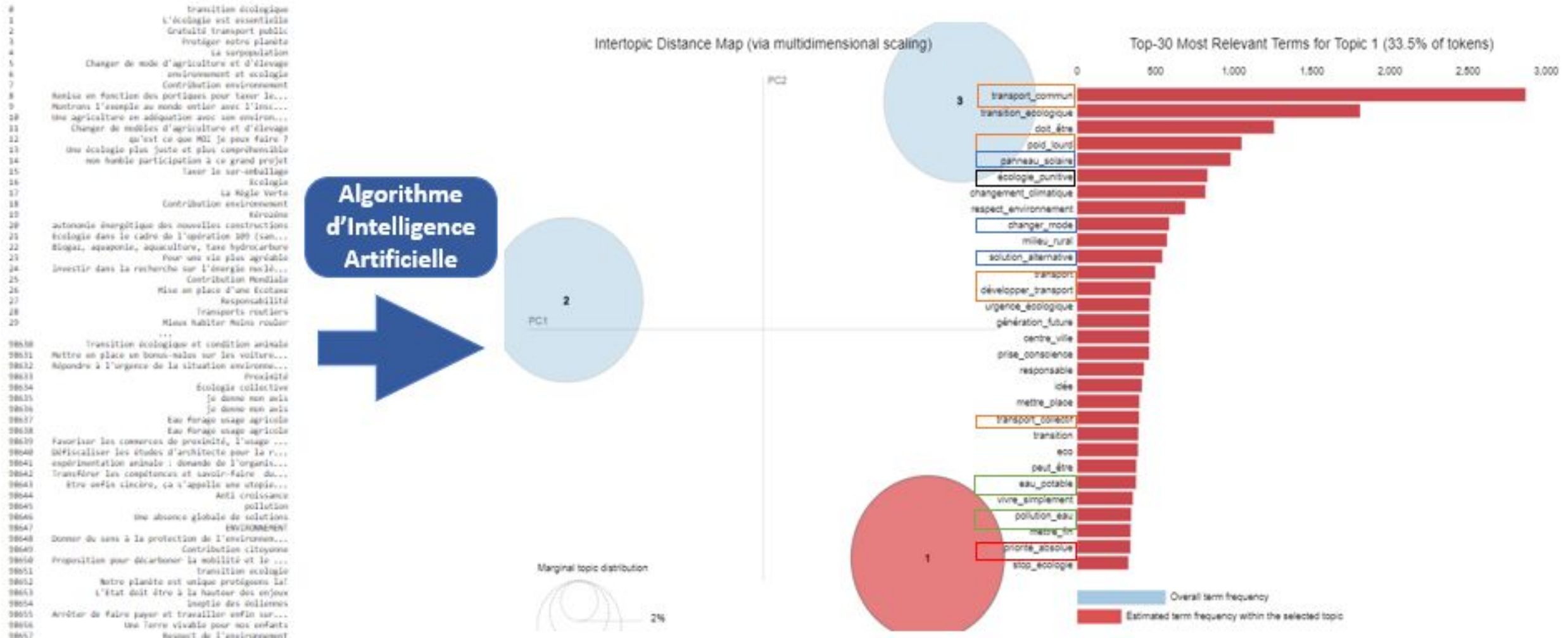
## Données socio-politiques par commune



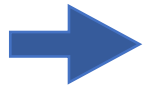
Analyse du  
langage



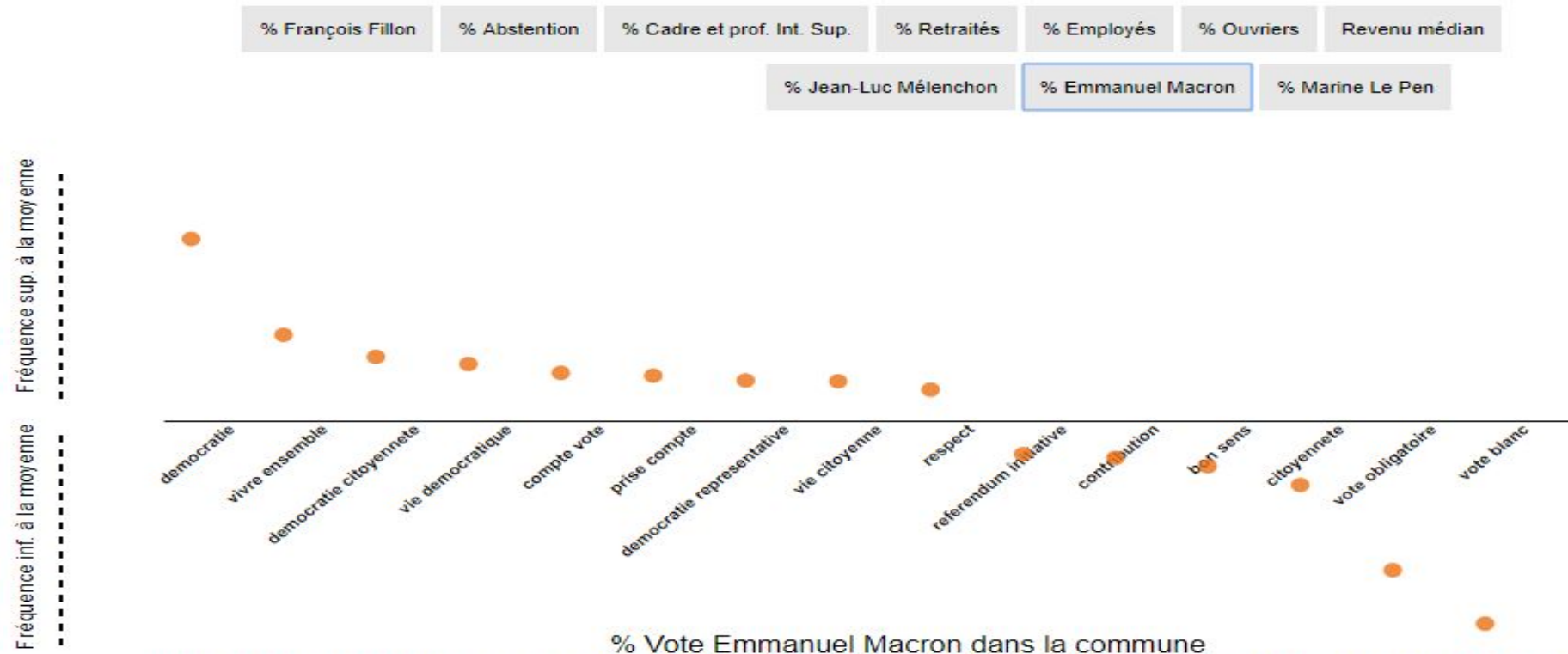
# Latent Dirichlet Allocation



# Permettre à tous les citoyens de faire leur analyse



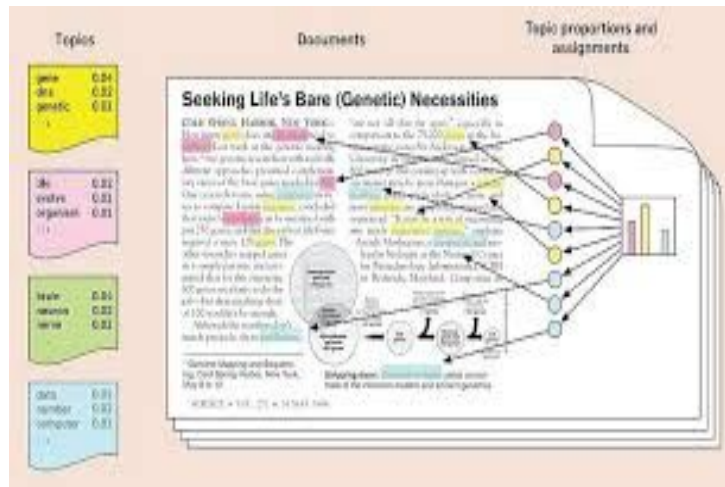
Un site web décrivant les réponses aux questions fermées et les sujets des contributions fonction des variables socio-politiques des Français



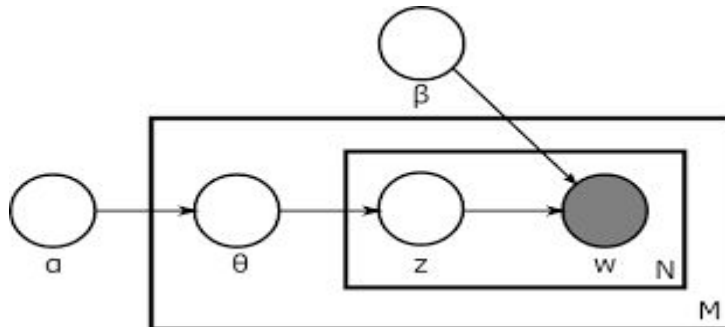
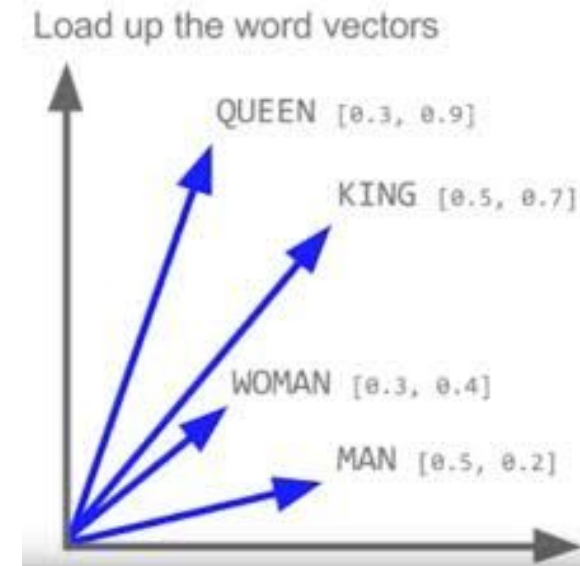
Note de lecture: les mots au-dessus de l'axe apparaissent plus que la moyenne dans les titres d'un participant habitant dans une commune avec un % Vote Emmanuel Macron élevé, les mots au-dessous de l'axe apparaissent moins que la moyenne dans les titres d'un participant habitant dans une commune avec un % Vote Emmanuel Macron élevé.

# LDA2vec

LDA + word2vec = lda2vec



+



# Conclusions et Futures Développements

## ●Résultats:

<https://louisv123.github.io/analyses-gdn/index.html#participation>

<https://github.com/nicolasdugue/hackatal2019>

## ●Conclusion:

- Un diagnostic donnant des pistes de réflexions pour inciter tous les citoyens à participer aux conclusions du Grand Débat National
- Outil d'aide à décision mais pas un diagnostic définitif