

# Métriques et TAL

N. Friburger  
Master 2 Pro

# Besoins

---

- ▶ **Evaluation des systèmes**
  - ▶ Recherche d'information
  - ▶ Classification et catégorisation de l'information
  - ▶ Extraction d'information
  - ▶ Résumé automatique
  - ▶ Indexation
  - ▶ Recherche d'information multilingue
  - ▶ Aide à la compréhension
- ▶ **Mais aussi**
  - ▶ Analyse syntaxique, étiquetage, etc.

# Exemple

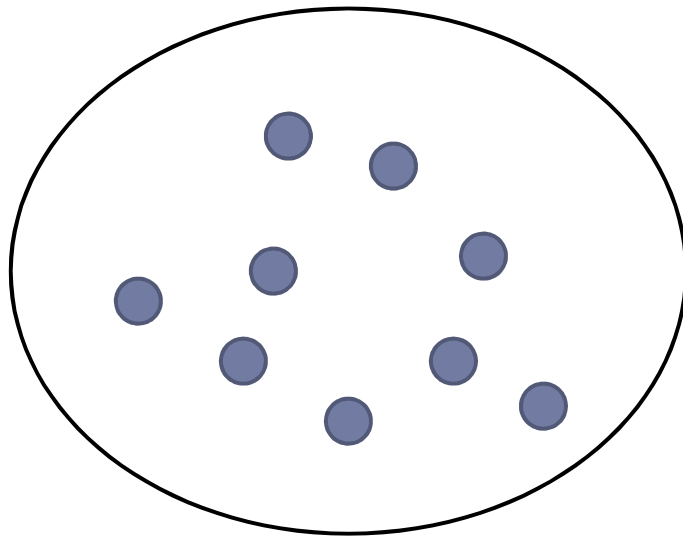
---

- ▶ Soit un ensemble de documents
  - ▶ 9 textes sur le football français ●
  - ▶ 9 textes sur la politique 😊
  - ▶ 3 textes sur le cinéma français ★
  - ▶ 6 textes sur l'économie 🌙
  
- ▶ Nous testons les résultats donnés par 4 systèmes de recherche de documents différents
  - ▶ Requête : football français

# Exemple

---

- ▶ Un utilisateur utilise un système 1 pour chercher les documents portant sur le football français
  - ▶ Résultat

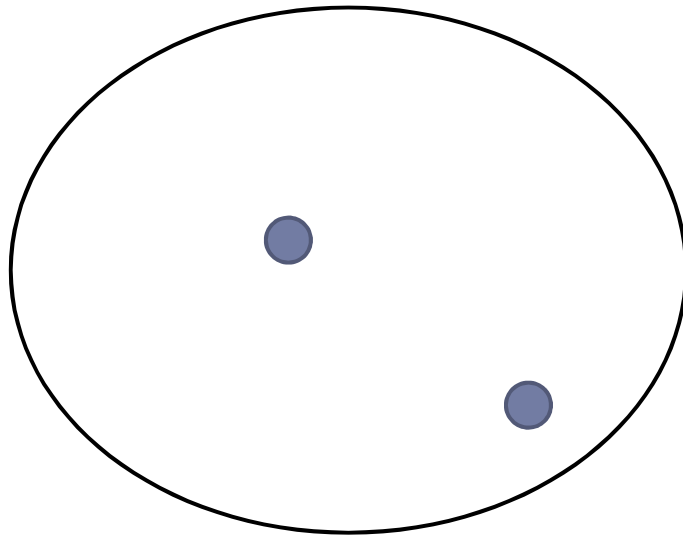


- ▶ Situation idéal
  - ▶ Tous les documents sur le football français sont trouvés,
  - ▶ Et seulement ceux là !

# Exemple

---

- ▶ Un utilisateur utilise un système 2 pour chercher les documents portant sur le football français
  - ▶ Résultat

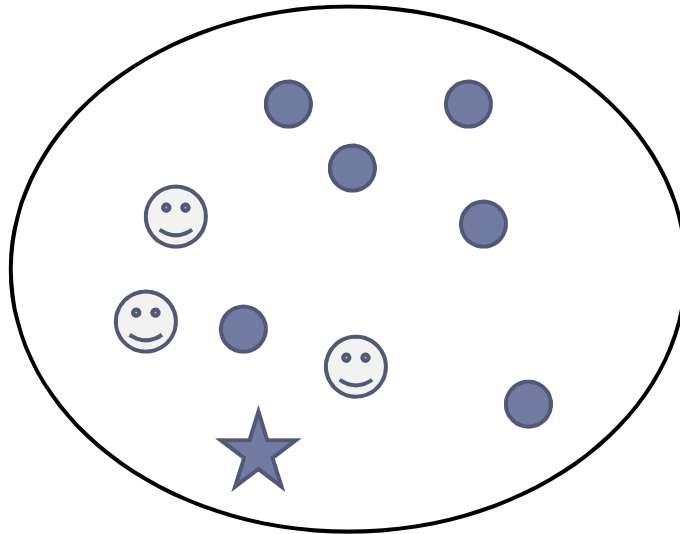


- ▶ 2 documents trouvés
- ▶ pas d'erreur

# Exemple

---

- ▶ Un utilisateur utilise un système 3 pour chercher les documents portant sur le football français
  - ▶ Résultat

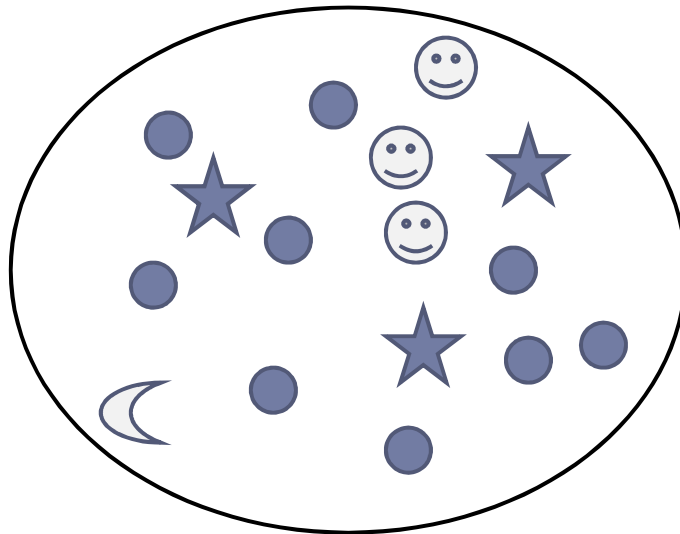


- ▶ Une partie des documents sont trouvés
- ▶ Présence d'erreurs

# Exemple

---

- ▶ Un utilisateur utilise un système 4 pour chercher les documents portant sur le football français
  - ▶ Résultat



- ▶ Tous les documents sont trouvés
- ▶ Mais accompagnés de beaucoup d'erreurs

## *La mesure de Rappel*

---

$$\text{Rappel} = \frac{\text{nbre de docs pertinents retrouvés}}{\text{nbre de docs pertinents à retrouver}}$$

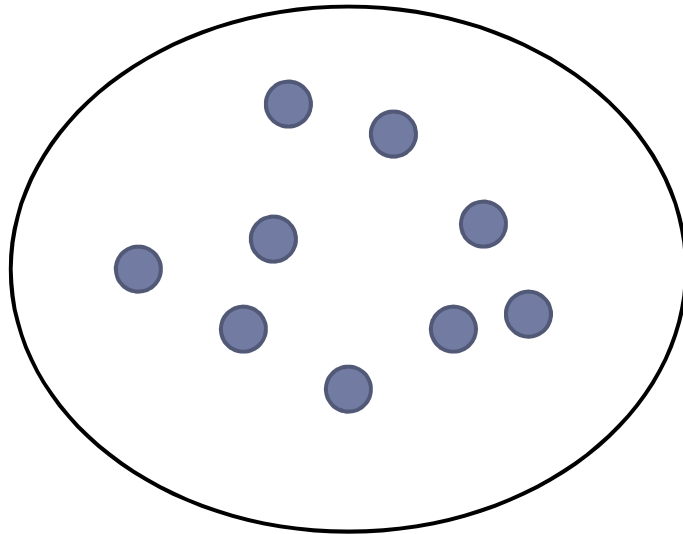
$$0 \leq \text{Rappel} \leq 1$$



## *La mesure de Rappel*

---

- ▶ Un utilisateur utilise un système 1 pour chercher les documents portant sur le football français
  - ▶ Résultat

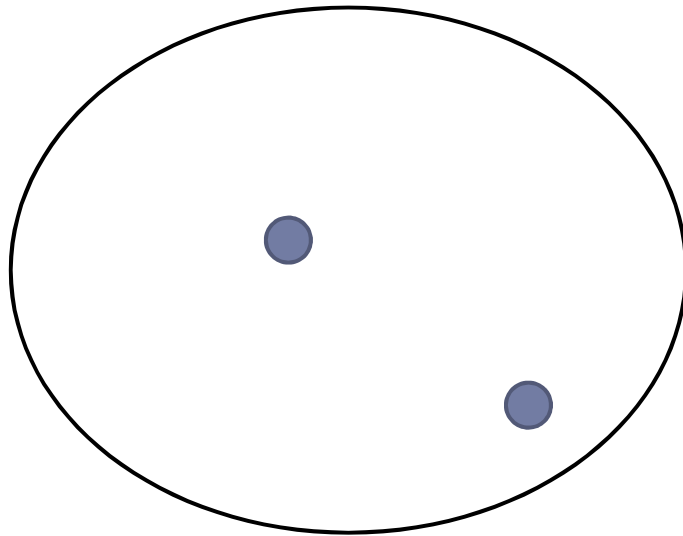


- ▶ Rappel =  $9/9 = 1$

## *La mesure de Rappel*

---

- ▶ Un utilisateur utilise un système 2 pour chercher les documents portant sur le football français
  - ▶ Résultat



- ▶ Rappel =  $2/9 = 0,22$
- ▶ Il manque des documents ; il y a donc des "silences"

# *La mesure de Rappel*

---

## ▶ **Silence**

- ▶ Réponses pertinentes qui ne sont pas proposées par le système alors qu'elles existent
- ▶  $\text{silence} + \text{rappel} = 1$ 
  - ▶ Plus le silence est grand, plus le rappel est faible.

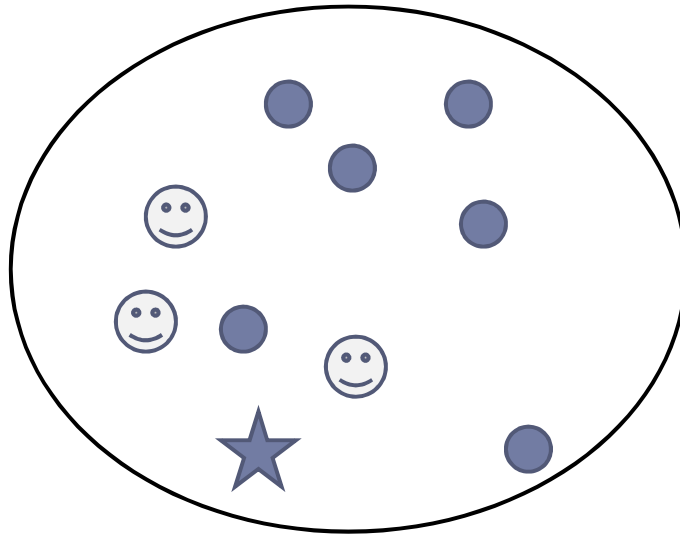
## ▶ **Causes du silence**

- ▶ Requête trop "précises"
- ▶ Indexation des docs insuffisante

## *La mesure de Rappel*

---

- ▶ Un utilisateur utilise un système 3 pour chercher les documents portant sur le football français
  - ▶ Résultat

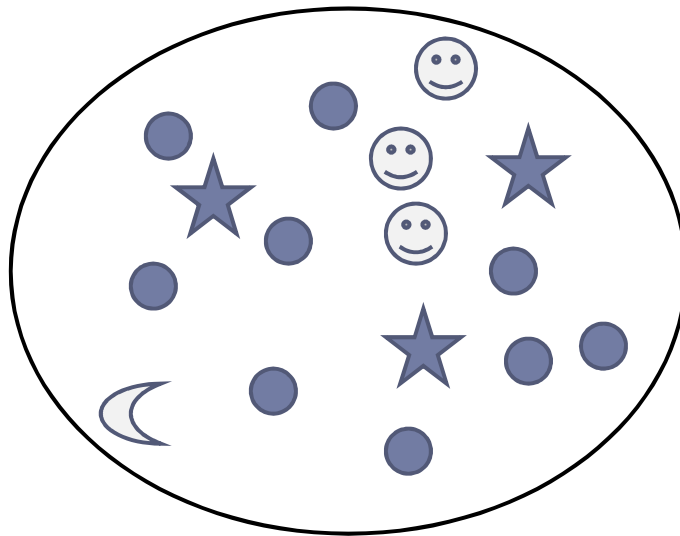


- ▶  $Rappel = 6/9 = 0,67$
- ▶ Des silences et des erreurs ...

# La mesure de Rappel

---

- ▶ Un utilisateur utilise un système 4 pour chercher les documents portant sur le football français
  - ▶ Résultat



- ▶ Rappel =  $9/9 = 1$
- ▶ Pas de silences mais beaucoup d'erreurs !
- ▶ En terme de rappel, le système est excellent, il faudrait pouvoir rendre compte des erreurs dans une mesure

# La mesure de Précision

---

- ▶ Le rappel n'est pas une mesure suffisante (systèmes 3 et 4)
  - ▶ Le rappel ne montre pas qu'il y a des erreurs : documents non pertinents !

Précision =  $\frac{\text{nbre de docs pertinents retrouvés}}{\text{nbre de docs total proposé par le moteur de recherche pour une requête donnée}}$

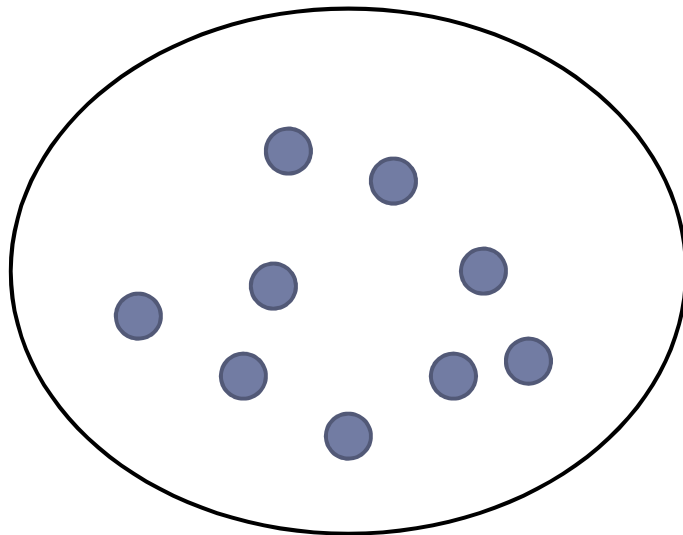
$$0 \leq \text{précision} \leq 1$$

- ▶ Si la précision est élevée, peu de documents inutiles sont proposés par le système → système "précis"

## *La mesure de Précision*

---

- ▶ Un utilisateur utilise un système 1 pour chercher les documents portant sur le football français
  - ▶ Résultat

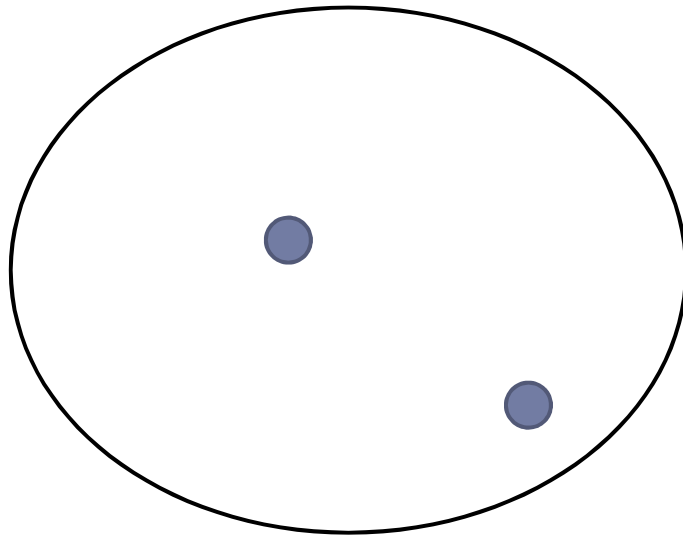


- ▶ Rappel =  $9/9 = 1$
- ▶ Précision =  $9/9 = 1$

## *La mesure de Précision*

---

- ▶ Un utilisateur utilise un système 2 pour chercher les documents portant sur le football français
  - ▶ Résultat



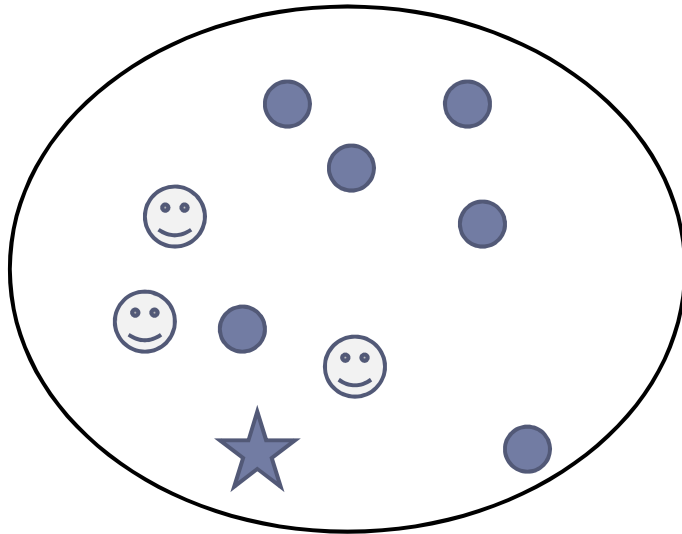
- ▶ Rappel =  $2/9 = 0,22$
- ▶ Précision =  $2/2 = 1$



# La mesure de Précision

---

- ▶ Un utilisateur utilise un système 3 pour chercher les documents portant sur le football français
  - ▶ Résultat



- ▶ Rappel =  $6/9 = 0,67$
- ▶ Précision =  $6/10 = 0,6$
- ▶ Le système présente des documents inutiles par rapport à la requête, il y a donc du bruit

# *La mesure de Précision*

---

## ▶ Bruit

- ▶ Réponses non-pertinentes proposées par un système
- ▶ bruit + précision = 1
  - ▶ Plus le bruit est grand, plus la précision est faible

## ▶ Causes du bruit

- ▶ Requête trop "floue"
  - ▶ Manque de termes dans une requête,
  - ▶ ou termes coordonnés par "ou"

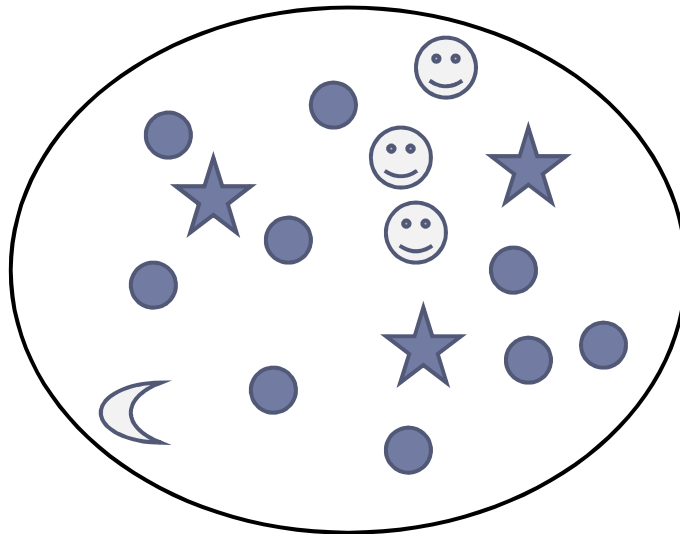
## ▶ Remarque

- ▶ Les résultats du système 3 contiennent à la fois du bruit et du silence

## *La mesure de Précision*

---

- ▶ Un utilisateur utilise un système 4 pour chercher les documents portant sur le football français
  - ▶ Résultat



- ▶ Rappel =  $9/9 = 1$
- ▶ Précision =  $9/16 = 0,56$

# *F-mesure*

---

## ▶ F-mesure

- ▶ combine la précision et le rappel avec une pondération
  - ▶ De manière générale, F-Mesure  $F_\beta$  ( $\beta > 0$ )

$$F_\beta = (1 + \beta^2) \cdot (\text{precision} \cdot \text{rappel}) / (\beta^2 \cdot \text{precision} + \text{rappel}).$$

- ▶ Si  $\beta = 1$ ,

$$F = 2 \cdot (\text{precision} \cdot \text{rappel}) / (\text{precision} + \text{rappel}).$$

# Rappel et précision de plusieurs classes

---

- ▶ Ex : On veut calculer le rappel et la précision d'un système repérant les entités nommées dans des textes
  - ▶ Les EN sont de 3 sortes : organisation, personne, lieu
  - ▶ On calcule le rappel et la précision pour chacune de ces catégories
  - ▶ On veut calculer la moyenne du rappel et de la précision sur ces 3 classes !
  
- ▶ Pour n classes :

$$\frac{\sum_{i=1}^n \text{rappel}_i}{n} \quad (\text{idem pour la précision})$$

## *Augmenter le rappel et/ou la précision*

---

- ▶ Requête d'origine, ex : football français
- ▶ Relâchement de contrainte
  - ▶ ex. requête moins précise : football
  - ▶ ex. requête "synonyme" : foot en France
  - ▶ Tendance à augmenter le rappel, à diminuer la précision
- ▶ Resserrement de contrainte
  - ▶ ex : football français France
  - ▶ Tendance à diminuer le rappel, à augmenter la précision

# *Augmenter le rappel et/ou la précision*

---

- ▶ Dans un système représentant un document comme un sac de mots

(égalité des mots de la requête avec ceux du document)

- ▶ Comment augmenter le rappel et la précision ?

- ▶ En calculant une proximité entre mots

- ex : château ≈ châteaux

- ex : annulation ≈ annuler

- ▶ En lemmatisant les mots

- Soit les 2 textes suivants

- ▶ Les vols sont annulés en raison des grèves à Air France

- ▶ A cause de la grève, Air France a annulé le vol.

- ▶ 2 mots communs (air france)

- Version lemmatisée (sans les mots vides)

- ▶ <vol> <être> <annuler> <raison> <grève> <Air> <France>

- ▶ <cause> <grève> <Air> <France> <annuler> <vol>

- ▶ 5 mots communs

# *Augmenter le rappel et/ou la précision*

---

- ▶ Dans des systèmes utilisant le TAL
- ▶ Problèmes de la variation
  - ▶ morpho-syntaxique
    - ▶ en nombre, ex : journal, journaux
    - ▶ en genre, ex : petit, petite
    - ▶ conjugaisons, ex : juger, jugée, jugés, jugeront ...
    - ▶ etc.
  - ▶ Synonymie
    - ▶ ex : maladie héréditaire      maladie génétique
    - ▶ ex : magistrat                      juge
    - ▶ ex : législation                    réglementation
  - ▶ Dérivation
    - ▶ ex : france                      française
  - ▶ etc.
- ▶ Solution : relâcher les contraintes en admettant des variantes



# *Augmenter le rappel et/ou la précision*

---

## ▶ Problème de l'ambiguïté

- ▶ une même forme peut avoir plusieurs sens différent

- ▶ ex : vol

- Le vol a eu lieu dans une bijouterie parisienne.
- Il a embarqué sur le vol Paris–Londres.

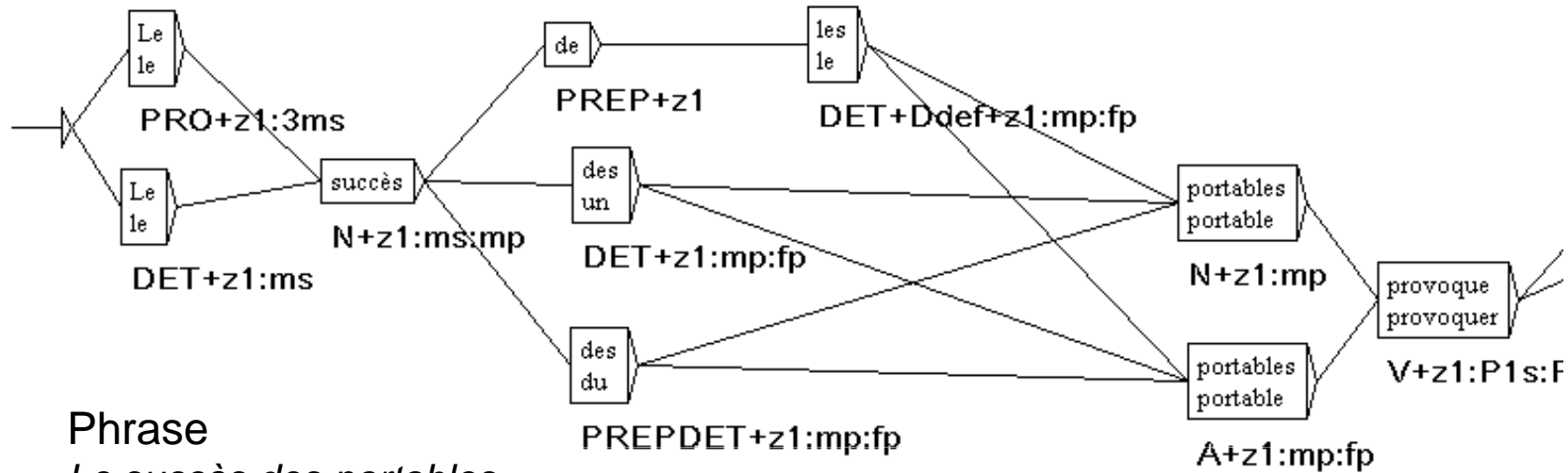
- ▶ ex : émission

- L'émission d'un nouveau timbre a eu lieu la semaine dernière.
- Cette émission de télévision a beaucoup de succès.

## ▶ Solution : resserrer la contrainte, en décrivant le contexte

- ▶ ex : rechercher "émission de télévision"

# Augmenter le rappel et/ou la précision



Phrase

*Le succès des portables  
provoque une saturation  
des réseaux*

Beaucoup de bruit dans  
toutes ces réponses !  
Rappel très bon  
Précision très mauvaise

