

Alain Joubert & Mathieu Lafourcade  
Université Montpellier 2 – France



# ENRICHISSEMENT DE LA LONGUE TRAINE D'UN RÉSEAU EXICAL GRACE À UN OUTIL D'ÉVALUATION

# UNE DOUBLE PROBLÉMATIQUE

**EVALUER QUALITATIVEMENT  
UN RÉSEAU LEXICAL**

**&**

**OBTENIR DES RELATIONS DE FAIBLE POIDS  
PERTINENTES  
(AKA LONGUE TRAINE)**

# ACQUISITION LEXICALE

## ⊙ Quelles informations ?

- Termes
- Inventaire de **sens** (dictionnaire) ou d'**usages** (locuteur)
- Relations entre termes et entre usages
  - ... vers des éléments de connaissances

## ⊙ Souvent...

- Long un terme => une personne + négociation
- Coûteux les payer
- Normatif peu fidèle aux utilisateurs
- Statique pas/peu d'évolution dans le temps

# ACQUISITION LEXICALE

## ... QUELLES RELATIONS ?

- ◎ Associations **libres** (idées reliées d'une façon ou d'une autre)
  - chat → animal, chien, souris, ...      vache → lait, meuh, corne, pis, ...
  - Tolstoï → Guerre et Paix, russe, ...      doubler → voiture, autoroute, ...
- ◎ Relations **ontologiques**
  - Hyperonymes – hyponymes – partie-de – tout - ...
- ◎ Relations **lexicales**
  - Syn – contraire – locutions – même famille – magn/antimagn ...
- ◎ **Rôles** sémantiques (**typiques** ou **possibles**)
  - Agent – patient – instrument – lieu – cause – conséquence – temps...
- ◎ Usages
  - sapin (arbre), sapin (sapin de Noël), sapin (bois)
  - sapin( cercueil), sapin (fiacre)

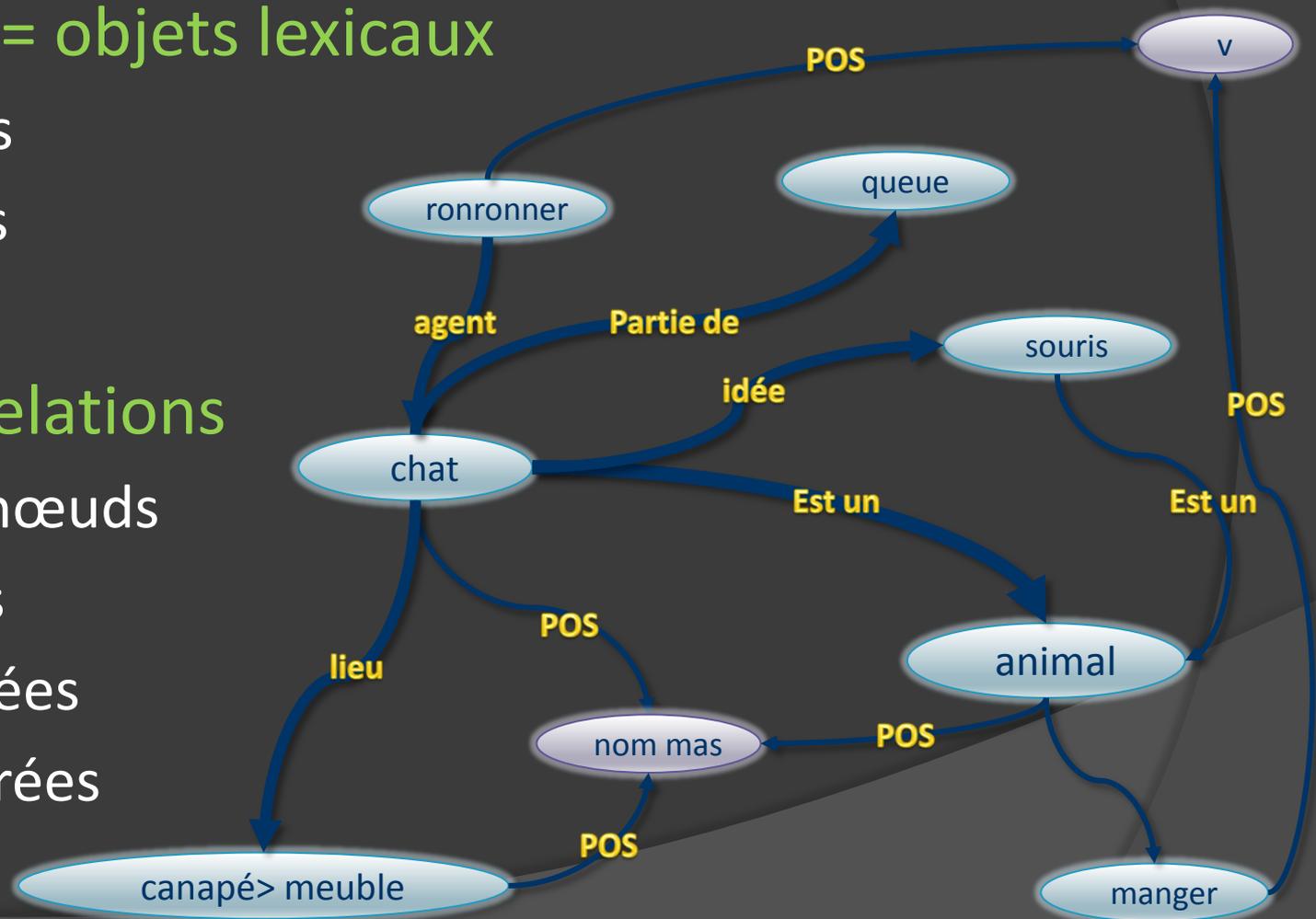
# ... SOUS LA FORME D'UN RÉSEAU LEXICAL

## ⊙ Nœuds = objets lexicaux

- Termes
- Usages

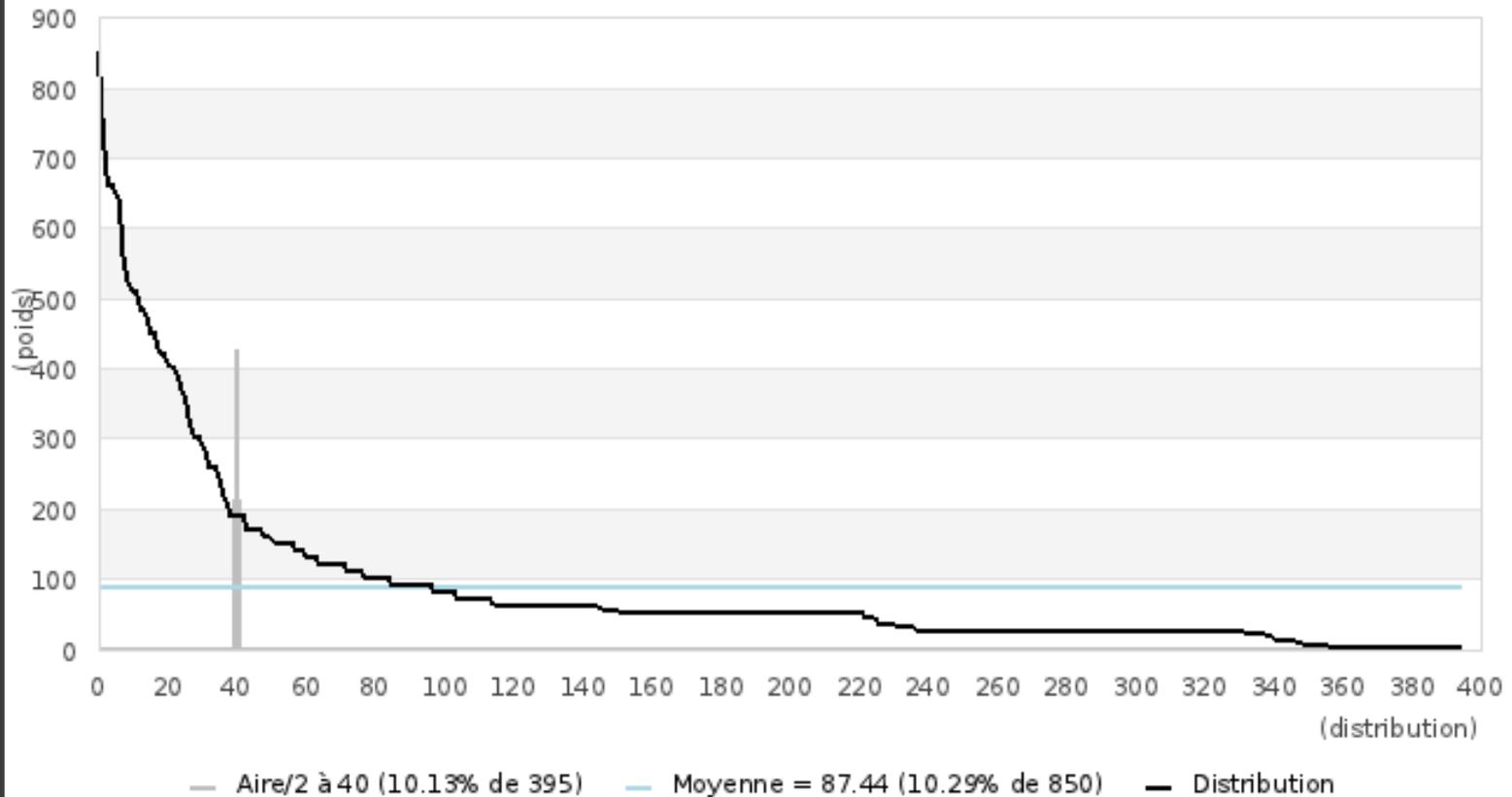
## ⊙ Arcs = relations

- Entre nœuds
- Typées
- Orientées
- Pondérées



# LONGUE TRAINE D'UN RÉSEAU LEXICAL ?

Terme 'chat' et relation all sortantes (395 données)



**GWAP**

**&**

**CONSENSUS POPULAIRE**

DONNER DES IDEES ASSOCIEES AU TERME QUI SUIT :

pomme

Temps  
48 s

30s

OK



Dernier terme proposé : tarte aux pommes

Ce terme a plusieurs sens ou il en manque ? Demandez de l'aide à vos amis

- tarte aux pommes
- < verger (agriculture)
- tomber dans les pommes
- verger >
- pommier
- < pépin (fruit)
- pépin >
- Blanche-Neige
- Golden
- < queue (fruit)
- quartier >
- queue >
- < fruit (botanique)
- fruit >

14/15 +

+10%

pommier • fruit

Apple • vert • quartier • vodka • charlotte • paradis • jus de fruit

apple • vert • quartier • vodka • charlotte • paradis • jus de fruit

pommier • fruit

Mots usagés

Inspiration

Inspiration

Mots usagés



Déconnexion

Niveau : 47

DONNER DES IDEES ASSOCIEES AU TERME QUI SUIT :

pomme

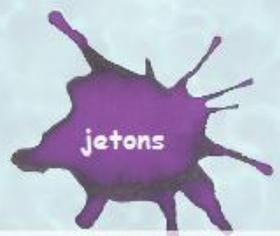
Réponses données par vous (kaput) pour cette partie (191503) : tarte aux pommes - verger (agriculture) - tomber dans les pommes - verger - pommier - pépin (fruit) - pépin - Blanche-Neige - Golden - queue (fruit) - quartier - queue - fruit (botanique) - fruit

Réponses données par le créateur (Amanyte) de cette partie (191503) : arrosoir - terre - bonne - tarte - pin - Golden - rouge - verte - api - fruit - compote - pommier - ommier

Golden - fruit - pommier

Vous gagnez 139 crédits et 4 point(s) d'honneur

Like Be the first of your friends to like this.

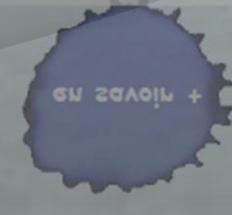
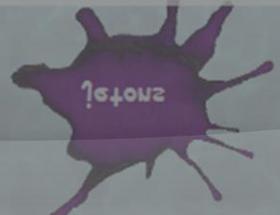


6.742 s

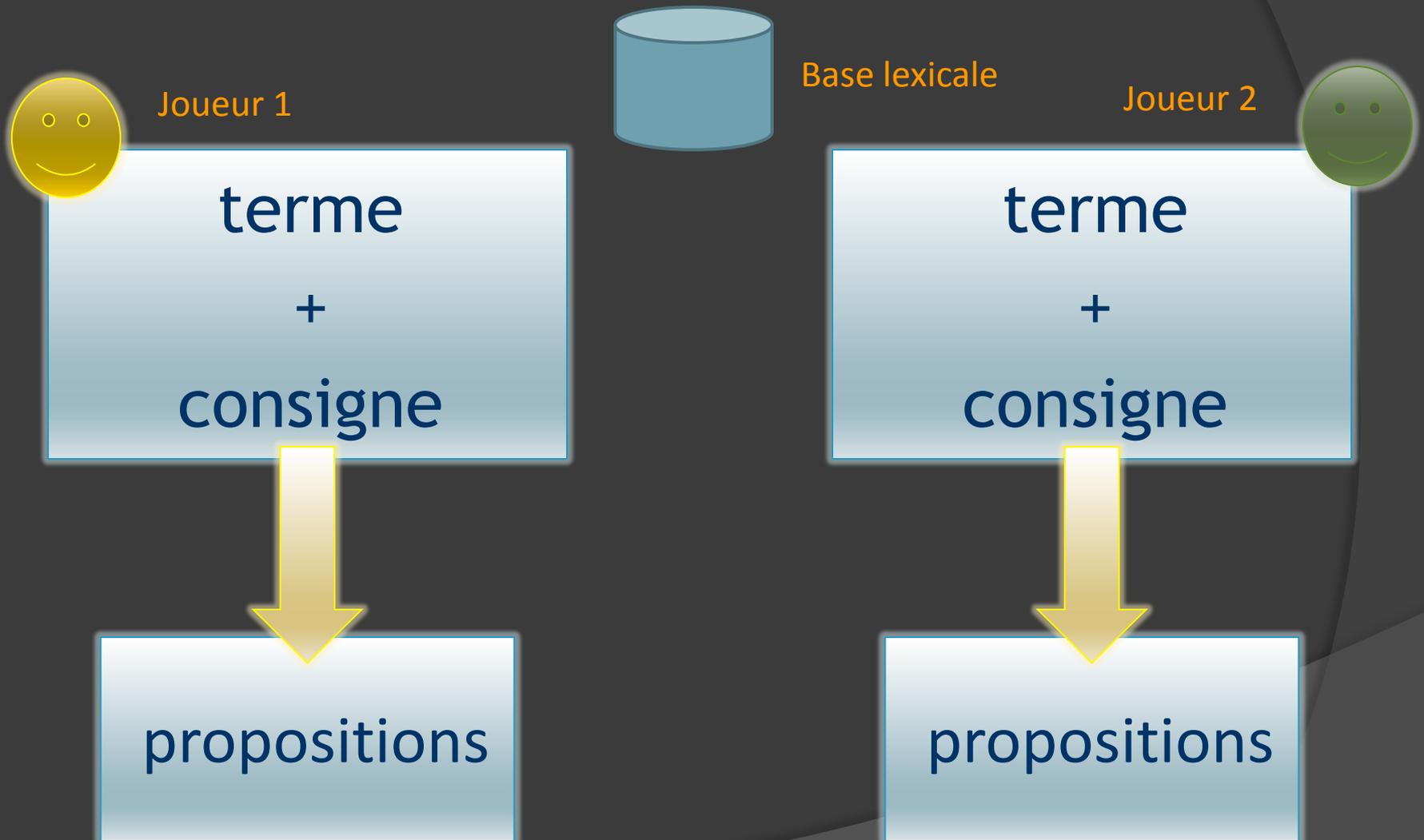
les joueurs 'Amanyte' et 'kaput' ont gagné 139 crédits avec 'pomme'.

2.547.2

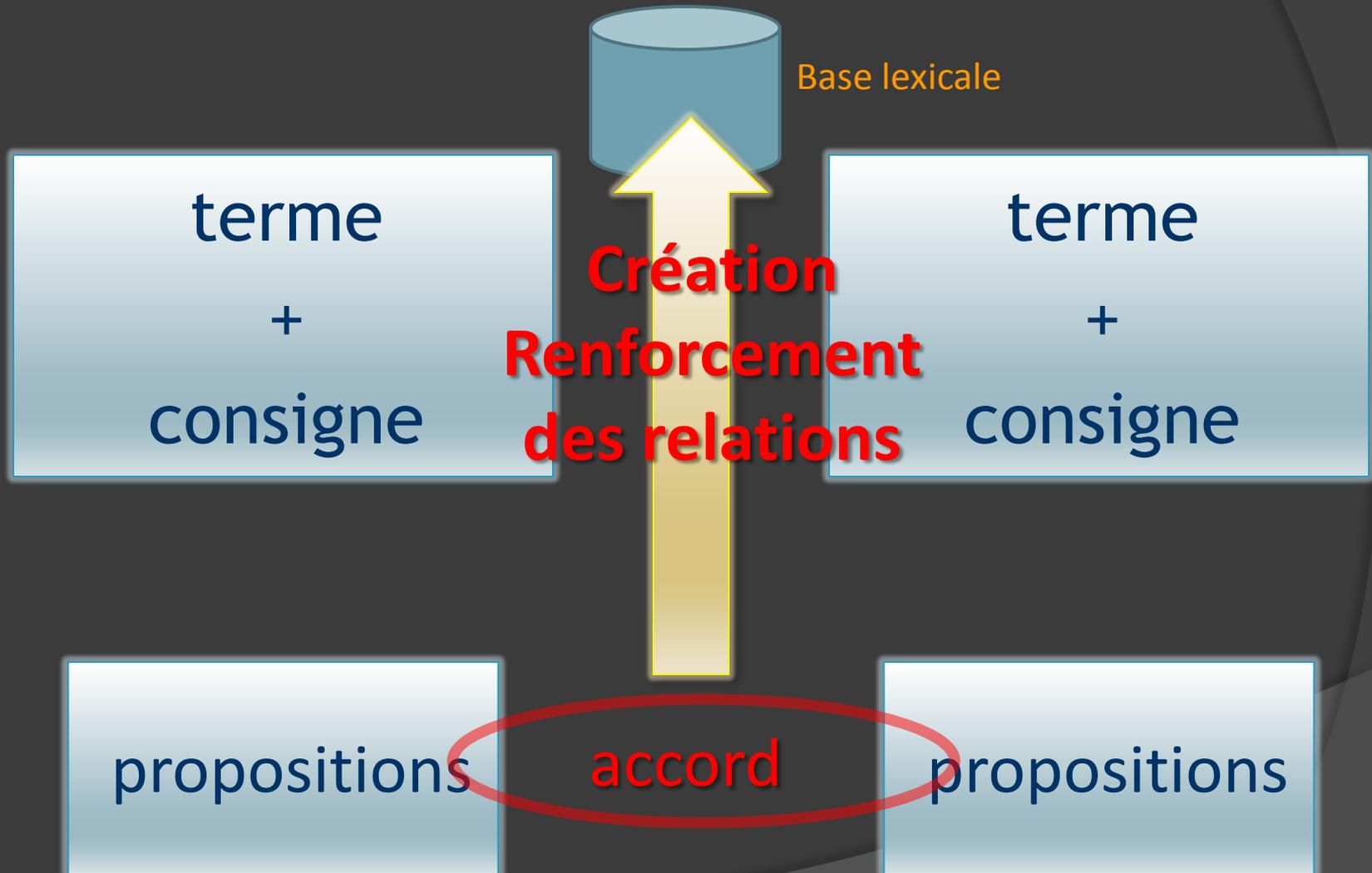
les joueurs 'Amanyte' et 'kaput' ont gagné 139 crédits avec 'pomme'.



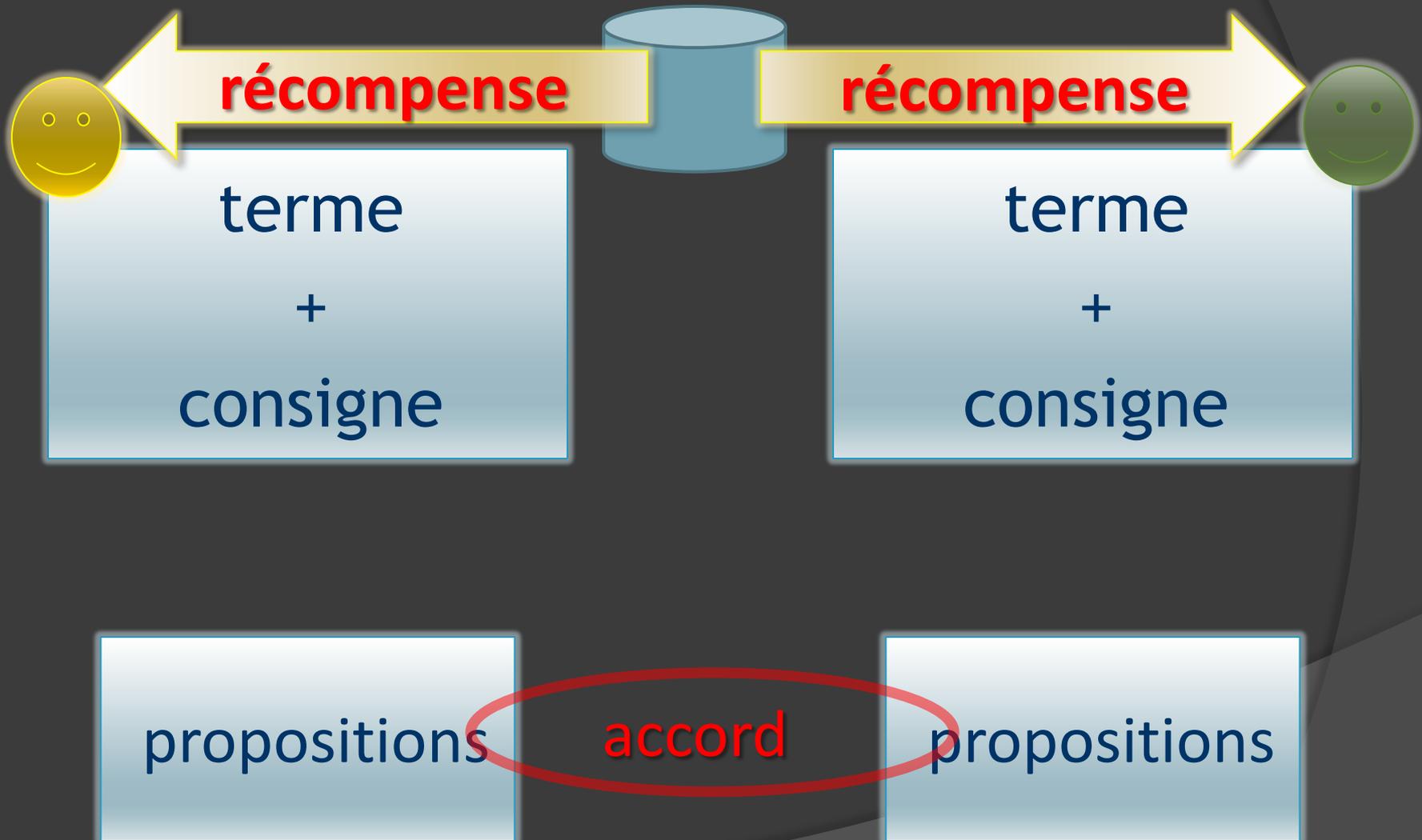
# MODÈLE D'INTERACTION



# MODÈLE D'INTERACTION



# MODÈLE D'INTERACTION



# RÉGULATION

## Filtrage - accord par paires

- Consensus minimal itéré (pondération)
- Minimiser le bruit, maximiser le rappel (longue traine)

## Caractéristiques

- Mot tiré au hasard
- Partenaire inconnu durant la partie
- Partie asynchrone

## Points

- + si relation peu pondérée
- - si relation très pondérée
- Relations taboues

# QUELQUES CHIFFRES

- ⊙ Environ **1 200 000** de parties ont été jouées depuis le début de JeuxDeMots (septembre 2007)  
(> **1 700 000** parties : JeuxDeMots + PtiClic + Askit + AKI)
- ⊙ Environ **18000h** de temps de jeu cumulé (moyenne de 1 minute/partie) : **750** jours
- ⊙ ~ **3000** joueurs inscrits  
temps moyen de rotation = 3 semaines
- ⊙ Coût estimé = **75 k euro**
  - Estimation WordNet (sur 25 ans) =  $300 \text{ k} * 25 = 7,5 \text{ M euro}$
  - Rapport de 100

**EVALUATION**  
**DU RÉSEAU LEXICAL**  
**&**  
**RENFORCEMENT DE LA LONGUE TRAINE**

# EVALUATION QUANTITATIVE

- ◎ **240 000 termes** dans la base.  
109 851 ont au moins une relation de type idée associée.
- ◎ **1 375 000 relations lexicales** dans la base.  
153 000 termes ont au moins une relation sortante ( $A \rightarrow B$ ).  
92 000 termes ont au moins une relation entrante ( $A \leftarrow B$ ).
- ◎ **212 000 termes** renseignés avec un **POS**.
- ◎ **3850 termes raffinés** et **13000 usages**.
- ◎ **4200 termes** étiquetés comme polysémiques (couverture 92 %).

# EVALUATION QUALITATIVE

- ⦿ Comment évaluer un réseau lexical ?

Pas de référence (**étalon** or inexistant)... sinon on aurait déjà la ressource que l'on cherche à construire

- ⦿ Quels termes vérifier ?

Échantillonnage ?

Trouver les termes **intéressants**

- ⦿ Evaluation positive

Vérifier si les **bonnes** relations y sont

- ⦿ Evaluation négative

Débusquer les **mauvaises** relations

# MAIS COMMENT ÉVALUER CELA ?

## chat

Adjectif masculin singulier, Nom, Nom masculin singulier

**Lemme** chat **Informations diverses** terme polysémique, vivant, lieu, concret, objet naturel, terme général

**Raffinements sem** > chat (félin), chat (discussion), chat (jeu)

**Gloses** > animal, chat perché, discussion, félin, jeu

**Idées associées** > 93 affectueux, aiguille, angora, **animal**, animal de compagnie, animal domestique, animale, autre, caisse, calin, caresse, caresser, chat sauvage, chat (félin), chat (discussion), chat (jeu), **chaton**, chatte, chatte (chat), **chien**, chien (animal), compagnie, coussinet, croquette, devinette, discussion, domestique, dormir, doux, ennemi, être vivant, félidé, **félin**, Felix, Félix, fourbe, gamelle, gouttière, **griffe**, griffe (ongle), griffes, gris, jeu, jeux, lait, langue, litière, maison, mammifère, mammifères, matou, miaou, miaulement, miauler, mignon, minet, minet (chat), minou, mou, moustache, moustaches, MSN, noir, oiseau, pâtée, **patte**, pattes, pelage, pelote, perché, persan, **poil**, poils, poilu, puce, puce (insecte parasite), **queue**, queue (queue d'animal), rat, rat (rongeur), ronronnement, **ronronner**, siamois, siamois (chat), sieste, **souris**, souris (rongeur), sphynx (chat), tigre, vibrisse, virtuelle, yeux, zoologie < 167 :-), acheter chat en poche, aiguille, angora, angora (chat), **animal**, animal de compagnie, animal domestique, animaux, animaux de compagnie, animaux de la ferme, arbre à chat, Azraël, bête, bête (animal), boîte à meuh, botte, botté, caniche, caracal, caresse, **caresser**, cataire, chartreuse, Chat, Chat botté, chat botté, chat de gouttière, chat domestique, Chat sauvage, **chat sauvage**, chat-tigre, chat (félin), chat (discussion), chat (jeu), **chaton**, chaton (chat), **chatroulette**, chats, **chatte**, chatte (chat), **chatter**, chatterie, **chien**, chien (animal), compagnie, coussinet, coussinet (patte), croquette, croquette (nourriture pour animaux), crotte, Curl, désensibilisation, devinette, discussion, domestique, donner sa langue au chat, dressage, Egypte, émoticône, émoticône, fauve (félin), **félidé**, **félin**, Felix, Félix, Félix Ier, fouetter, fourrure, Friskies, Gaston Lagaffe, gouttière, greffier, greffier (chat), **griffe**, griffe (ongle), **griffer**, griffes, gris, Gros Minet, Grosminet, haret, Hercule, IRC, Je suis un chat, jeu, jeu d'enfant, Katz, lait, lait (produit laitier), langue, langue de chat, lapin blanc, Le Maître chat ou le Chat botté, **les animaux**, Les Musiciens de Brème, litière, lynx, Ma Dalton, mammifère, manul, matou, Maurice et Patapon, mdr, messagerie instantanée, messenger, miaou, miaulement, miauler, **minet**, minet (chat), **minou**, mistigri (chat), mou, mou (poumon), moustache, **MSN**, msn, Natsume Soseki, noir, nourriture pour animaux, oeil de chat, oiseau, orientale, pâtée, **patte**, pattes, pelage, persan, persan (chat), Pierre et le Loup, **poil**, puce, puce (insecte parasite), puces, **quadrupède**, Quadrupèdes, queue, queue (queue d'animal), rat, rat (rongeur), répulsif, retomber sur ses pieds, rétractible, Ronron, ronronnement, ronronner, serval, siamois, siamois (chat), sortir ses griffes, **SOURIS**, souris (rongeur), **sphinx**, sphynx (chat), Sylvester, tapetum, tchat, Teal'c, **tigre**, tigre (félidé), **titi**, Tom et Jerry, **Vétérinaire**, vétérinaire, vibrisse

**Thèmes/domaines** > 21 Anatomie, animal, Biologie, botanique, carnivore, cinéma, cuisine, Elevage, Histoire, ichtyologie, jeu, littérature, mammifère, marine, mécanique, minéralogie, Physiologie, poisson, Vétérinaire, Zoologie, zoologie

**Synonymes** > 8 chaton, dialogue, matou, minet, minou, mistigri, MSN, tchat **Contraires** > chien

**Hyperonymes** > 24 **animal**, animal de compagnie, animal domestique, bête, carnivore, \* cerf, chasseur, compagnon, être vivant, félidé, félidés, **félin**, ingrédient de cuisine, lieu, mamifère, mammifère, \* mulot, \* ongulé, persan, poil, quadrupède, vertébré, vivant, \* zébu **Hyponymes** > 49 abyssin, angora, angora (chat), angorra, Azraël, birman, \* campagnol, chartreux, chat angora, chat botté, chat de gouttière, chat de Pallas, **chat domestique**, chat fourré, chat marsupial, chat noir, chat perché, chat persan, **chat sauvage**, chat sauvage d'Europe, chat sauvage d'Europe, chat serval, chat siamois, chat tigre, chat tigré, **chaton**, \* chaton (botanique), chatte, chatte (chat), chien, domestique, **félin**, gouttière, greffier, greffier (chat), gris, Hercule, matou, minet, minou, persan, persan (chat), queue, sauvage, siamois, siamois (chat), souris, sphynx (chat), **tigré**

**partie de chat** > 30 bouche, chaton, corps, coussinet, coussinet (patte), dent, dents, **griffe**, griffes, gueule, langue, moustache, moustaches, museau, nez, oeil, oreille, oreilles, organes, **patte**, pattes, peau, pelage, **poil**, poils, queue, queue (queue d'animal), tête, vibrisse, yeux

**Plus intense que chat** > félin, lion, mammifère, tigre

**Termes de la même famille** > chaton **Locutions/termes composés** > 25 à bon chat bon rat, chat à neuf queues, chat angora, chat botté, chat de gouttière, chat de mer, chat de Pallas, chat domestique, chat échaudé, chat en eau froide, chat en poche, chat fourré, chat marin, chat marsupial, chat noir, chat perché, chat persan, chat sauvage, chat sauvage d'Europe, chat sauvage d'Europe, chat serval, chat siamois, chat tigre, chat tigré, langue au chat, langue de chat

**Caractéristiques de chat** > 14 agile, blanc, câlin, crevé (mort), doux, félin, indépendant, joueur, mignon, noir, persan, persan (chat), sauvage, tigré

**Où se trouve chat ?** > 8 arbre, arbre (végétal), canapé, coussin, cuisine, jardin, lit, maison **Que trouve-t-on dans/sur chat ?** > 11 dos, griffe, langue, moustaches, oreille, patte, poil, queue, souris, tête, yeux

**Que peut-on faire dans/sur chat ?** > caresser, discuter, nourrir, ronronner

# ... ou pire, cela ?

chat' (id=150 => chat ; niveau = 47.5987 ; luminosité = 3286)

346 relations ==>

- chat ---r\_associated#0:790--> félin
- chat ---r\_isa#6:780--> félin
- chat ---r\_associated#0:680--> chien
- chat ---r\_isa#6:641--> animal
- chat ---r\_associated#0:640--> animal
- chat ---r\_associated#0:640--> souris
- chat ---r\_has\_part#9:630--> queue
- chat ---r\_associated#0:560--> miauler
- chat ---r\_has\_part#9:520--> poil
- chat ---r\_associated#0:500--> chatte
- chat ---r\_associated#0:490--> chaton
- chat ---r\_agent-1#24:480--> dormir
- chat ---r\_has\_part#9:461--> moustache
- chat ---r\_flpot#12:458--> \_FL:9
- chat ---r\_associated#0:430--> griffe
- chat ---r\_agent-1#24:410--> miauler
- chat ---r\_associated#0:400--> ronronner
- chat ---r\_associated#0:390--> minou
- chat ---r\_agent-1#24:390--> chasser
- chat ---r\_flpot#12:383--> \_FL:0
- chat ---r\_associated#0:380--> moustache
- chat ---r\_has\_part#9:380--> patte
- chat ---r\_flpot#12:363--> \_FL:6
- chat ---r\_has\_part#9:360--> griffe
- chat ---r\_agent-1#24:360--> manger
- chat ---r\_agent-1#24:340--> ronronner
- chat ---r\_associated#0:300--> poil
- chat ---r\_flpot#12:292--> \_FL:24
- chat ---r\_hypo#8:290--> siamois
- chat ---r\_associated#0:280--> queue
- chat ---r\_has\_part#9:280--> pattes
- chat ---r\_has\_part#9:270--> oreille
- chat ---r\_isa#6:260--> mammifère
- chat ---r\_has\_part#9:250--> yeux
- chat ---r\_infopot#36:240--> \_INFO-NO-MORE-QUESTION
- chat ---r\_associated#0:210--> animal de compagnie
- chat ---r\_associated#0:210--> mammifère
- chat ---r\_has\_part#9:200--> griffes
- chat ---r\_associated#0:190--> patte
- chat ---r\_isa#6:190--> félidé
- chat ---r\_hypo#8:190--> grince

365 relations <==

- félin ---r\_associated#0:1050--> chat
- MSN ---r\_associated#0:910--> chat
- chien ---r\_associated#0:780--> chat
- être vivant ---r\_hypo#8:620--> chat
- chat sauvage ---r\_isa#6:480--> chat
- animal de compagnie ---r\_hypo#8:440--> chat
- patte ---r\_associated#0:430--> chat
- chat domestique ---r\_isa#6:420--> chat
- animal ---r\_associated#0:400--> chat
- dormir ---r\_agent#13:400--> chat
- chaton ---r\_family#22:370--> chat
- souris ---r\_associated#0:360--> chat
- tigre ---r\_associated#0:360--> chat
- caresser ---r\_associated#0:350--> chat
- chatter ---r\_associated#0:330--> chat
- minou ---r\_associated#0:330--> chat
- chat sauvage ---r\_associated#0:310--> chat
- mammifère ---r\_associated#0:270--> chat
- canapé ---r\_lieu-1#28:260--> chat
- chaton ---r\_associated#0:250--> chat
- animal ---r\_hypo#8:240--> chat
- mammifère ---r\_hypo#8:240--> chat
- griffe ---r\_lieu#15:220--> chat
- roupiller ---r\_agent#13:220--> chat
- sphinx ---r\_associated#0:220--> chat
- se laver ---r\_agent#13:200--> chat
- griffer ---r\_associated#0:190--> chat
- minet ---r\_associated#0:190--> chat
- chatte ---r\_associated#0:180--> chat
- félin ---r\_hypo#8:180--> chat
- les animaux ---r\_associated#0:180--> chat
- Qui pourrait renifler une odeur ? ---r\_agent#13:170--> chat
- félidé ---r\_hypo#8:170--> chat
- griffe ---r\_associated#0:170--> chat
- titi ---r\_associated#0:170--> chat
- poil ---r\_associated#0:160--> chat
- Vétérinaire ---r\_associated#0:150--> chat
- animal de compagnie ---r\_associated#0:150--> chat
- dévorer ---r\_agent#13:150--> chat
- faire sa toilette ---r\_agent#13:150--> chat
- félidé ---r\_associated#0:150--> chat

# EVALUATION QUALITATIVE

## Approches indirectes via une application

- Jeux de **devinettes** : indices → terme

  - Utilisateur** = donneur d'indices (+ relation optionnelle)

  - AKI** = devineur (+ apprentissage + mémoire à court terme)

- Utilisation du réseau lexical

  - Intersection des listes de termes liés

    - Moyenne géométrique** sur les poids des relations

  - Rattrapage sur deux coups

    - Somme booléenne (chaque relation a un poids de 1)

# AKI – un jeu de devinette tout bête

Soumettre

l'indice

Après 5 indice(s), il s'agit peut-être de :

**médoc** (vin)

C'est la bonne réponse !

[abandon]

si ce n'est pas ça vous pouvez proposer un nouvel indice...

Vos indices	Mes propositions
vin rouge Bordeaux cabernet sauvignon AOC	bouteille blanc vin rouge Médoc médoc (vin)

# AKI – un jeu de devinette tout bête

Soumettre l'indice

Après 3 indices, il s'agit sûrement de :

## maladie de Parkinson

C'est la bonne réponse !

si ce n'est pas ça vous pouvez proposer un nouvel indice...

Vos indices	Mes propositions
maladie vieillesse tremblement	grippe mort maladie de Parkinson

Après 4 indices, il s'agit sûrement de :

## cothurne

C'est la bonne réponse !

si ce n'est pas ça vous pouvez proposer un nouvel indice...

Vos indices	Mes propositions
chaussure théâtre antiquité montante	pied clown lieu cothurne

Après 4 indice(s), je suis perdu, désolé

Il s'agissait de

(faites attention aux accents et aux majuscules/n)

Vos indices	Mes propositions
chemin de fer voie de garage wagon wagon de marchandise	rail sous-station électrique transport ferroviaire

# HYPOTHÈSE

La performance de l'outil

reflète

la qualité du réseau lexical

# PRINCIPE DE L'ALGORITHME

Processus itératif

= sélection du plus fort poids

⊙ 1<sup>er</sup> indice : terme  $i_1$

Signature :  $S(i_1) = S_1 = t_1, t_2 \dots$  (termes reliés à  $i_1$ )

Proposition :  $p_1 = t_1 = \max(S_1)$

Signature courante :  $S'_1 = S_1 - (p_1, i_1)$

⊙ 2<sup>ème</sup> indice : terme  $i_2$

Signature :  $S_2 = (S'_1 \cap S(i_2)) - i_2$

Proposition :  $p_2 = \max(S_2)$

Signature courante :  $S'_2 = S_2 - p_2$

# PRINCIPE DE L'ALGORITHME

## Processus itératif

- $n^{\text{ème}}$  indice : terme  $i_n$

$$\text{Signature} : S_n = (S'_{n-1} \cap S(i_n)) - i_n$$

$$\text{Proposition} : p_n = \max(S_n)$$

$$\text{Signature courante} : S'_n = S_n - p_n$$

- Arrêt de l'algorithme si  $p_n =$  terme recherché

- Procédure de rattrapage = plus forte majorité

$$\text{Si } S_n = \emptyset$$

$$\text{Alors } S_n = (S'_{n-1} \oplus S(i_n)) - i_n$$

## ⦿ Parties avec **mots tabous**

- Certains **indices** sont interdits
- **Indices** = termes fortement reliés dans le réseau

Votre défi est de faire deviner à AKI le  
terme :

**bathyscaphe**

mots tabou : bathyscaphe - sous-marin -  
submersible - principe d'Archimède - fosse des  
Mariannes - bateau - fonds marins

Vous êtes en mode tabou !

## ⦿ Parties avec mots tabous

- Certains **indices** sont interdits
- **Indices** = termes fortement reliés dans le réseau

AKI le

Votre défi est de faire deviner à AKI le

**Force les joueurs à chercher  
des indices plus éloignés**

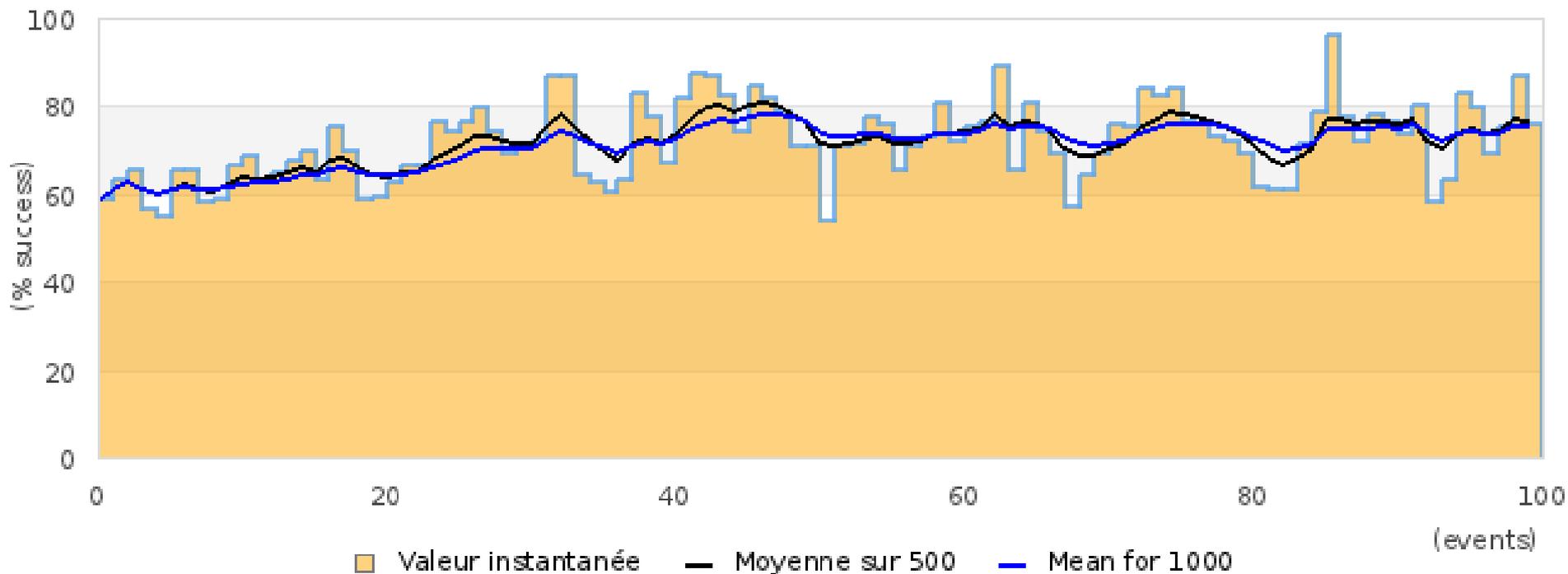
mots tabou : bathyscaphe - sous-marin -  
submersible - principe d'Archimède - fosse des  
Mariannes - bateau - fonds marins

**Renseigne la longue traine**

Vous êtes en mode tabou !

# RÉSULTATS POUR DU TOUT VENANT

AKI Data (13620 data with chunk size of 136)



**AKI ~ 75%**

Test sur > 12 000 parties dont les termes sont choisis par les utilisateurs

**Utilisateur ~ 46%**

Test sur environ 200 entrées et les termes tabous de AKI donnés comme indices

# QUELQUES CHIFFRES & ANALYSE

- > 20 000 parties de AKI
- > 2000 termes introduits
- > 50 000 relations introduites pour la longue traine
- ◎ Echantillonnage
  - Termes choisis par les joueurs
  - Termes courants sont surévalués (1/3 des parties)
- ◎ Test sévère
  - Les joueurs visent à mettre en défaut le système
  - 25 % des parties en tabou
    - Plus difficile pour AKI de trouver
    - Permet d'acquérir de nouvelles associations

# CONCLUSION

## ◎ Approche

- AKI est volontairement simpliste
  - Évaluer les données (et non pas l'algorithme) - **Hypothèse**
- Evaluation et acquérir des données
  - Au long cours (permanente)
  - À grande échelle (vocabulaire ouvert, bcp d'évalauteurs divers)
  - Dynamiquement (le réseau évolue dans le temps)

## ◎ Données

- Evolution des performances de 60% à environ >75% en 6 mois

## ◎ Question ouverte

- **Asymptote de la performance? 80 % - 90 % - 95 % ?**

# CONCLUSION

ON PEUT **CONSTRUIRE** UN RÉSEAU LEXICAL  
AVEC DES JEUX

&

ON PEUT AUSSI **ÉVALUER** UN RÉSEAU LEXICAL  
AVEC DES JEUX

&

ON PEUT **ACQUÉRIR** DES RELATIONS DE LA **LONGUE TRAINE**  
AVEC DES JOUEURS

MERCI