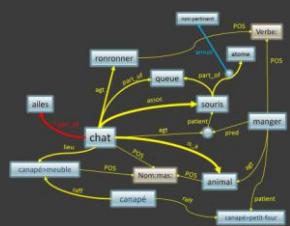


Mathieu LAFOURCADE  
TEXTE team, LIRMM, U. of Montpellier, France



## GAMES WITH A PURPOSE

THE Jeux de mots PROJECT

January 2017



Attached to Montpellier University and the French National Center for Scientific Research (CNRS). Its activity develops through three scientific research departments (**Informatics**, Microelectronics, Robotics) and 19 teams.

**TEXTE team** : Exploration et exploitation de données textuelles – 11 people

Syntax, Textual Semantic, Lexical Semantics, Algebraic Models, Vector Models, Dialog Models

# What are GWAPs?

- Games (meant to be funny, addictive, pleasant...)
- Designed for
  - Data acquisition
  - Problem solving
- Dubbed **collective intelligence**
- Core assumption

A large number of ordinary people  
is more efficient than  
a small number of specialists

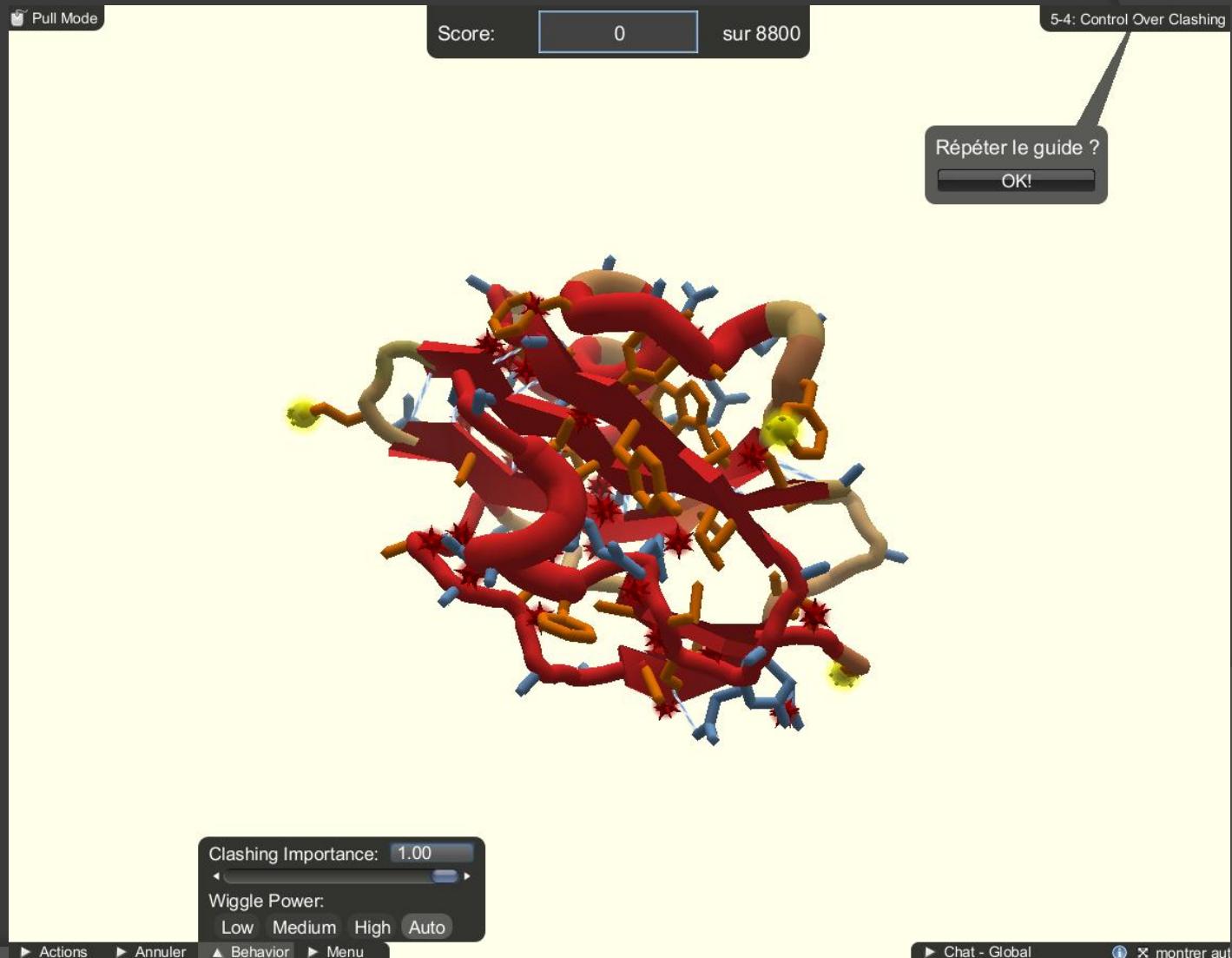
# Amazon Mechanical Turk?



- Online crowdsourcing, Microworking
- Legal issues
  - Piece work is not legal in many countries
- Ethical issues
  - Some people try to live from their work for AMT
- Quality issues:
  - Very poor quality (people maximize number of microtasks done)
  - Requires effort and money to check data
    - Not so economical in the end after all...
- see « Amazon Mechanical Turk: Gold Mine or Coal Mine? »  
by Karen Fort, Gilles Adda, K. Bretonnel Cohen

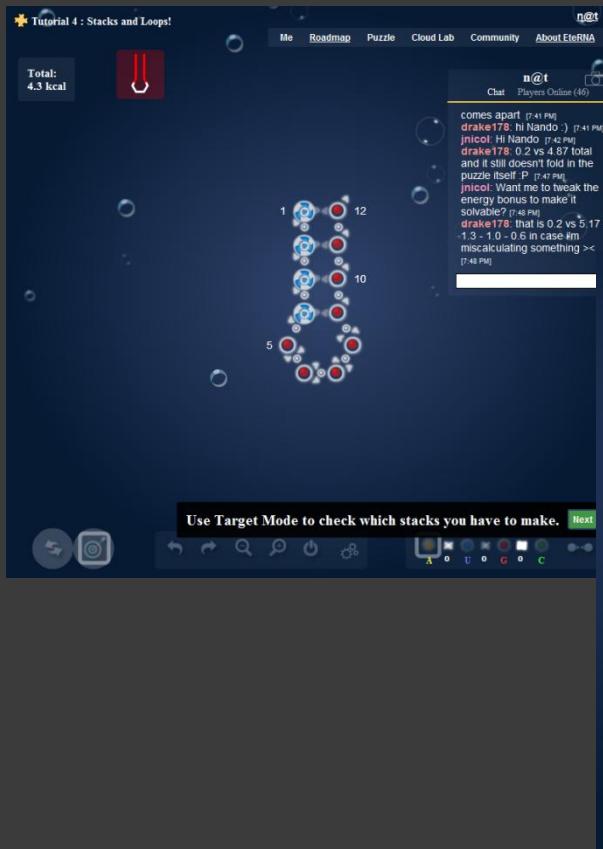
# Some GWAPs in Biology

## ○ Foldit



# Some GWAPs in Biology

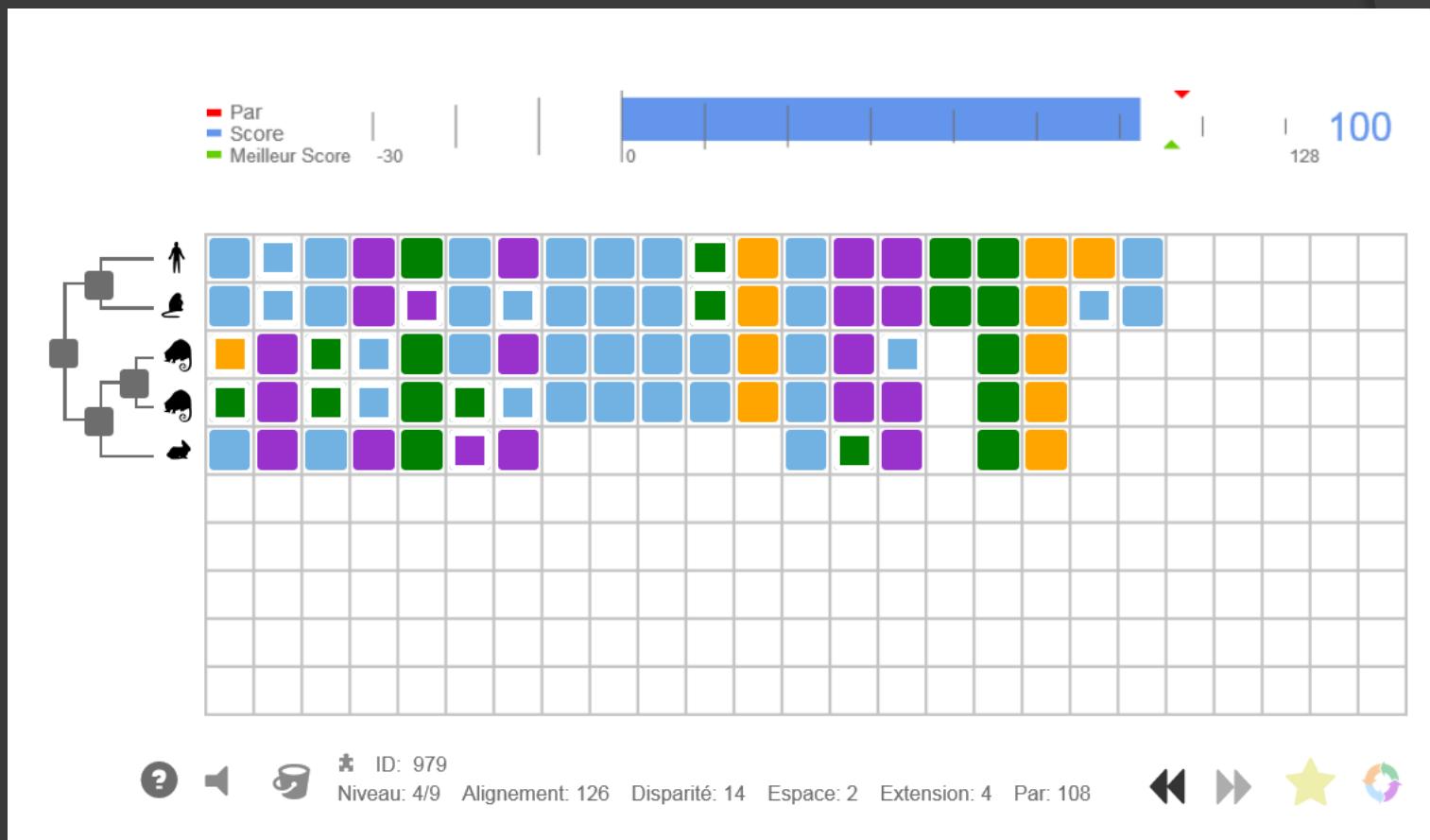
## ● Eterna



The screenshot shows a vertical RNA structure with numbered nucleotides (1, 5, 10, 12). The total energy is 0 kcal. A chat window shows a conversation between players n@t and jnicol about tweaking energy bonuses. The message from jnicol reads: "Want me to tweak the energy bonus to make it solvable?" and "that is 0.2 vs 5.17 + -1.3 + 1.0 - 0.6 in case I'm miscalculating something ><". The message from drake178 reads: "this is just the top stack though, my solution is not good enough to solve both, it was just a proof of concept that the top will not solve anyway." The player n@t is currently active.

# Some GWAPs in Biology

## ○ Phylo



# Some GWAPs in Biology

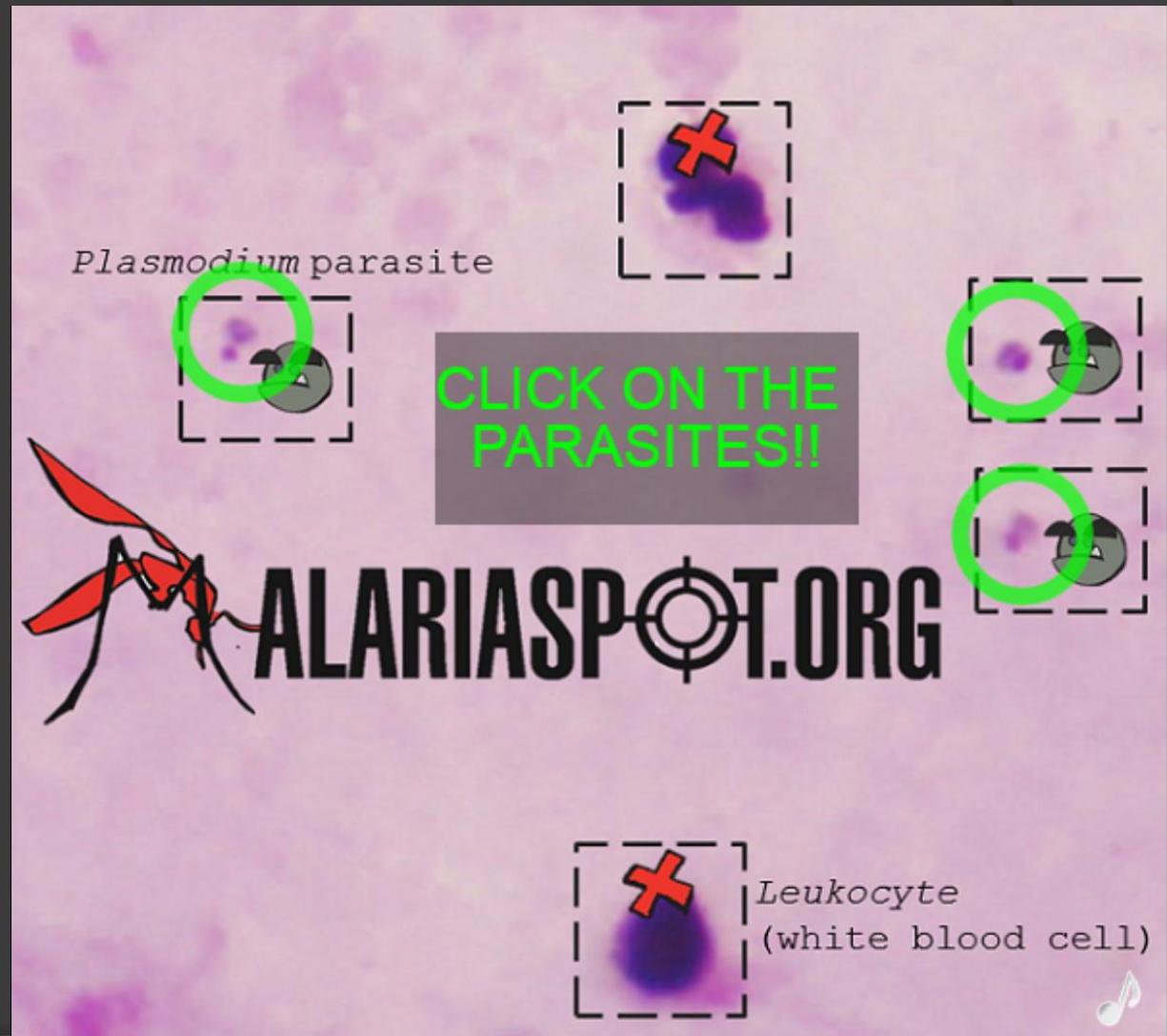
## ⦿ Nightjar



I give up

# Some GWAPs in Medicine

- Malaria Spot



# Some GWAPs in Arts

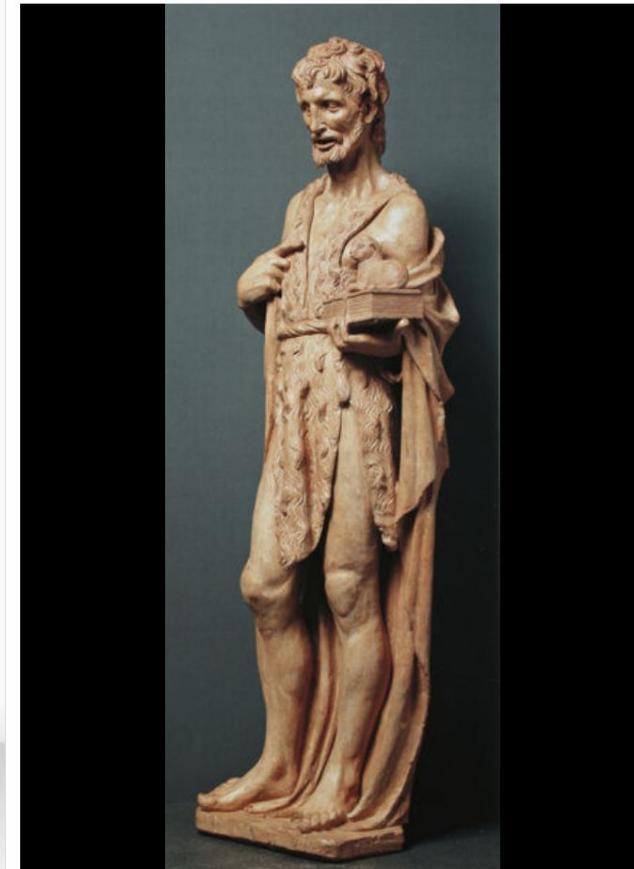
## ○ Artigo

MENTIONS LÉGALES PROFIL DÉCONNEXION R

# ARTigo

À PROPOS D'ARTIGO  
BLOG / /   
[AU TABLEAU D'HONNEUR](#)

[JOUER À NOUVEAU](#)  
 [AU TABLEAU D'HONNEUR](#)



Artiste  
Giovanni de Fonduli  
Titre  
Johannes der Täufer  
Lieu  
Turin / Privatsammlung  
Datation  
1470/1480  
Score  
30  
Mots-clés déjà données  
[STATUE](#)  
[HOMME](#)  
[SCULPTURE](#)  
[LIVREE](#)  
[BARBE](#)

# Some GWAPs in NLP

## ○ Wordrobe

The screenshot shows a game interface for 'wordrobe'. At the top, there's a navigation bar with a user icon, 'Welcome natalice / Account / Sign out', and links for 'play', 'games', 'scores', and 'about'. Below the header, a purple circular icon with a person symbol and the text 'Others' is followed by the message 'Questions left until drawer is completed: 8'. A question text is displayed: 'The tape did not show the killing, and has not been authenticated. The diplomat was kidnapped off a Baghdad street on Saturday as he stopped to buy a newspaper. Meanwhile, Iraqi **police** say twin car **bombs** have killed at least 13 **people** and wounded nearly 30 **others** south of Baghdad in the town of Mashruh.' Below the text is a list of four options with radio buttons: 1) police, 2) bombs, 3) people (which is selected), and 4) Not in the text. At the bottom, there's a slider labeled 'Place your bet: low' on the left and 'high' on the right, with a blue track and a silver slider handle. To the right of the slider are two buttons: 'answer' and 'skip'.

Others Questions left until drawer is completed: 8

The tape did not show the killing, and has not been authenticated. The diplomat was kidnapped off a Baghdad street on Saturday as he stopped to buy a newspaper. Meanwhile, Iraqi **police** say twin car **bombs** have killed at least 13 **people** and wounded nearly 30 **others** south of Baghdad in the town of Mashruh.

1) police  
2) bombs  
3) people  
4) Not in the text

Place your bet: low  high

answer skip

# Some GWAPs in NLP

## ○ Zombilingo

Il faut trouver le sujet du verbe surligné. Si le sujet est un groupe nominal, c'est le mot principal qu'il faut sélectionner.

- "Une première réunion a eu lieu sur la question de l'indemnisation."
- "Aucune adaptation de la dose n'est nécessaire"
- "Le juge Desmure se déclare alors incomptént pour instruire des faits visant le chef de l'État."

Note : pour une personne, le mot principal est 1- son titre (Monsieur, Président, etc) 2- son prénom (et non son nom)

- "Phileas Fogg vivait seul dans sa maison de Saville-row."

Quant à Mr. Fogg, il reposa aussi paisiblement que s'il eût été dans sa tranquille maison de Saville-row.

59

# GWAPs... some properties

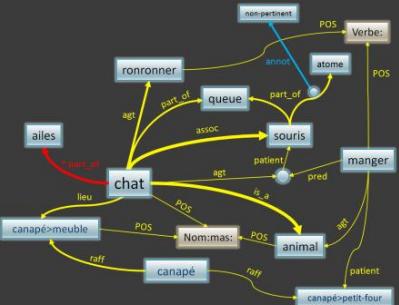
- A good player  $\Leftrightarrow$  good data
- Beware of **various biases**
- Difficult to be **funny AND efficient**
- In general, short life span (many gwaps are dead before long)
- Often the expected results are overestimated

# CONTEXT OF THE JEUXDEMOTS PROJECT

SEMANTIC ANALYSIS OF TEXTS  
WITH PROPAGATION ALGORITHMS  
ON/WITH  
A LEXICO-SEMANTIC NETWORK

=

A KNOWLEDGE (DATA)BASE,  
A GRAPH  
WITH WORDS,  
SENSES/USAGES/RELATIONS



# ACQUISITION OF LEXICAL, TERMINOLOGICAL, ONTOLOGICAL INFORMATION ...

## ● What for ?

- **applications** needing lexical, common sens and specialized field knowledge
  - Report analysis in medical imaging (**Imaios**)
  - Offer/demand matching in tourism (**Bedycasa**)
  - Debate management (**SucceedTogether**)
  - Class factorization in software eng. (**Orange**, **Berger Levraut**)

## ● How ?

- Automatically (extracting for corpora) ?  
knowledge is not always explicitly present in texts  
not exclusively, not totally – a lot of implicit knowledge
- By hand?              Long – (too) costly – normative – static data

# ... AS A LEXICAL NETWORK

## Nodes

Terms, textual segment, NP

Usages, concepts

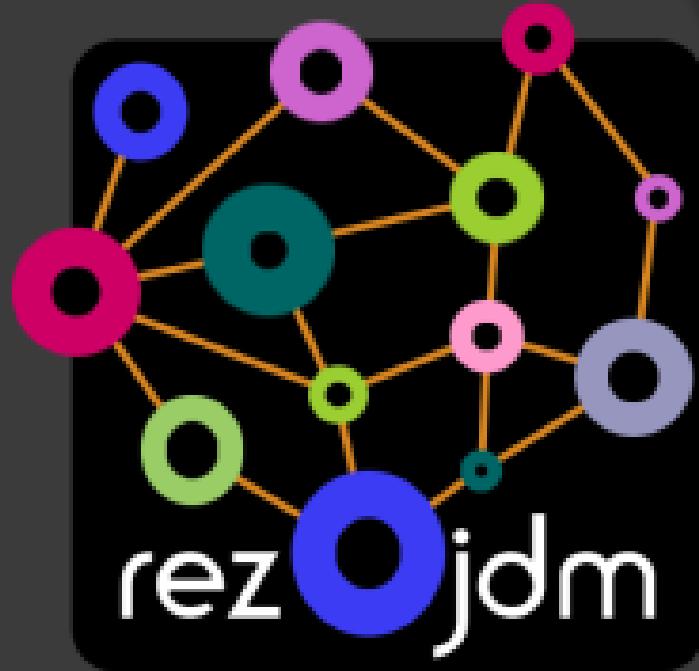
Various symbolic informations

## Relations

Typed

Directed

Weighted



free idea associations

hyperonyms – hyponymes – part-of – whole– mater/substance ...

synonymes – antonyms– locutions – magn/antimagn ...

agent - patients – instruments – locations– causes/consequences – telic role– temporal values...

# ... AS A LEXICAL NETWORK

## Nodes

Terms, textual segment, NP

Usages, concepts

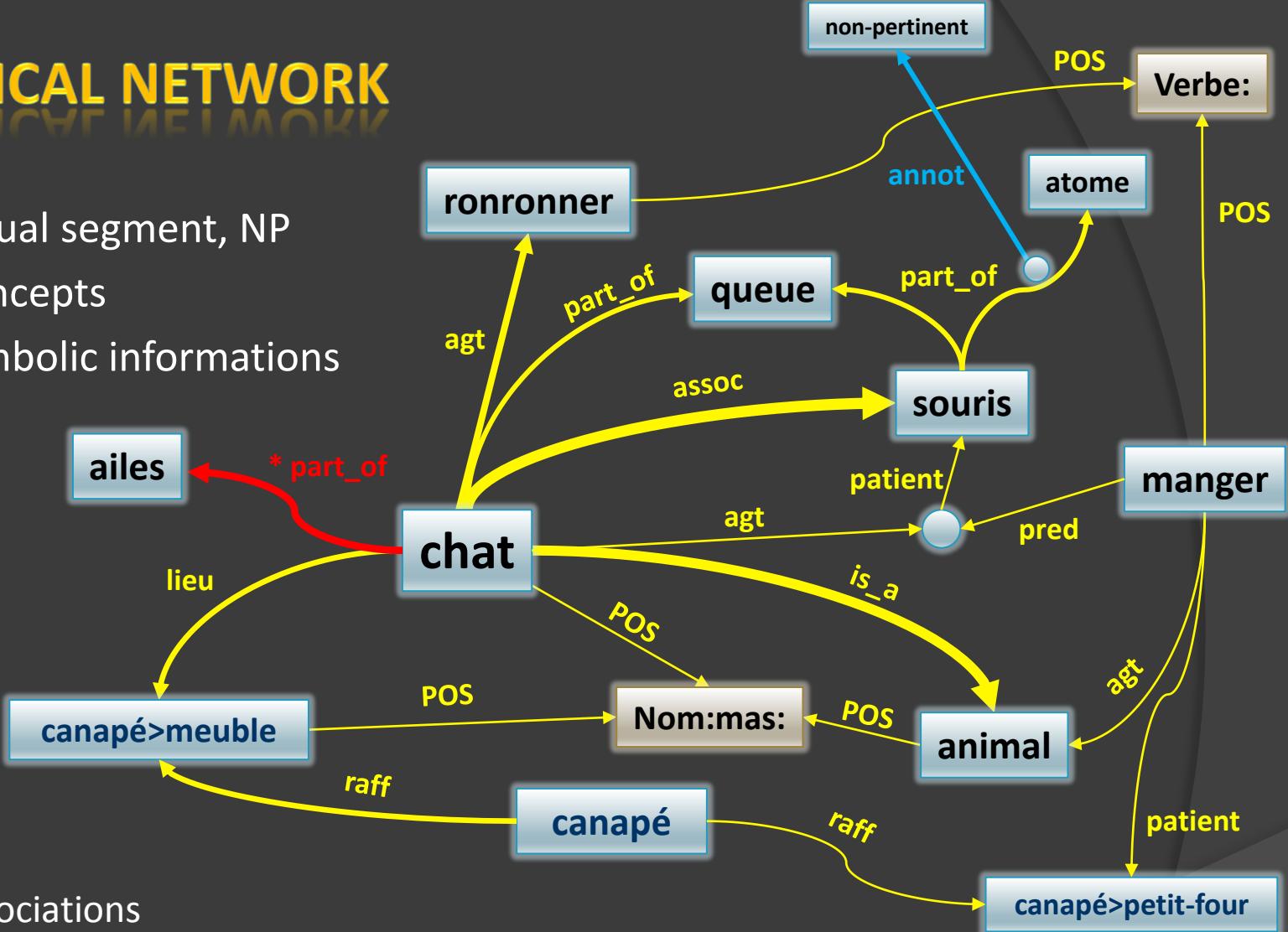
Various symbolic informations

## Relations

Typed

Directed

Weighted



free idea associations

hypernyms – hyponymes – part-of – whole– mater/substance ...

synonymes – antonyms – locutions – magn/antimagn ...

agent - patients – instruments – locations – causes/consequences – telic role – temporal values...

# LEXICAL ACQUISITION WITH GWAPS

GAME WITH A PURPOSE

# Jeux de mots

HYPOTHESIS :

FOR THE DEVELOPER, DATA ACQUISITION THROUGH GWAP IS  
FREE • FAST • EFFICIENT • PROVIDES NON-NEGOCIATED ANSWERS

AND FOR PLAYERS THE GAME WILL/MUST BE  
EXCITING • REWARDING • ADDICTIVE

## DONNER DES IDEES ASSOCIEES AU TERME QUI SUIT :

kaput  
Niveau : 209  
Crédits : 199400  
Honneur : 194682  
● 2782  
● 1000



Niveau : 51

# lutin



Temps

49 s

30s

OK



Dernier terme proposé : **petit**

Raffinements possibles :

1. petit (taille)
2. petit (jeune)
3. petit (minuscule)
4. petit (modique)

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

petit >  
agaçant  
taquin  
pétulant  
espiègle  
enfant >  
gobelins  
gnome  
farfadet  
tourmenter  
esprit follet  
Bretagne  
être fantastique  
lutiner  
petit >

15/28



## JDM PLAY EXAMPLE

# lutin



Réponses données par kaput : petit • agaçant • taquin • pétulant • espiègle • enfant • gobelin • gnome • farfadet • tourmenter • esprit follet • Bretagne • être fantastique • lutiner • petit

Réponses données par egaillat : grelot • oreille • habit • Schtroumpfs • malicieux • nez • créature légendaire • taquin • oreilles • elfe fée • forêt • nuton • Puck • Père Noël • espiègle • mythologie • légende • vert • esprit follet • bonnet • créature • conte • être imaginaire • gnome • chapeau • petit • troll • korrigan • farfadet

petit • taquin • espiègle • gnome • farfadet • esprit follet



**Vous gagnez 300 crédits et 8 point(s) d'honneur**



Soyez le premier de vos amis à indiquer que vous aimez ça.



## RESULTS

Esprit follet

farfadet

gnome

lutin

petit

taquin

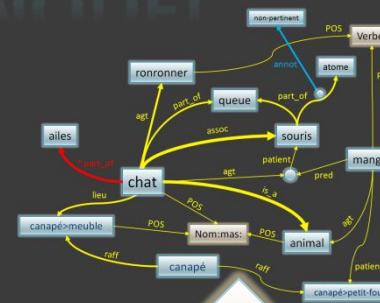
espiègle

assoc

# INTERACTION MODEL

game 1

term  
+  
instruction



lexical network

game 2

term  
+  
instruction

**creation / strengthening  
of relations**

confrontation

propositions      **Intersection**      propositions

player 1



**rewards**

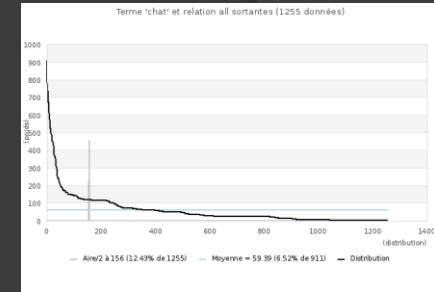


player 2

# PLAYER ACTIVITY REGULATION

## Filtering - matching of player pairs

- Iterated Minimal Consensus (weighting)
- Minimizing noise, maximizing *recall* (long tail)



## Features

- Word pseudo-randomly selected
- Other player(s) unknown during play
- Asynchronous games

## Points

- more if relation is weak
- less if relation is strong

## PLAYER DILEMNA

THINKING LIKE WE THINK  
THE OTHER PLAYER THINKS  
AND  
BEING ORIGINAL

# QUANTITATIVE EVALUATION

- > 1 000 000 terms + many word forms in the network
  - > 1 270 000 terms with at least one associated idea
- > 67 000 000 lexical relations
  - > 1 340 000 terms with at least one outgoing relation ( $A \rightarrow B$ )
  - > 990 000 terms with at least one incoming relation ( $A \leftarrow B$ )
- > 1 000 000 with one or several POS (part of speech)  
(4.5 % no pos)
- > 24 000 refined terms and > 66 000 usages
- > 25 200 labelled as polysemous (coverage 98 %)
- > 829 000 inhibitory (negative) relations

never ending learning process

new words, NP, refinements... new relations

# QUALITATIVE EVALUATION WITH COUNTER-GAMES



Indirect approaches :

Totaki – a guessing game/ AskIt – a question game/ ...

- **Totaki** : {clues} => term

Player = clue giver (+ optional relation type)

Totaki = guesser (lexical network + learning + short term memory)

- Looking for quasi intersection in the lexical network

- **Hypothesis** : if the target term is found

→ the network is properly built/informed for this term

Après 4 indice(s), je suis perdu, désolé je donne ma langue au chat

Il s'agissait de gare de triage

Ok

(faites attention aux accents et aux majuscules/minuscules surtout si c'est un i)

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

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

# REVERSED TABOU – FREQUENT TERMS



Can we find terms from the clue?

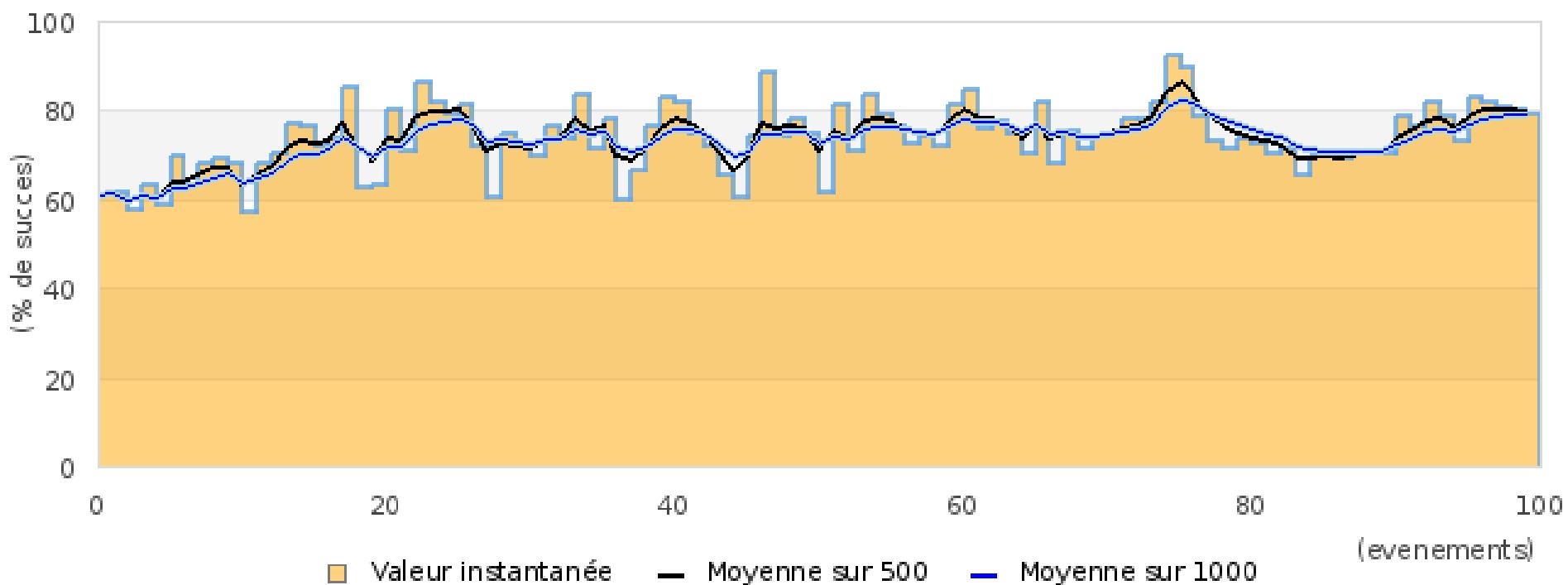
With the 500 riddles of the original game : AKI 494 – humans : 398

Totaki 98,8 %

Humans 79,6 %

# RESULTS FOR OPEN VOCABULARY

Données Totaki (25 153 données de taille 251)



Totaki ~ 80 %

Tests over 25 000 games  
where terms are chosen by  
players

Humans ~ 43 %

Tests over 300 terms on  
which players made some  
guesses (controled env.)

# Other games



SEXIT



ask  
you



# Counter games

# Other games



# Counter games

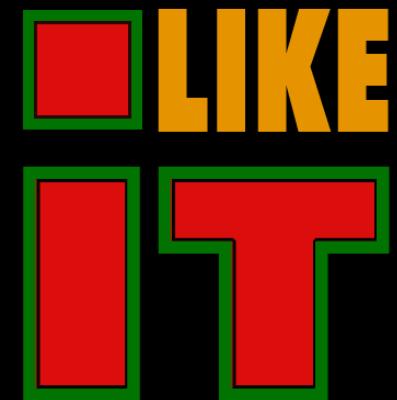
# LIKE IT

## A POLARITY GAME

Est-ce que vous aimez l'idée de

aller chez le dentiste





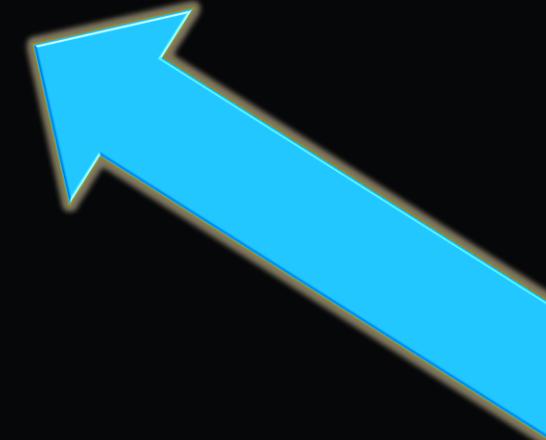
# A POLARITY GAME

aller chez le dentiste

votre réponse précédente était non comme 92% des personnes

Est-ce que vous aimez l'idée de

aloès du Cap



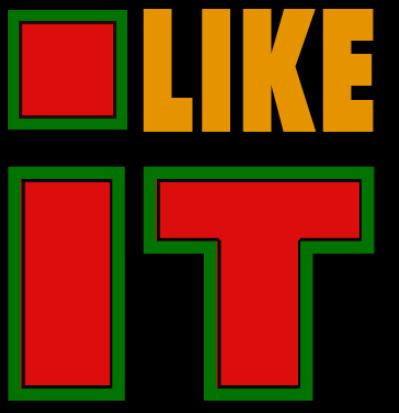


# CANDIDATE SELECTION THROUGH A PROPAGATION ALGORITHM

(pseudo random walk in the network)

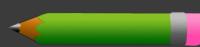
## Infinite iteration of

- Random selection of a term **T** having a **positive** or a **negative** polarity (or both)
- 50% proposing **T**  
50% proposing one neighbor of **T** in the network
- Seed with:
  - bien** = 1 **positive** vote
  - mal** = 1 **negative** vote



657,843 polarized terms

- 551,871 positive polarity                            655,003 neutral polarity  
366,913 negative polarity
- Total number of votes = 146,080,950
  - 70,698,908 positive votes (48.4 %)
  - 51,185,949 neutral votes (35 %)
  - 24,196,093 negative votes (16.6 %)
- mean of 220 votes per entry  
(beware! power law)



**Associations d'idées** > 34 fracture ▶ tibia ▶ jambe ▶ fracture (lésion) ▶ plâtre ▶ traumatisme ▶ fracture spiroïde ▶ lésion ▶ os (squelette) ▶ os ▶ médecine ▶ lésion physique ▶ ostéosynthèse ▶ blessure sportive ▶ traumatisme (physique) ▶ accident ▶ blessure ▶ plâtre (médecine) ▶ douleur (physique) ▶ lésion osseuse ▶ fracture du plateau tibial ▶ fracture de Segond ▶ clou centro-médullaire ▶ fissure (médecine) ▶ douleur ▶ fracture ouverte ▶ chute ▶ jambe (membre) ▶ blessé ▶ blessure (lésion physique) ▶ traumatologie ▶ orthopédie ▶ fracture du fracture ▶ tibia ▶ traumatologie ▶ orthopédie ▶ médecine ▶ lésion physique ▶ fracture (lésion) ▶ lésion ▶ lésion osseuse

**Est accompagné par** ▶ fracture du péroné ▶ fracture de la fibula

**Thèmes/domaines** ▶ médecine (science) ▶ médecine ▶ traumatologie ▶ orthopédie ▶ radiologie

**Génériques** H ▶ fracture (lésion) ▶ fracture ▶ lésion osseuse ▶ lésion physique ▶ lésion ▶ \* fracture (sociologie)

**Symptôme(s)** ▶ déformation (médecine) ▶ déformation ▶ douleur (physique) ▶ douleur ▶ **Diagnostique(s)** ▶ scanner (médecine, technique) ▶ scanner (médecine) ▶ radiographie (cliché) ▶ radiographie ▶

**Plus intense que fracture du tibia** ▶ fracture double ▶ double fracture **Moins intense que fracture du tibia** ▶ foulure ▶ entorse ▶

**Locutions/termes composés** tibia ▶ fracture ▶ fracture du ▶ fracture (lésion)

**Caractéristiques de fracture du tibia** ▶ fermée ▶ ouverte ▶ grave ▶ non déplacée ▶ spiroïde ▶ douloureuse (souffrance) ▶ complexe (compliqué) ▶ diaphysaire ▶ comminutive ▶ douloureuse ▶ complexe ▶ \* hépatique ▶

**A quoi fracture du tibia peut-il s'opposer/ combattre ?** ▶ marche (mouvement) ▶ marche ▶

**Lieux incluant/contenant fracture du tibia ?** ▶ tibia ▶ jambe (membre) ▶ jambe ▶ corps ▶ \* bras ▶

**Que peut faire fracture du tibia ? (agent)** ▶ faire souffrir ▶ faire mal ▶ **Que peut-on faire à/de fracture du tibia ? (patient)** ▶ réduire ▶ visualiser ▶ radiographier ▶ plâtrer ▶ opérer ▶ opérer (chirurgie) ▶ diagnostiquer

**Causes associées à fracture du tibia** ▶ ski (sport) ▶ ski ▶ se blesser ▶ se battre ▶ sport ▶ sport (activité physique) ▶ traumatisme (physique) ▶ traumatisme ▶ tomber ▶ glisser ▶ coup (choc) ▶ accident de ski ▶ accident de moto ▶ accident de la route ▶ accident ▶ activité physique ▶ blessure sportive ▶ coup ▶ chute ▶ choc ▶ Sport

**Conséquences associées à fracture du tibia** ▶ radio ▶ radiographie ▶ soin ▶ soin (acte médical) ▶ plâtre (médecine) ▶ plâtre ▶ broche (médecine) ▶ douleur (physique) ▶ immobilité ▶ marcher avec des béquilles ▶ broche ▶

**Sentiments/émotions associés à fracture du tibia** ▶ fatalité ▶ contrariété ▶ amertume (tristesse) ▶ malchance ▶ ennui ▶ ennui (contrariété) ▶ mécontentement ▶ tracas ▶ souffrance ▶ rage ▶ dépit ▶ découragement ▶ consternation ▶ colère ▶ calamité ▶ culpabilité ▶ dépendance (assujettissement) ▶ abattement ▶ déception ▶ douleur ▶ amertume ▶ horrible

**Rôles agentifs fracture du tibia** ▶ se faire ▶ provoquer ▶ occasionner

## Sentiments/émotions associés à fracture du tibia ➤ 26 colère ■ fatalité ■

amertume (tristesse) ■ contrariété ■ malchance ■ ennui (contrariété) ■ ennui  
 ➤ ■ mécontentement ■ rage ➤ ■ triste (malheureux) ■ tracas ■ souffrance ■ dépit  
 ■ peur ■ découragement ■ consternation ■ calamité ■ angoisse (médecine) ■  
 amertume ➤ ■ culpabilité ■ douleur ➤ ■ déception ■ dépendance  
 (assujettissement) ■ douleur (physique) ■ abattement ➤ ■ horrible

**Associations d'idées** ➤ 34 fracture ➤ ■ tibia ■ jambe ➤ ■ fracture (lésion) ■ plâtre ➤ ■ traumatisme ➤ ■ fracture spiroïde ■  
 lésion ■ os (squelette) ■ os ➤ ■ médecine ➤ ■ lésion physique ■ ostéosynthèse ■ blessure sportive ■ traumatisme (physique) ■ accident ➤ ■  
 blessure ➤ ■ plâtre (médecine) ■ douleur (physique) ■ lésion osseuse ■ fracture du plateau tibial ■ fracture de Segond ■ clou centro-médullaire ■  
 fissure (médecine) ■ douleur ➤ ■ fracture ouverte ■ chute ➤ ■ jambe (membre) ■ blessé ■ blessure (lésion physique) ■ traumatologie ■  
 orthopédie ■ fracture du fracture ➤ ■ tibia ■ traumatologie ■ orthopédie ■ médecine ➤ ■ lésion physique ■ fracture (lésion) ■ lésion ■  
 lésion osseuse

**Est accompagné par** ➤ fracture du péroné ■ fracture de la fibula

**Thèmes/domaines** ➤ ■ médecine (science) ■ médecine ➤ ■ traumatologie ■ orthopédie ■ radiologie

**Génériques** H ➤ fracture (lésion) ■ fracture ➤ ■ lésion osseuse ■ lésion physique ■ lésion ■ \* fracture (sociologie)

**Symptôme(s)** ➤ ■ déformation (médecine) ■ déformation ➤ ■ douleur (physique) ■ douleur ➤ ■ **Diagnostique(s)** ➤ ■ scanner  
 (médecine, technique) ■ scanner (médecine) ➤ ■ radiographie (cliché) ■ radiographie ➤

**Plus intense que fracture du tibia** ➤ ■ fracture double ■ double fracture **Moins intense que fracture du tibia** ➤ ■ foulure ■ entorse ➤

**Locutions/termes composés** ■ tibia ■ fracture ➤ ■ fracture du ■ fracture (lésion)

**Caractéristiques de fracture du tibia** ➤ ■ fermée ■ ouverte ■ grave ➤ ■ non déplacée ■ spiroïde ■ douloureuse (souffrance) ■  
 complexe (compliqué) ■ diaphysaire ■ comminutive ■ douloureuse ➤ ■ complexe ➤ ■ \* hépatique ➤

**A quoi fracture du tibia peut-il s'opposer/combattre ?** ➤ ■ marche (mouvement) ■ marche ➤

**Lieux incluant/contenant fracture du tibia ?** ➤ ■ tibia ■ jambe (membre) ■ jambe ■ corps ➤ ■ \* bras ➤

Chercher

la forme

fracture du tibia



P	ch
pi	re
ta	yo
CO2	te
UKN	se
CAR	xy
AGT	ca
ISA	

tæ
n

F
R

D
32

## fracture du tibia

Nom

- Natures grammaticales

ok

Lemme

Forme incorrecte de



Informations diverses

wiki

- Informations lexicales

ok

- Informations sémantiques

ok

polarité



Masculin équivalent



Féminin équivalent



Variantes



Raffinements morpho

Raffinements sem

Gloses (termes évoquant les sens possibles)



Inhib

Associations d'idées



béquille



fracture ▷ tibia ▷ jambe ▷ fracture (lésion)

plâtre ▷ traumatisme ▷ fracture spiroïde ▷ lésion ▷ médecine ▷ lésion physique ▷ os ▷ lésion osseuse ▷ ostéosynthèse ▷ traumatisme (physique) ▷ accident ▷ plâtre (médecine) ▷ blessure (lésion physique) ▷ os (squelette) ▷ jambe (membre) ▷ douleur (physique) ▷ fissure (médecine) ▷ douleur ▷ clou centro-médullaire ▷ blessé ▷ chute ▷ blessure ▷ blessure sportive ▷ fracture du plateau tibial ▷ fracture de Segond ▷ fracture ouverte ▷ Médecine ▷ traumatologie ▷ radiologie ▷ orthopédie ▷ médecine (science) ▷ fracture du ↵12 fracture ▷ tibia ▷ médecine ▷ médecine (science) ▷ Médecine ▷ radiologie ▷ orthopédie ▷ traumatologie ▷ lésion physique ▷ fracture (lésion) ▷ lésion

1éision osseuse ↵ fracture tassemement [>] [-] [x] [!]) ▷ maladie (médecine) [>] [-] [x] [!]) [?] ▷ fatigue [>] [-] [x] [!]) ou mais bli et bloblu, non ? couture [>] [-] [x] [!]) ▷ panser [>] [-] [x] [!]) ▷ maladie [>] [-] [x] [!]) ▷ pansement [>] [-] [x] [!]) ▷ plaie [>] [-] [x] [!]) ▷ stress [>] [-] [x] [!]) ▷ avoir mal [>] [-] [x] [!]) ▷ médecine (remède) [>] [-] [x] [!]) [?] ▷ (sociologie) [>] [-] [x] [!]) [?] ▷ urgences [>] [-] [x] [!]) ▷ cassé [>] [-] [x] [!]) ▷ hospitalisation [>] [-] convalescence [>] [-] [x] [!]) ▷ AVP [>] [-] [x] [!]) ▷ pérone ● ski ▷ tibia ● genou ● fracture ▷ Genou ● cheville ▷ consolidation

Fracture

double fracture

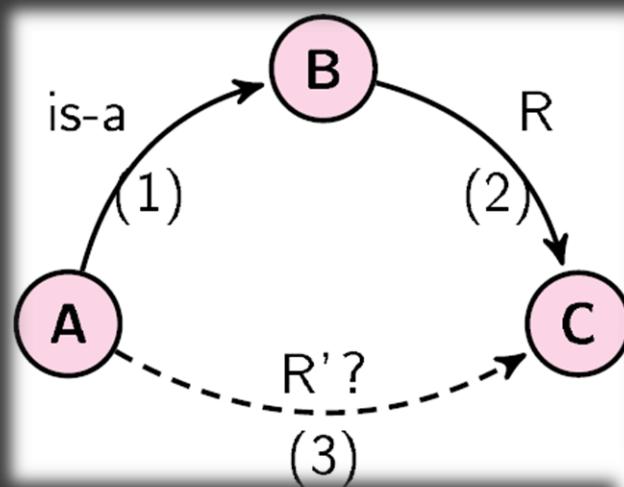
Est accompagné par



fracture du péroné ● fracture de la fibula ● Liens sociaux/familiaux



# CONSOLIDATION AND CHECKING WITH INFERENCES - DEDUCTION

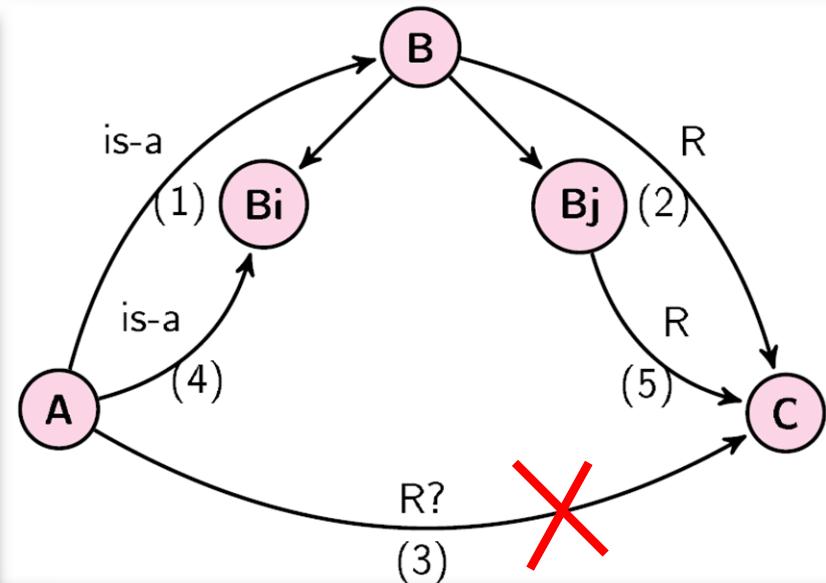


From the most general toward the most specific

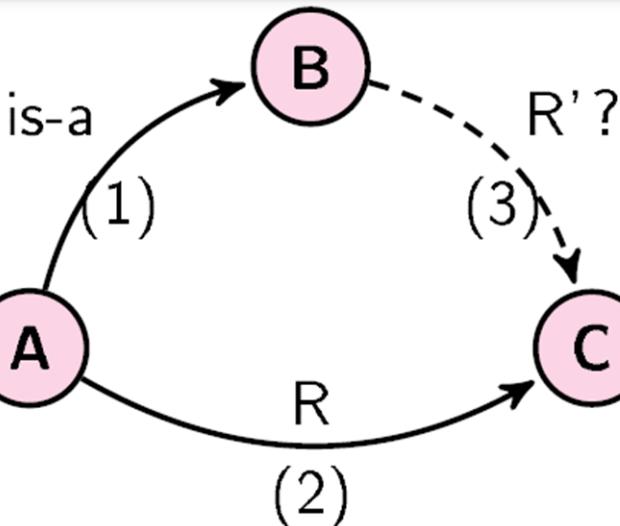
Logical and statistical blocking because of polysemy - for example:

- livre > lecture
- livre > monnaie
- livre > masse

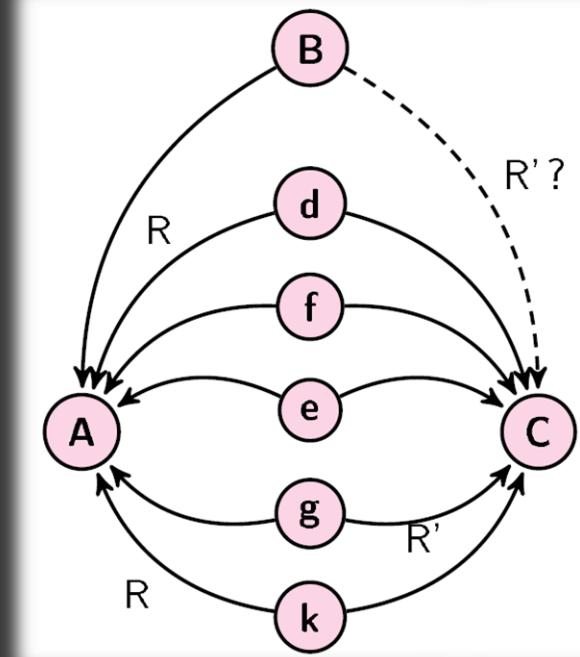
\* Bible *is-a* livre & livre *carac* convertible  
=> Bible *carac* convertible



# CONSOLIDATION AND CHECKING USING INFERENCES – INDUCTION AND ABDUCTION



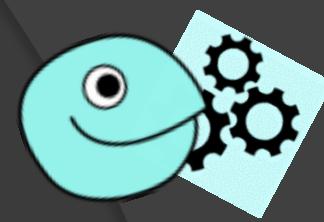
From specific to general



imitation of  
examples

About 93 % of the inferred relations are correct and relevant

# SOME REFINEMENTS



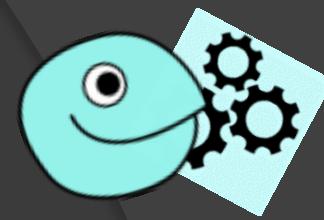
- For polysemy and word usages

- avocat --r-raff\_sem--> avocat>**fruit** ← gloses
- avocat --r-raff\_sem--> avocat>**justice** ←
- grippe --r-raff\_sem--> grippe>maladie
- grippe --r-raff\_sem--> grippe>virus

➤ 24 000 termes raffinés  
➤ et > 66 000 usages

# SOME REFINEMENTS

2014 REFINEMENTS



- ◎ Decision tree, example with **frégate**

--r-raff\_sem--> frégate>**navire**

--r-raff\_sem--> frégate>**navire**>**moderne**

--r-raff\_sem--> frégate>**navire**>**ancien**

--r-raff\_sem--> frégate>**oiseau**

...

Navire de guerre ancien ou moderne.

(Navigation) (Marine) (Militaire) Navire de guerre qui n'avait qu'une seule batterie couverte et qui portait de vingt à soixante bouches à feu. [...]

(Marine) (Militaire) Navire de guerre (moderne) de taille moyenne, capable d'assurer plusieurs types de missions. [...]

(Ornithologie) Oiseau de mer palmipède, d'une très grande envergure, et qui saisit à la surface des poissons dont il se nourrit. [...]

Automobile du constructeur Renault.

(Argot polytechnicien) (Désuet) Bicorne. [...]

(Sexualité) Jeune homosexuel, giton. [...]

WSD = selecting proper refinements  
Activation algorithm through relations with other terms



# NEGATIVE RELATIONS

Allow to represent

> 270 000 negative relations in the lexical network

- exceptions
  - autruche --r-agent-1<0--> voler
- Inductive inferences potentially relevant, but wrong
  - ver de terre --r-agent-1<0--> mordre
- contrastive informations between refinements
  - avocat>**fruit** --r-agent-1<0--> plaider
  - avocat>**justice** --r-has-part<0--> noyau>**fruit**

Negative relations can be used as inhibition in WSD

► what's specific?



# INHIBITORY RELATIONS

A term linked to a refinement **inhibites**  
its co-refinements if not linked

chat (félin)

<-- r\_inhib --

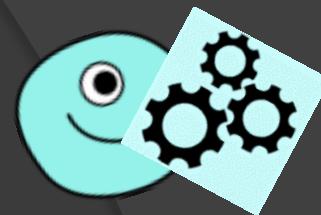
siamois (Siam) • siamois (jumeaux) • sphynx (sphinx) •  
coussinet (chemin de fer) • minou (affection) • coussinet  
(mécanique) • coussinet (architecture) • minou (sexe de  
la femme) • coussinet (coussin) • persan (Perse)

> 370 000 inhibitory  
relations in the  
lexical network

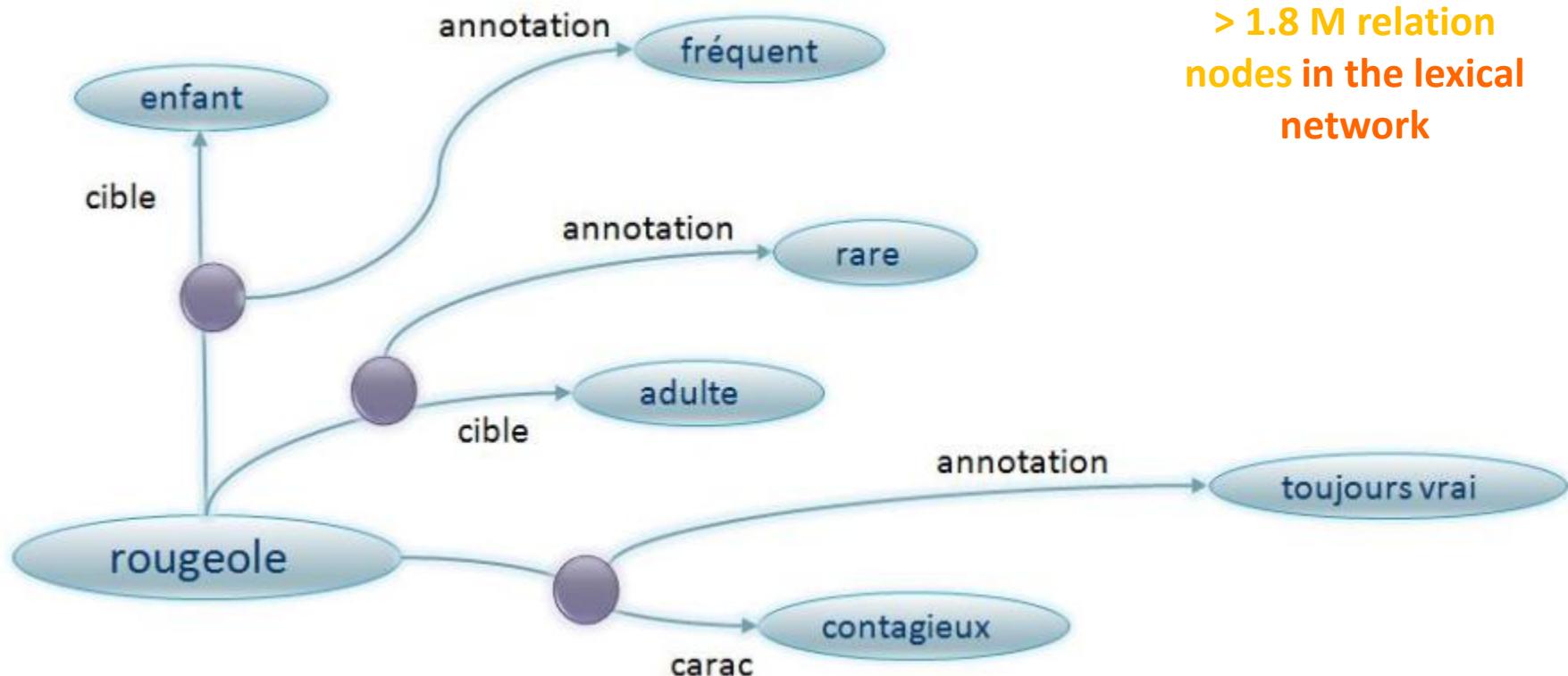
Negative relations can be used as inhibition  
in WSD based on a thematic approach

► what's specific?

# RELATION ANNOTATIONS



- Reification of a relation  
with new associated informations

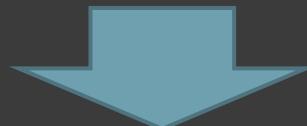




# AGGREGATIONS

## ◎ Another reification form

lion --r\_agent-1 -> dévorer



> 3366 aggregated forms in the lexical network

lion [agent] dévorer

-- r\_patient -> gazelle, zèbre

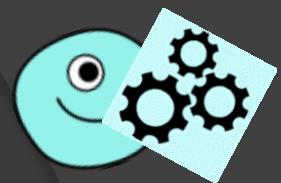


(lion [agent] dévorer) [patient] gazelle

-- r\_action lieu -> savane

-- r\_manner -> férolement

► what's specific?



# CONCEPTUAL INFORMATIONS

\_INFO-COUNTABLE-NO

\_INFO-COUNTABLE-YES

\_INFO-SEM-ACTION

\_INFO-SEM-CARAC

\_INFO-SEM-COLOR-RELATED

\_INFO-SEM-EMOTION-RELATED

\_INFO-SEM-IMAGINARY

\_INFO-SEM-LIVING-BEING

\_INFO-SEM-NAMED-ENTITY

\_INFO-SEM-ORGA

\_INFO-SEM-PERS \_INFO-SEM-PERS-  
FEM \_INFO-SEM-PERS-MASC

\_INFO-SEM-PLACE

ABSTRACT ANATOMICAL

GEO HUMAN

\_INFO-SEM-PROPERTY-NAME

\_INFO-SEM-QUANTIFIER

\_INFO-SEM-SET \_INFO-SEM-SUBST

\_INFO-SEM-THING

...-ABSTR

...- ARTEFACT

...- CONCRETE

...- NATURAL

\_INFO-SEM-TIME \_INFO-SEM-EVENT

> 4.6M conceptual  
relations

► what's specific?

# APPLICATIONS

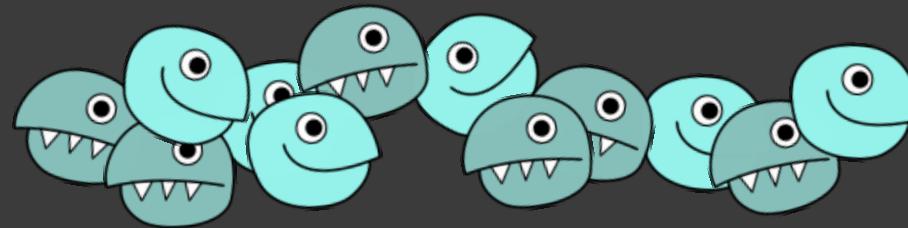


- Analysis of medical imaging reports (Imaios)  
=> Indexation = weighted term list
- On a specific domain medicine-radiology-anatomy  
(weighting TFIDF or Okapi)
  - F1-score = 70%
  - With augmentation (syn, generic, cause, consequence, etc.) => F1-score = 75%
  - Adding general domain => + 12 %
  - Refinement selection => +7%
  - With inhibition => +3%

# FIRST GENERAL CONCLUSIONS

## ◎ Lexical resources

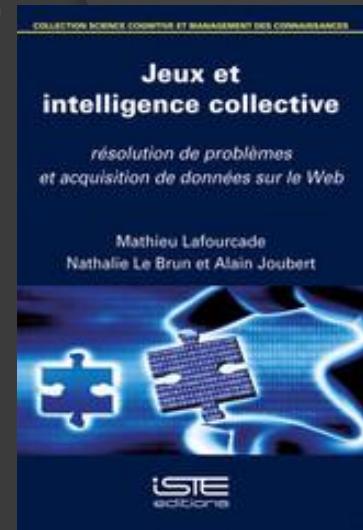
- If **not too specialized**
- Can be built with **native speakers**
- As many as you can, but a **dozen** can be enough...  
if motivated

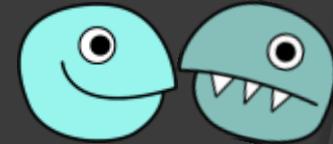


Not expensive – reliable ‘coz collected data are redundant

# MORE JDM CONCLUSIONS

- Since sept. 2007 (~ 10 years)
  - > 67 millions relations between over 1 000 000 terms and around 100 relation types
  - annotated relations (relevant, possible, not relevant)
    - the largest network of this type
    - already used for research and by some companies
- Evaluation
  - Collation of various points of view – negotiated (diko) or not (games)
  - Implicit relations (not present in texts) are captured by **instruction forcing** (players are invited to be explicit)



gwap  crowdsourcing 

principles are globally validated for lexical networks

- Relevant for general knowledge but also for specific domains (great surprise!)
- With GWAP (JDM, Askit, Likelt, ...) but also with contributions (Diko)
- In general, **virtuous circle** is difficult to identify
  - playing well ↔ producing proper data

# EVEN MORE CONCLUSIONS

## Some ethical aspects

- many involved players (some with more than 2000 hrs of play)
  - no memory in the lexical network of who has made what  
(only temporary storage of games still to be retrieved)
  - players are anonymous (login + pwd + email)
  - less than 1% troll / vandalism – corrected as soon as discovered
- 
- The data are made by the crowd...  
... and should return to the crowd



Freely available



THANK YOU

Jeux de mots

JEUXDEMOTS.ORG

DEMONSTRATIONS  
QUESTIONS



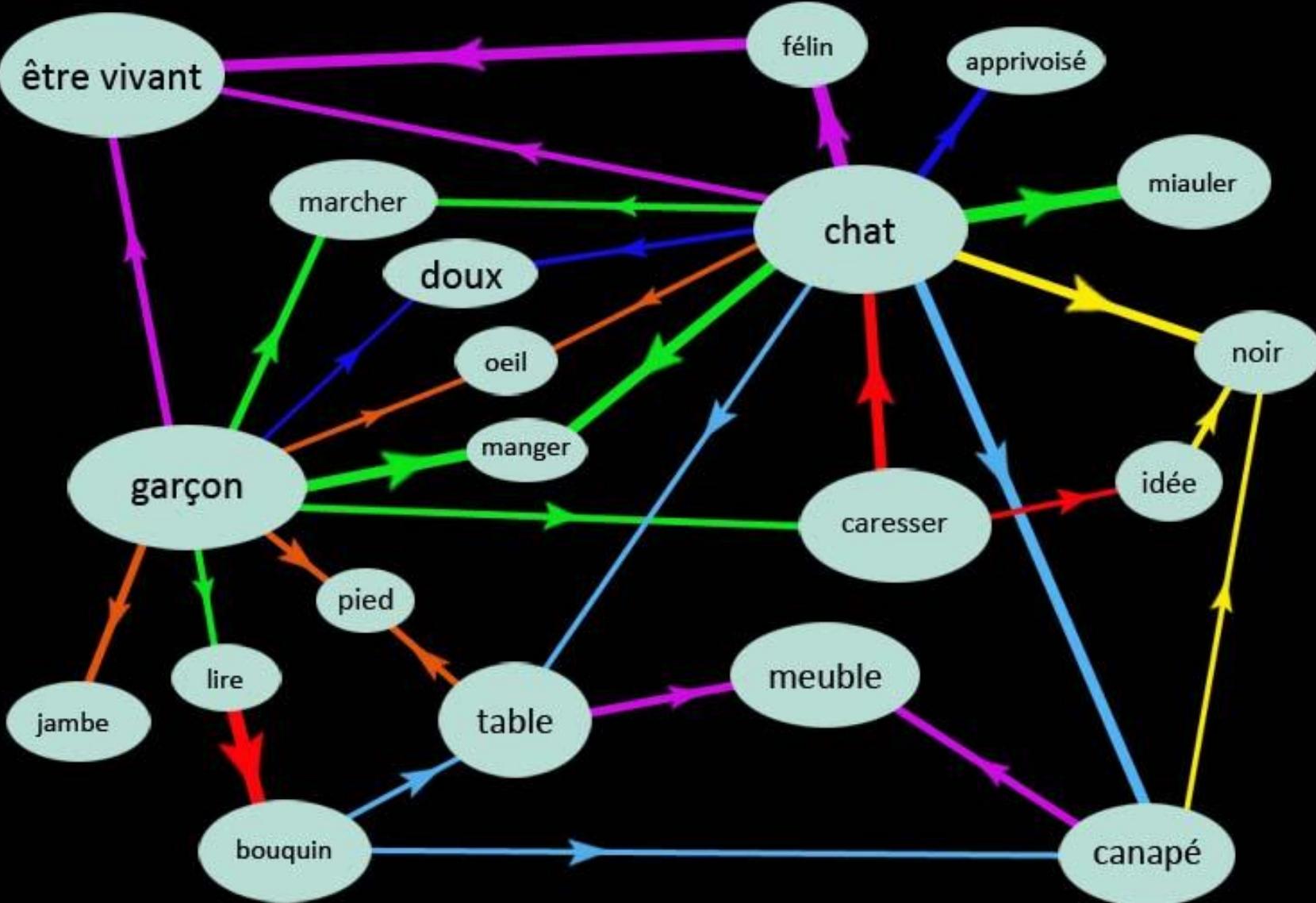
# DETAIL OF RELATIONS DISTRIBUTION

6542950 r\_associated (0) - 30.467 %  
 1531831 r\_pos (4) - 7.133 %  
 454401 r\_hypo (8) - 2.116 %  
 191855 r\_flop (12) - 0.893 %  
 24636 r\_instr (16) - 0.115 %  
 7168 r\_magn (20) - 0.033 %  
 259958 r\_agent-1 (24) - 1.21 %  
 93092 r\_lieu-1 (28) - 0.433 %  
 21873 r\_sentiment (32) - 0.102 %  
 1133563 r\_infopop (36) - 5.278 %  
 6540 r\_action-verbe (40) - 0.03 %  
 460 r\_verbe-adj (44) - 0.002 %  
 1245 r\_chunk\_instr (48) - 0.006 %  
 3075 r\_successeur-time (52) - 0.014 %  
 3397 r\_against-1 (56) - 0.016 %  
 154 r\_fem (60) - 0.001 %  
 121344 r\_instance (64) - 0.565 %  
 83 r\_set>item (68) - 0 %  
 1674 r\_syn Strict (72) - 0.008 %  
 435 r\_der\_morpho (99) - 0.002 %  
 6882 r\_has\_actors (103) - 0.032 %  
 215 r\_cible (107) - 0.001 %  
 112 r\_predecesseur-space (111) - 0.001 %  
 28 r\_descend\_de (151) - 0 %  
 1005096 r\_annotation (998) - 4.68 %  
 81 r\_termgroup (1002) - 0 %

25083 r\_raff\_sem (1) - 0.117 %  
 236897 r\_syn (5) - 1.103 %  
 640038 r\_has\_part (9) - 2.98 %  
 254829 r\_agent (13) - 1.187 %  
 113062 r\_carac (17) - 0.526 %  
 6644 r\_antimagn (21) - 0.031 %  
 10032 r\_instr-1 (25) - 0.047 %  
 7469 r\_chunk\_pred (29) - 0.035 %  
 923 r\_error (33) - 0.004 %  
 6909 r\_telic\_role (37) - 0.032 %  
 41072 r\_conseq (41) - 0.191 %  
 2813 r\_chunk\_sujet (45) - 0.013 %  
 1632 r\_time (49) - 0.008 %  
 4155 r\_make (53) - 0.019 %  
 403 r\_implication (57) - 0.002 %  
 1809 r\_equiv (61) - 0.008 %  
 67 r\_verb\_real (65) - 0 %  
 202 r\_item>set (69) - 0.001 %  
 1406 r\_bigger\_than (73) - 0.007 %  
 1489 r\_has\_auteur (100) - 0.007 %  
 52 r\_deplac\_mode (104) - 0 %  
 448 r\_symptomes (108) - 0.002 %  
 37 r\_successeur-space (112) - 0 %  
 82108 r\_aki (666) - 0.382 %  
 48238 r\_inhib (999) - 0.225 %  
 6 r\_learning\_model (2001) - 0 %

33727 r\_raff\_morpho (2) - 0.157 %  
 585836 r\_isa (6) - 2.728 %  
 632055 r\_holo (10) - 2.943 %  
 43111 r\_patient (14) - 0.201 %  
 11298 r\_data (18) - 0.053 %  
 17113 r\_familly (22) - 0.08 %  
 43098 r\_patient-1 (26) - 0.201 %  
 25166 r\_lieu\_action (30) - 0.117 %  
 15213 r\_maner (34) - 0.071 %  
 3921 r\_agentif\_role (38) - 0.018 %  
 41000 r\_causatif (42) - 0.191 %  
 3736 r\_chunk\_objet (46) - 0.017 %  
 5580 r\_object>mater (50) - 0.026 %  
 2634 r\_product\_of (54) - 0.012 %  
 635 r\_quantificateur (58) - 0.003 %  
 655 r\_maner-1 (62) - 0.003 %  
 0 r\_chunk\_head (66) - 0 %  
 280 r\_processus>agent (70) - 0.001 %  
 1546 r\_smaller\_than (74) - 0.007 %  
 80 r\_has\_personnage (101) - 0 %  
 1 r\_has\_interpret (105) - 0 %  
 744 r\_predecesseur-time (109) - 0.003 %  
 7 r\_social\_tie (113) - 0 %  
 5777889 r\_wiki (777) - 26.905 %  
 0 r\_prev (1000) - 0 %

188145 r\_domain (3) - 0.876 %  
 18417 r\_onto (7) - 0.085 %  
 532905 r\_locution (11) - 2.481 %  
 101387 r\_lieu (15) - 0.472 %  
 78976 r\_lemma (19) - 0.368 %  
 102743 r\_carac-1 (23) - 0.478 %  
 186970 r\_domain-1 (27) - 0.871 %  
 25063 r\_action\_lieu (31) - 0.117 %  
 38718 r\_meaning (35) - 0.18 %  
 6632 r\_verbe-action (39) - 0.031 %  
 468 r\_adj-verbe (43) - 0.002 %  
 0 r\_chunk\_loc (47) - 0 %  
 622 r\_mater>object (51) - 0.003 %  
 1411 r\_against (55) - 0.007 %  
 362 r\_masc (59) - 0.002 %  
 265 r\_agentive\_implication (63) - 0.001 %  
 466 r\_similar (67) - 0.002 %  
 4479 r\_variante (71) - 0.021 %  
 554 r\_accomp (75) - 0.003 %  
 241 r\_can\_eat (102) - 0.001 %  
 40742 r\_color (106) - 0.19 %  
 466 r\_diagnostique (110) - 0.002 %  
 40 r\_beneficiaire (150) - 0 %  
 6 r\_annotation\_exception (997) - 0 %  
 0 r\_succ (1001) - 0 %



action ➔ objet

terme ➔ caractéristique

sujet ➔ action

terme ➔ lieu

terme ➔ générique

terme ➔ partie

terme ➔ couleur

## A typed directed and weighted lexical network: example

