



LIRMM - www.lirmm.fr

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

The screenshot displays the Foldit game interface. At the top left, it says "Pull Mode". In the top center, the score is "Score: 0 sur 8800". At the top right, it indicates the current level: "5-4: Control Over Clashing". A tooltip on the right asks "Répéter le guide ?" with an "OK!" button. The central part of the screen shows a 3D protein structure with red ribbons and yellow and blue spheres. At the bottom center, there is a control panel with "Clashing Importance: 1.00" and a slider, and "Wiggle Power: Low Medium High Auto". At the bottom left, there are navigation buttons: "Actions", "Annuler", "Behavior", and "Menu". At the bottom right, there is a "Chat - Global" button and a "montrer aut" button.

Some GWAPs in Biology

Eterna

Tutorial 4 : Stacks and Loops!

Total: 4.3 kcal

Chat

comes apart [7:41 PM]
drake178: hi Nando :) [7:41 PM]
jnicol: Hi Nando [7:42 PM]
drake178: 0.2 vs 4.57 total and it still doesn't fold in the puzzle itself :P [7:42 PM]
jnicol: Want me to tweak the energy bonus to make it solvable? [7:48 PM]
drake178: that is 0.2 vs 5.17 + 1.3 - 1.0 - 0.6 in case im miscalculating something >< [7:48 PM]

Use Target Mode to check which stacks you have to make. [Next](#)

Tutorial 4 : Stacks and Loops!

Total: 0 kcal

Chat

jnicol: Want me to tweak the energy bonus to make it solvable? [7:48 PM]
drake178: that is 0.2 vs 5.17 + 1.3 - 1.0 - 0.6 in case i'm miscalculating something >< [7:48 PM]
drake178: 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 [7:52 PM]

As a starting point, let's put your RNA in Target Mode.

Tutorial 4 : Stacks and Loops!

Total: 0 kcal

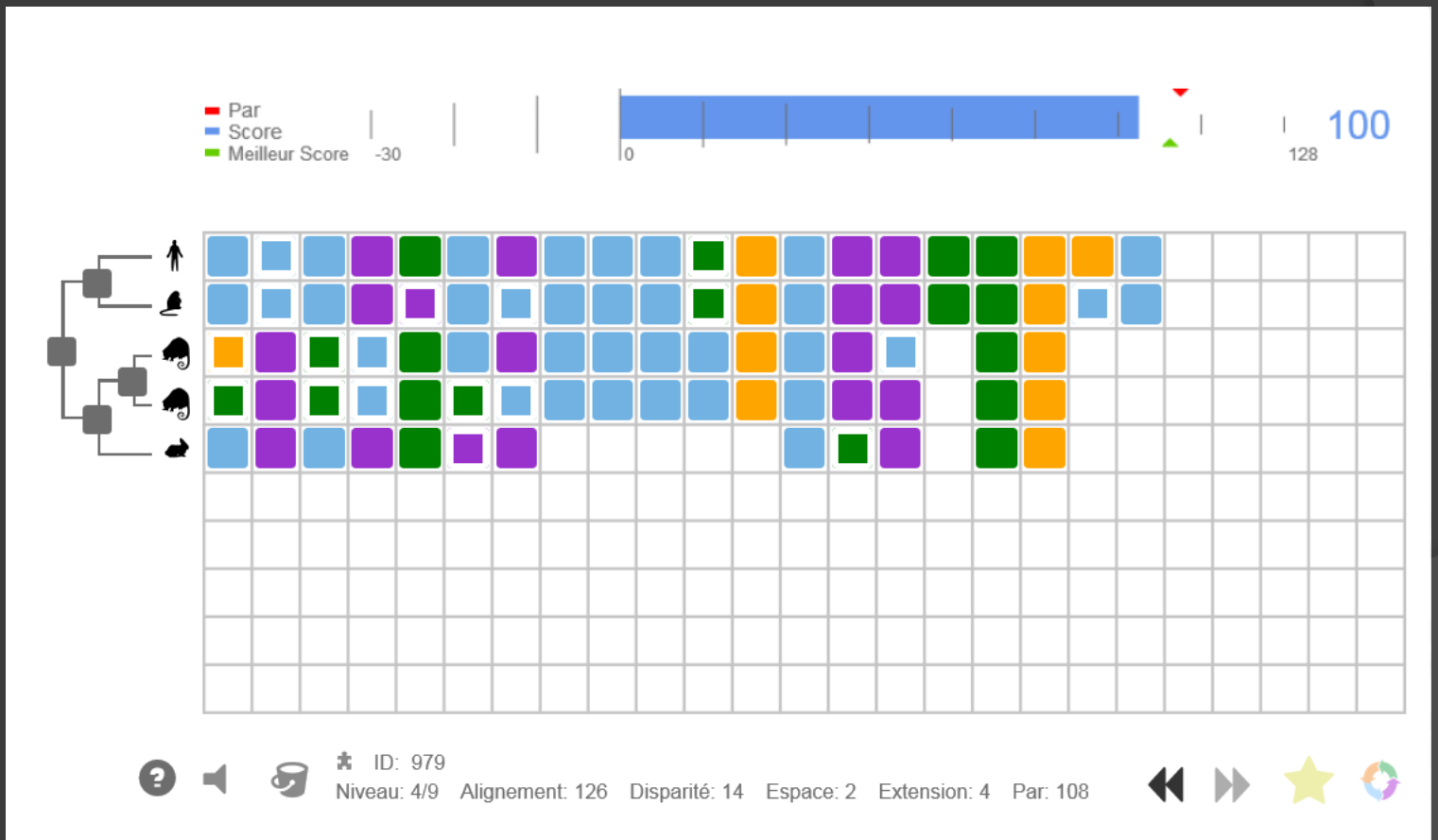
Chat

jnicol: Want me to tweak the energy bonus to make it solvable? [7:48 PM]
drake178: that is 0.2 vs 5.17 + 1.3 - 1.0 - 0.6 in case i'm miscalculating something >< [7:48 PM]
drake178: 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 [7:52 PM]

As a starting point, let's put your RNA in Target Mode.

Some GWAPs in Biology

Phylo



Some GWAPs in Biology

🕒 Nightjar



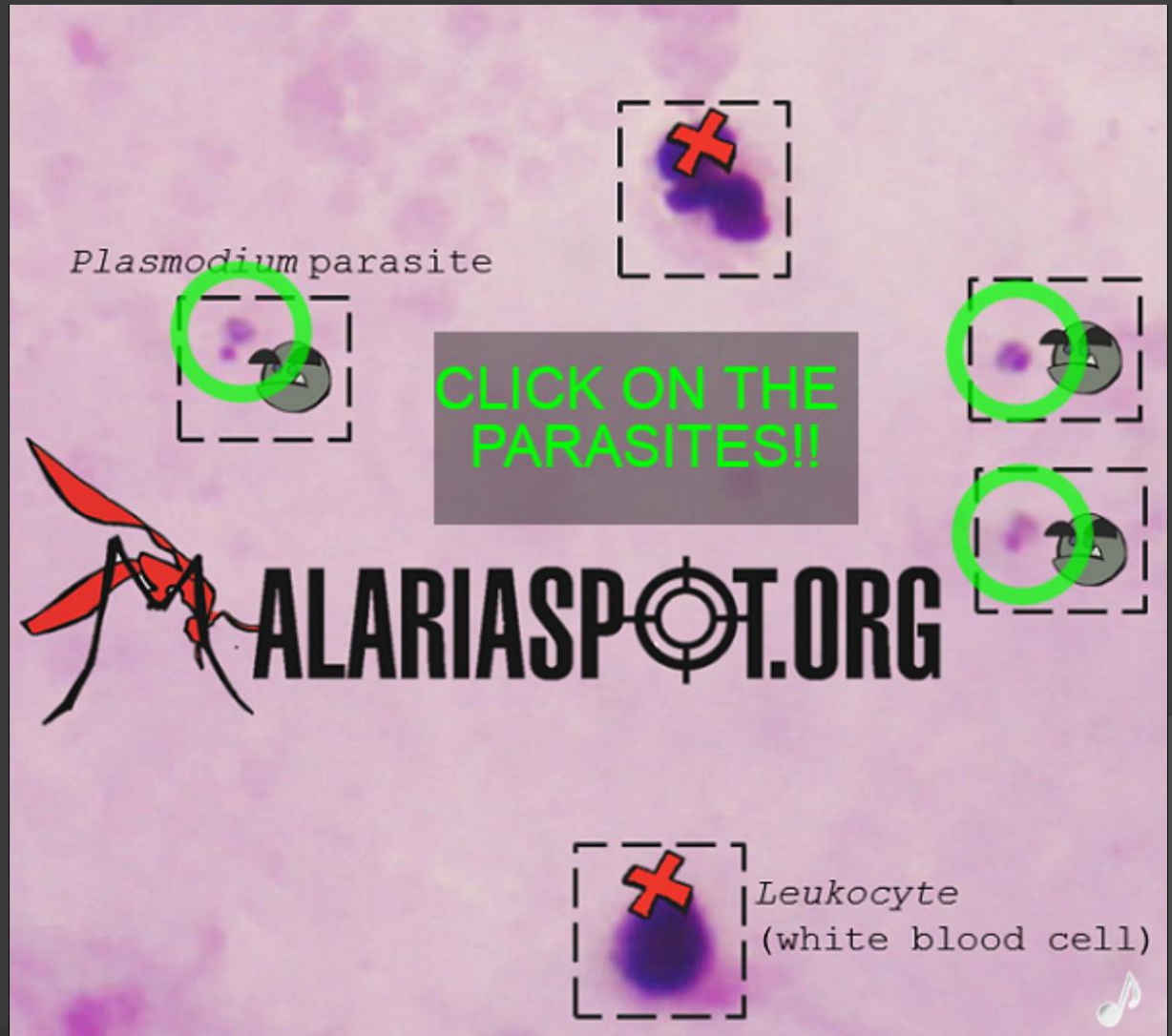
There is one nightjar hidden in every photo, touch it as soon as you see it



Start playing

Some GWAPs in Medicine

- Malaria Spot



Some GWAPs in Arts

Artigo

MENTIONS LÉGALES PROFIL DÉCONNEXION


ARTigo

À PROPOS D'ARTIGO
BLOG / [f](#) / [t](#)

AU TABLEAU D'HONNEUR

→ JOUER À NOUVEAU

→ AU TABLEAU D'HONNEUR



Artiste
Giovanni de Fonduli

Titre
Johannes der Täufer

Lieu
Turin / Privatsammlung

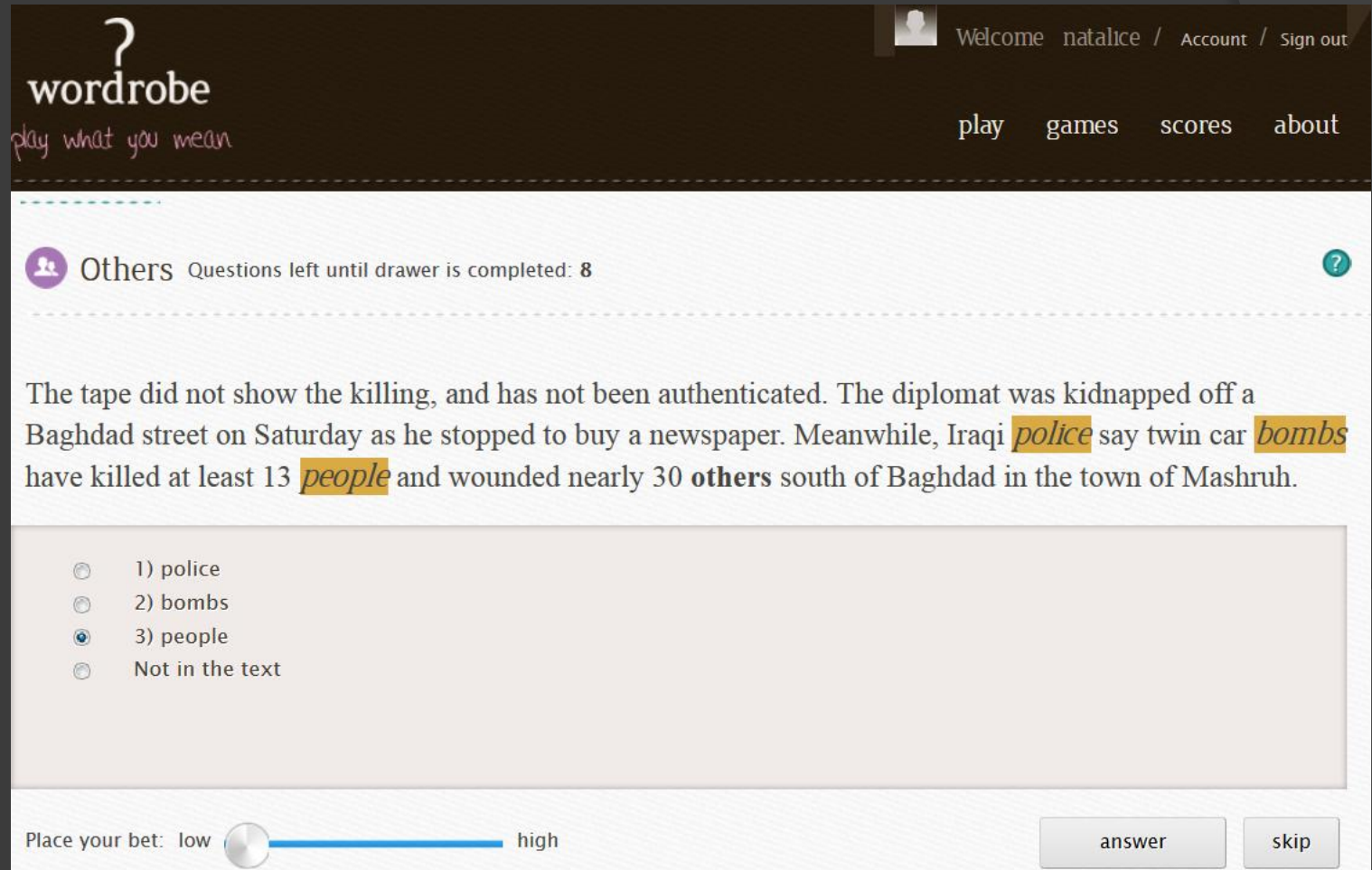
Datation
1470/1480

Score
30

Mots-clés déjà donnés
STATUE
HOMME
SCULPTURE
LIVREE
BARBE

Some GWAPs in NLP

Wordrobe



The screenshot shows the Wordrobe website interface. At the top, there is a navigation bar with the logo "wordrobe" and the tagline "play what you mean". The user is logged in as "natalice" and can access "Account" or "Sign out". There are links for "play", "games", "scores", and "about".

The main content area displays a question from the "Others" category. The text of the question is: "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." The words "police", "bombs", and "people" are highlighted in yellow.

Below the text, there are four radio button options for the answer:

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

At the bottom, there is a "Place your bet" slider ranging from "low" to "high", and two buttons: "answer" and "skip".

Some GWAPs in NLP

◉ Zombilingo



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 incompetent 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."



0%

Jouer Essayer

Quant à Mr. Fogg, il **repose** 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**

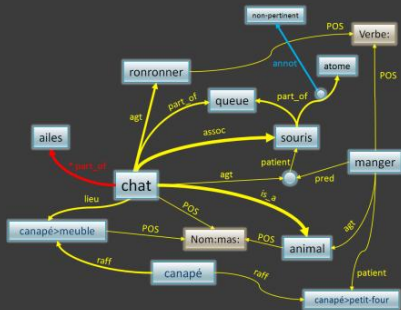
- ⦿ In general, short life span (many gwaps are dead before long)
- ⦿ Often the expected results are overestimated

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 Levrault**)

◎ 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

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

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

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

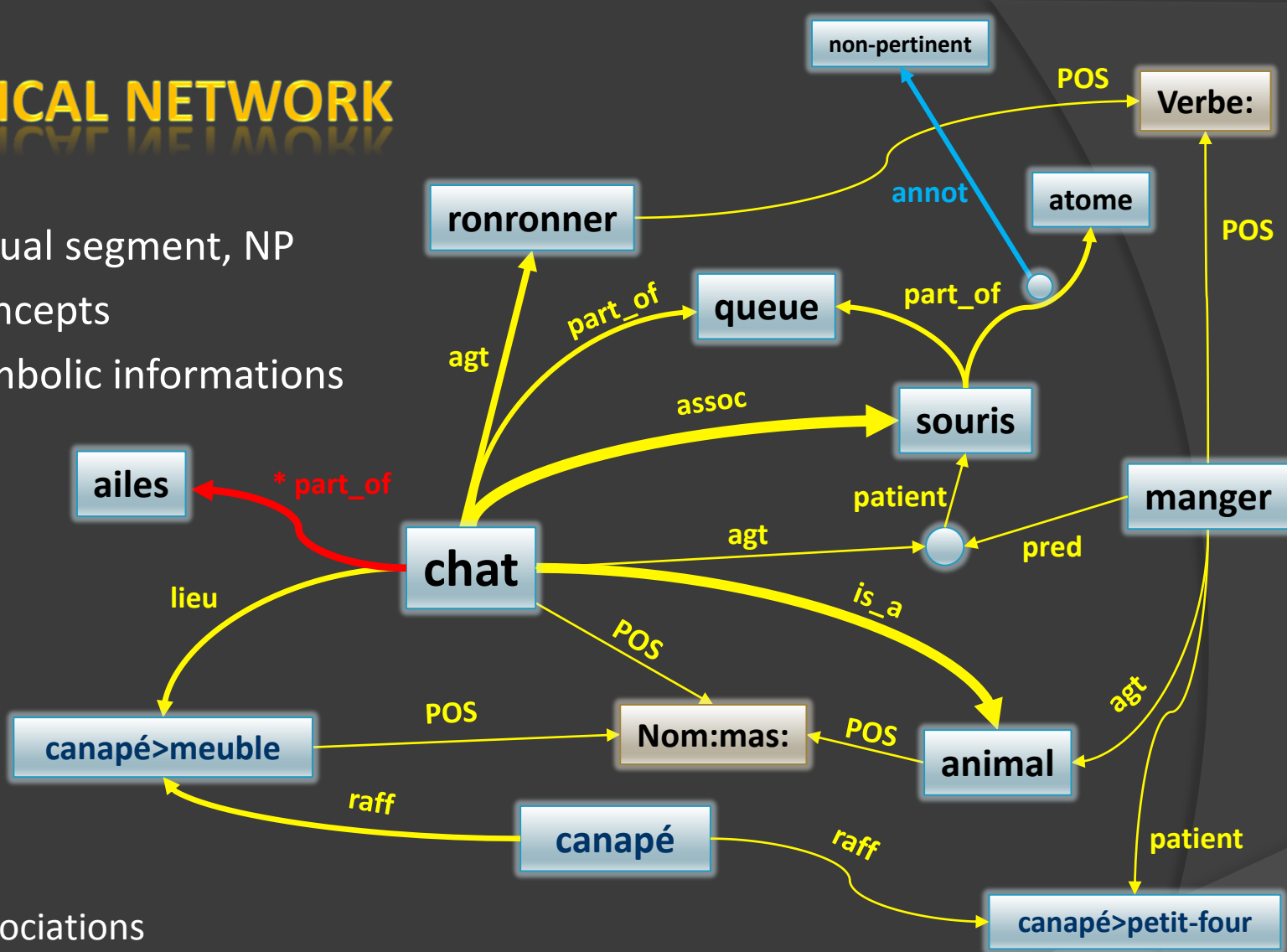
... AS A LEXICAL NETWORK

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 >
gobelin
gnome
farfadet
tourmenter
esprit follet
Bretagne
être fantastique
lutiner
petit >

15/28



JDM PLAY EXAMPLE

lutin

Niveau : 51



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

Réponses données par egailat : 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



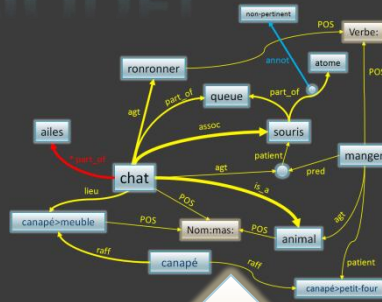
J'aime Soyez le premier de vos amis à indiquer que vous aimez ça.



RESULTS



INTERACTION MODEL



lexical network

game 1

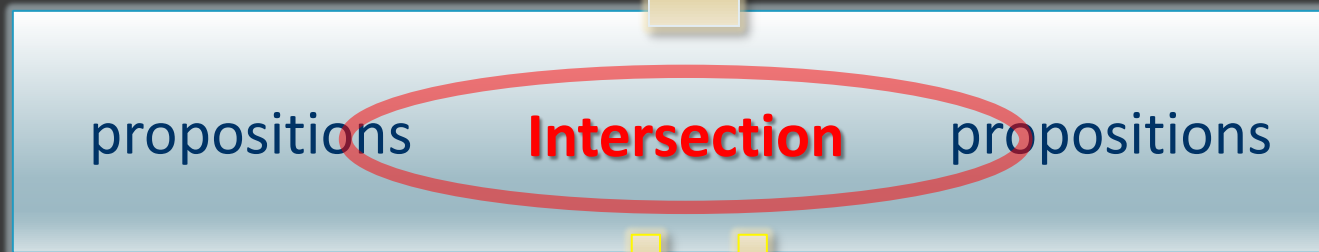
term
+
instruction

game 2

term
+
instruction

creation / strengthening
of relations

confrontation



player 1



rewards



player 2

PLAYER ACTIVITY REGULATION

Filtering - matching of player pairs

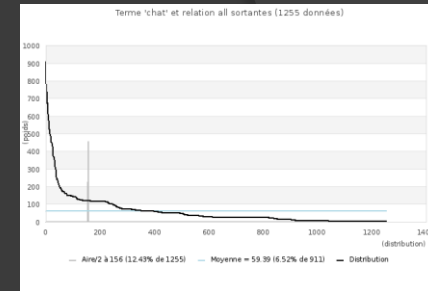
- Iterated Minimal Consensus (weighting)
- Minimizing noise, maximizing *recall* (long tail)

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 DILEMMA

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 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 je donne ma langue au chat

Il s'agissait de

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

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

REVERSED TABOU – FREQUENT TERMS



Can we find terms from the clue?

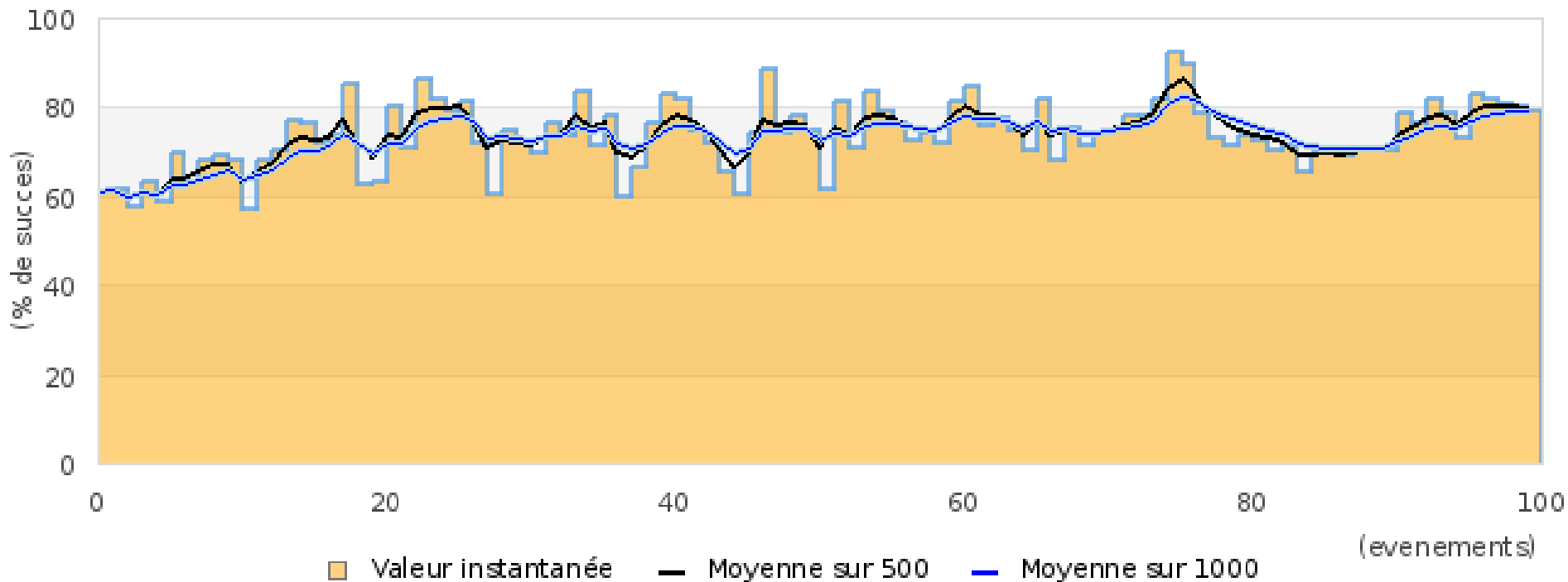
With the 500 riddles of the original game : AKI 494 – humans : 398

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

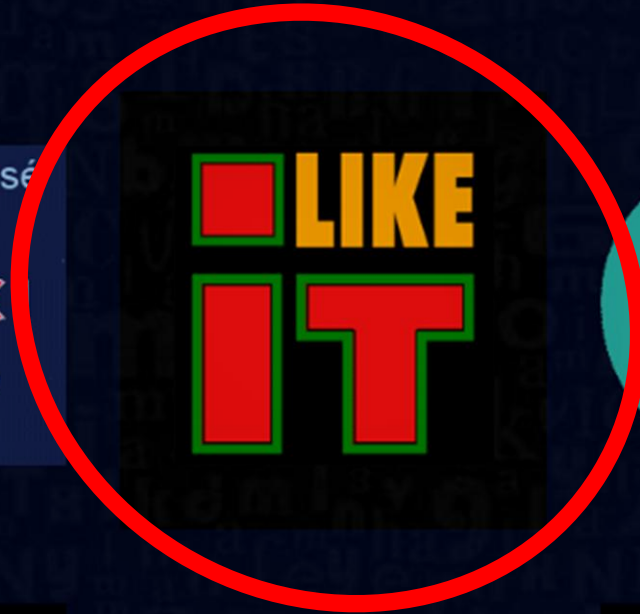


Counter games

Other games



SEXIT



ask
you



Counter games

LIKE

IT

A POLARITY GAME

Est-ce que vous aimez l'idée de

aller chez le dentiste



LIKE

IT

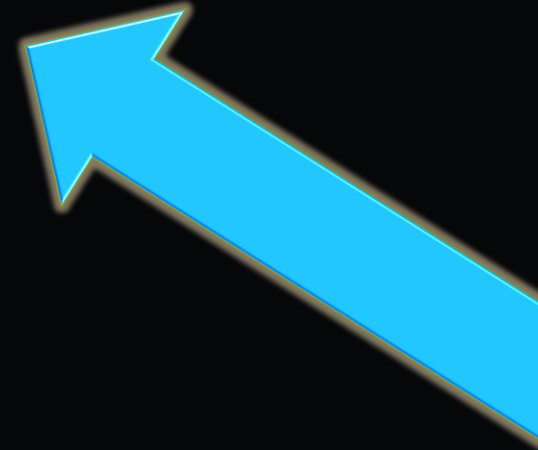
A POLARITY GAME

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

LIKE

IT

SOME RESULTS

657,843 polarized terms

- 551,871 **positive** polarity
- 366,913 **negative** polarity
- 655,003 **neutral** 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)



fracture du tibia

Nom Informations diverses wiki polarité

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 ▶ **Diagnostic(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 12 ▶ **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 21 ▶ 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 11 ▶ 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 22 ▶ **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

**fracture du tibia**

Nom Informations diverses wiki polarité

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 < 9 **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 > 12 **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** >

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**



Chercher la forme fracture du tibia



PI	ZO	VO	SE	GA
TA	CO	TE	PA	AS
CO2	UKN	CAR	AGT	ISA

tr	C	♥
n	F	R
...	D	32

fracture du tibia

Nom

- Natures grammaticales

-

- Informations sémantiques

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 **fracture** ▾ **tibia** ▾ médecine ▾ médecine (science) ▾ Médecine ▾ radiologie ▾ orthopédie ▾ traumatologie ▾ lésion physique ▾ fracture (lésion) ▾ lésion ▾ lésion osseuse **fracture tassement** ▾ **maladie** (médecine) ▾ **fatigue**

couture ▾ **panser** ▾ **maladie** ▾ **pansement** ▾ **plaie**

▾ **stress** ▾ **avoir mal** ▾ **médecine** (remède)

(sociologie) ▾ **urgences** ▾ **cassé** ▾ **hospitalisation**

▾ **convalescence** ▾ **AVP**

▾ **tibia** **péroné** ▾ **ski** ▾ **tibia** ▾ **genou** ▾ **fracture** ▾ **Genou** ▾ **cheville** ▾ **consolidation**

▾ **Fracture**

double fracture

fracture du péroné ▾ **fracture de la fibula** ▾

double fracture

fracture du péroné ▾ **fracture de la fibula** ▾

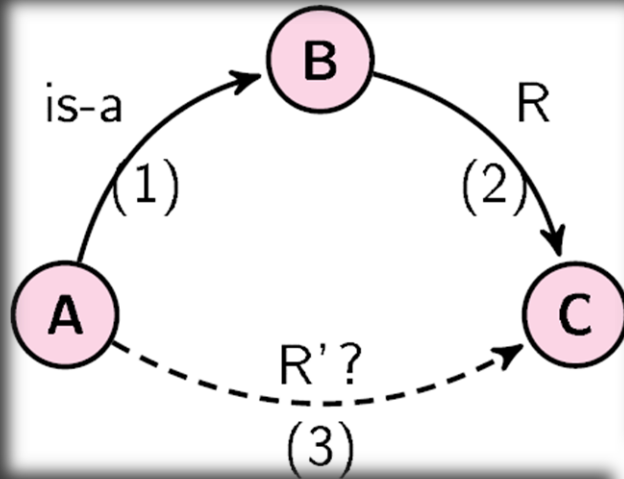
Oui mais
bi bi bla bla bla...
bi bi bla bla bla...



Oui mais
bié et bloblo,
non ?



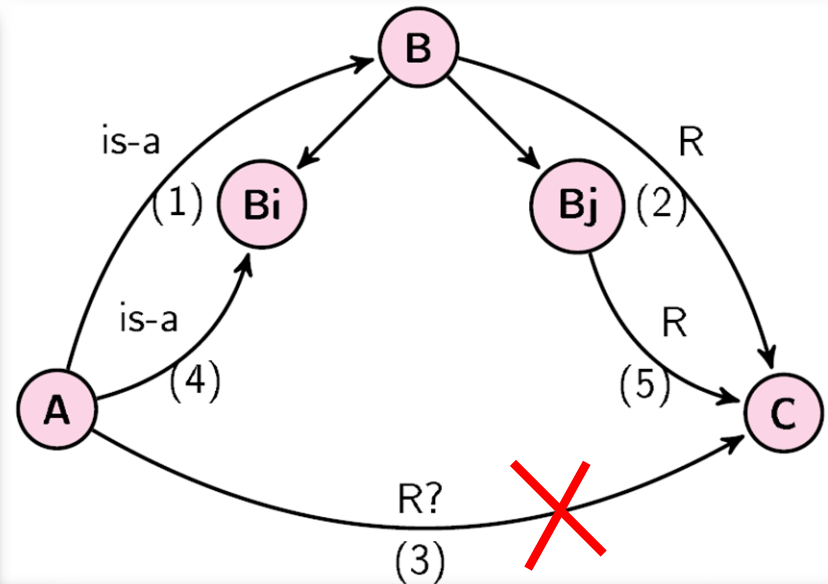
CONSOLIDATION AND CHECKING WITH INFERENCES - DEDUCTION



From the most
general toward the
most specific

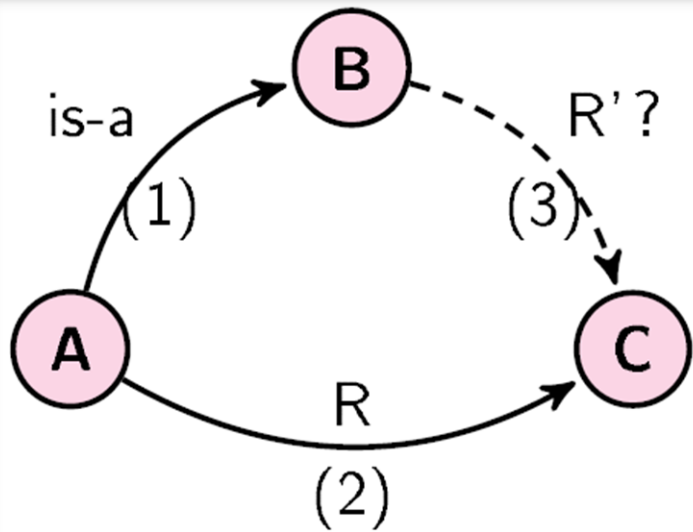
Logical and statistical blocking because of
polysemy - for example:

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

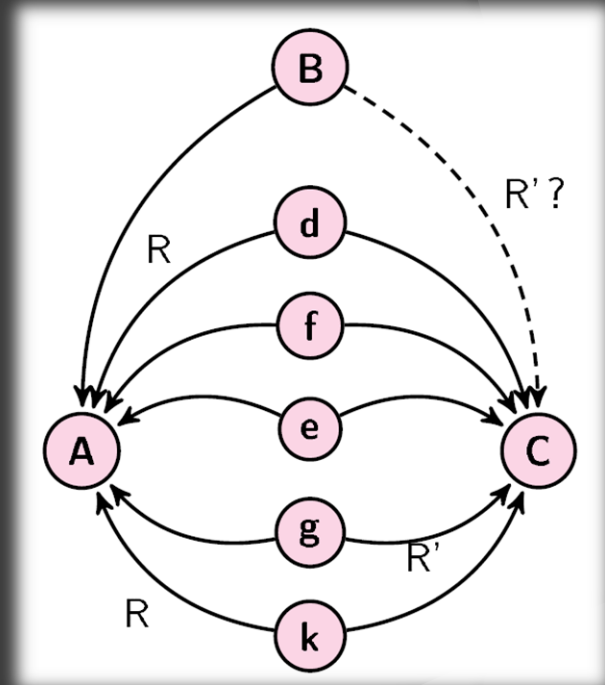


* Bible *is-a* livre & livre *carac* convertible
=> Bible *carac* convertible

CONSOLIDATION AND CHECKING USING INFERENCES – INDUCTION AND ABDUCTION



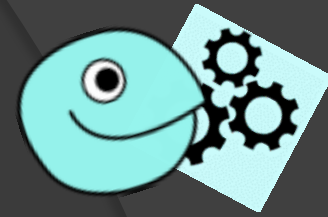
From specific to general



imitation of
examples

- The 3 inference types = **detector**
- of error in premises (1%)
 - of exceptions (< 1%)
 - of missing refinements (3%)
 - of irrelevant correct relations (3%)

About 93 % of the inferred relations are correct and relevant



SOME REFINEMENTS

◎ For polysemy and word usages

◎ `avocat --r-raff_sem--> avocat>fruit`

◎ `avocat --r-raff_sem--> avocat>justice`

◎ `grippe --r-raff_sem--> grippe>maladie`

◎ `grippe --r-raff_sem--> grippe>virus`

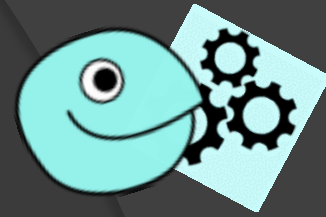
gloses

➤ 24 000 termes raffinés

➤ et > 66 000 usages

▶ what's specific?

SOME 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 les 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

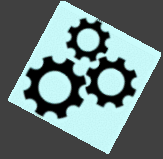
- ⊙ Inductive inferences potentially relevant, but wrong

- 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



INHIBITORY RELATIONS

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

> 370 000 inhibitory
relations in the
lexical network

chat (félin)

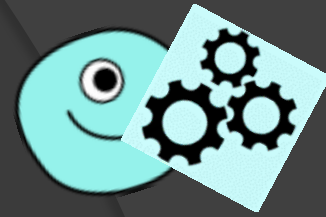
<-- r_inhib --

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)

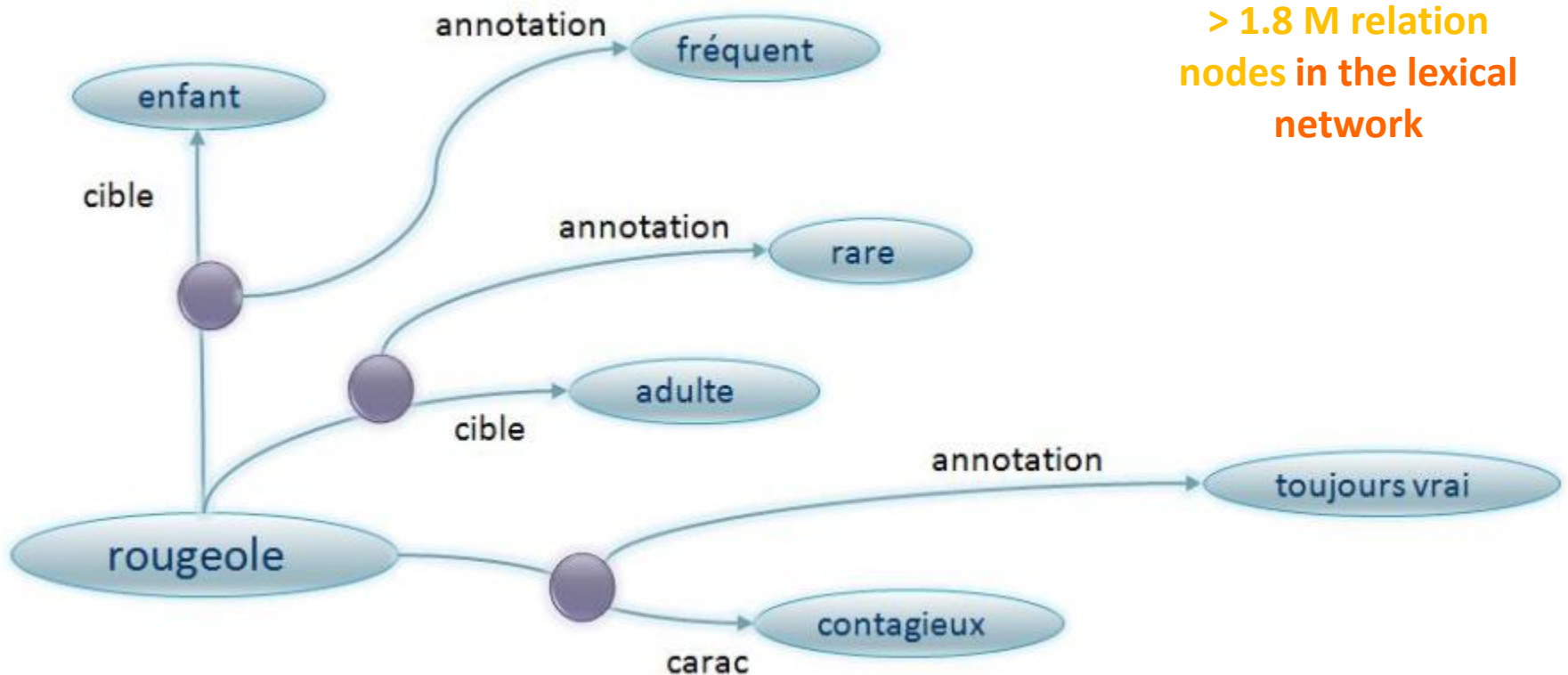
Negative relations can be used as inhibition
in WSD based on a thematic approach

▶ what's specific?

RELATION ANNOTATIONS



- ◎ Reification of a relation with new associated informations





AGGREGATIONS

◎ Another reification form

lion --r_agent-1 -> dévorer



lion [agent] dévorer

-- r_patient -> gazelle, zèbre



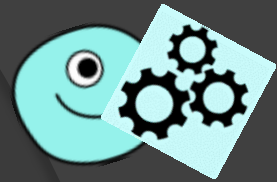
(lion [agent] dévorer) [patient] gazelle

-- r_action lieu -> savane

-- r_manner -> férocement

> 3366 aggregated
forms in the lexical
network

▶ 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

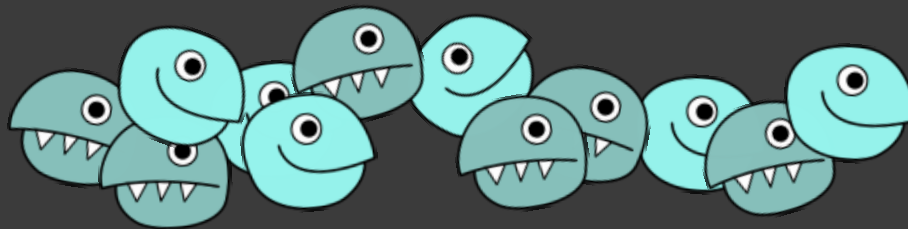
APPLICATIONS

- ◎ 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

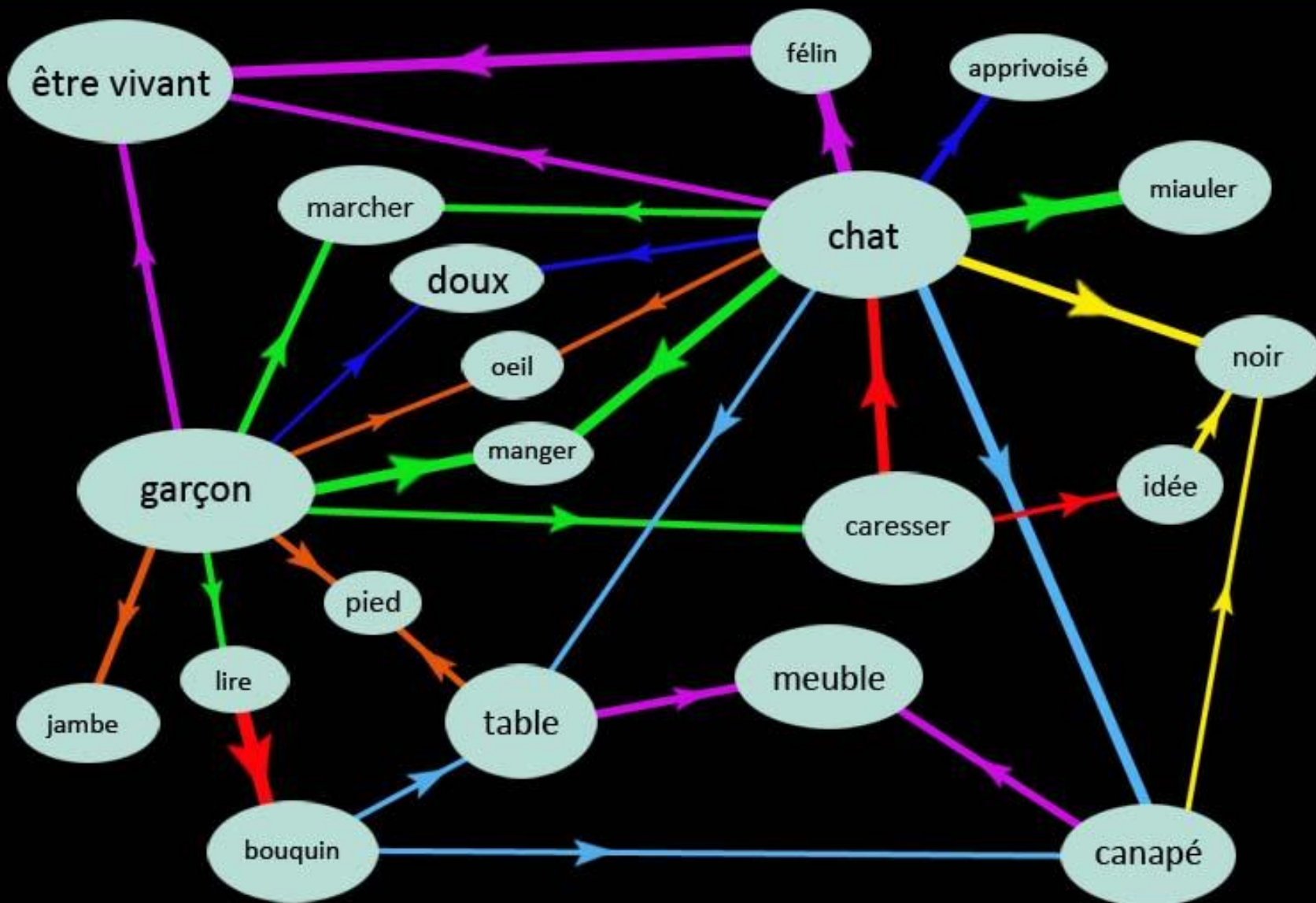
JEUXDEMOTS.ORG

DEMONSTRATIONS
QUESTIONS

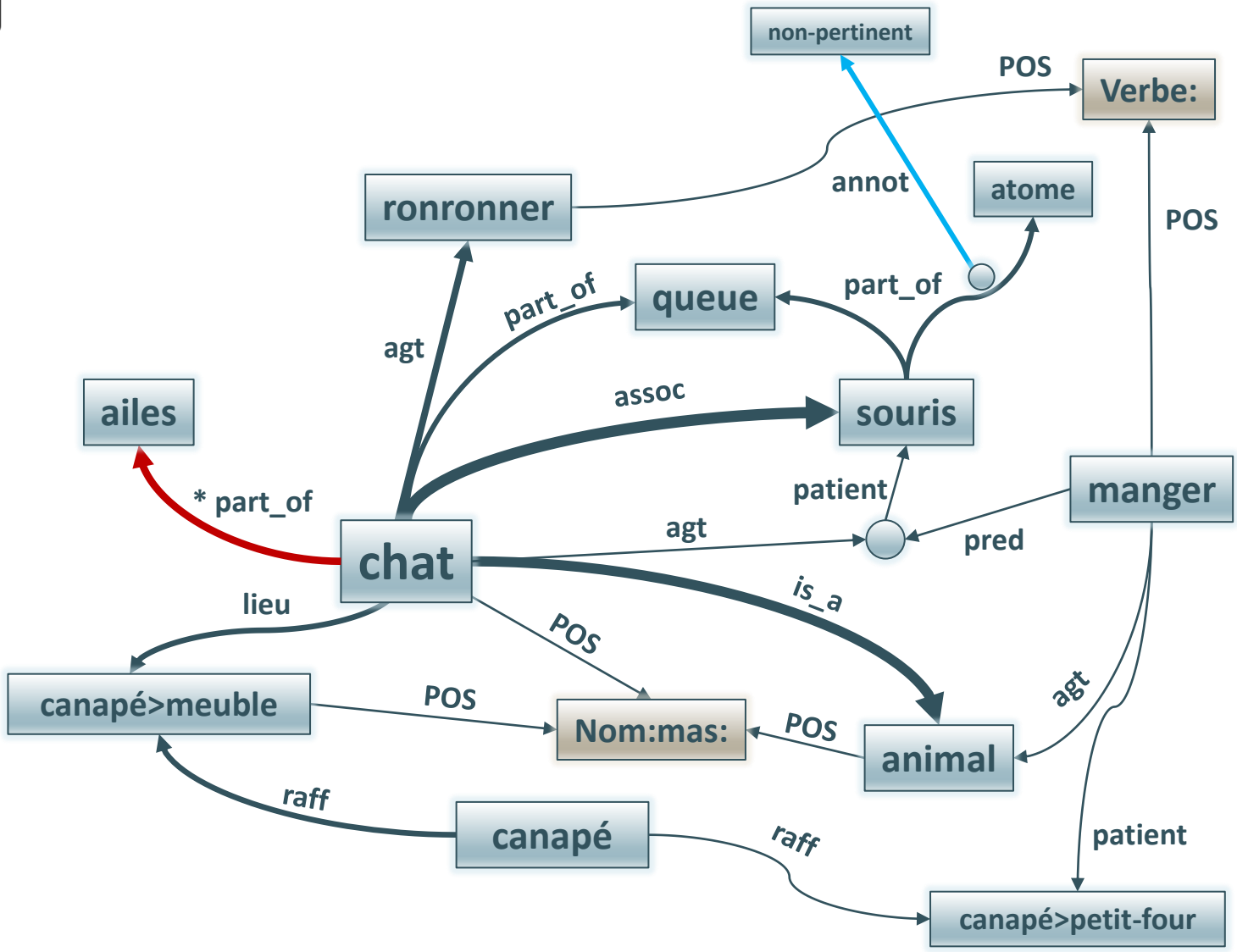


DETAIL OF RELATIONS DISTRIBUTION

6542950 r_associated (0) - 30.467 %	25083 r_raff_sem (1) - 0.117 %	33727 r_raff_morpho (2) - 0.157 %	188145 r_domain (3) - 0.876 %
1531831 r_pos (4) - 7.133 %	236897 r_syn (5) - 1.103 %	585836 r_isa (6) - 2.728 %	18417 r_anto (7) - 0.086 %
454401 r_hypo (8) - 2.116 %	640038 r_has_part (9) - 2.98 %	632055 r_holo (10) - 2.943 %	532905 r_locution (11) - 2.481 %
191855 r_fipot (12) - 0.893 %	254829 r_agent (13) - 1.187 %	43111 r_patient (14) - 0.201 %	101387 r_lieu (15) - 0.472 %
24636 r_instr (16) - 0.115 %	113062 r_carac (17) - 0.526 %	11298 r_data (18) - 0.053 %	78976 r_lemma (19) - 0.368 %
7168 r_magn (20) - 0.033 %	6644 r_antimagn (21) - 0.031 %	17113 r_family (22) - 0.08 %	102743 r_carac-1 (23) - 0.478 %
259958 r_agent-1 (24) - 1.21 %	10032 r_instr-1 (25) - 0.047 %	43098 r_patient-1 (26) - 0.201 %	186970 r_domain-1 (27) - 0.871 %
93092 r_lieu-1 (28) - 0.433 %	7469 r_chunk_pred (29) - 0.035 %	25166 r_lieu_action (30) - 0.117 %	25063 r_action_lieu (31) - 0.117 %
21873 r_sentiment (32) - 0.102 %	923 r_error (33) - 0.004 %	15213 r_maner (34) - 0.071 %	38718 r_meaning (35) - 0.18 %
1133563 r_infopot (36) - 5.278 %	6909 r_telic_role (37) - 0.032 %	3921 r_agentif_role (38) - 0.018 %	6632 r_verbe-action (39) - 0.031 %
6540 r_action-verbe (40) - 0.03 %	41072 r_conseq (41) - 0.191 %	41000 r_causatif (42) - 0.191 %	468 r_adj-verbe (43) - 0.002 %
460 r_verbe-adj (44) - 0.002 %	2813 r_chunk_sujet (45) - 0.013 %	3736 r_chunk_objet (46) - 0.017 %	0 r_chunk_loc (47) - 0 %
1245 r_chunk_instr (48) - 0.006 %	1632 r_time (49) - 0.008 %	5580 r_object>mater (50) - 0.026 %	622 r_mater>object (51) - 0.003 %
3075 r_successeur-time (52) - 0.014 %	4155 r_make (53) - 0.019 %	2634 r_product_of (54) - 0.012 %	1411 r_against (55) - 0.007 %
3397 r_against-1 (56) - 0.016 %	403 r_implication (57) - 0.002 %	635 r_quantificateur (58) - 0.003 %	362 r_masc (59) - 0.002 %
154 r_fem (60) - 0.001 %	1809 r_equiv (61) - 0.008 %	655 r_maner-1 (62) - 0.003 %	265 r_agentive_implication (63) - 0.001 %
121344 r_instance (64) - 0.565 %	67 r_verb_real (65) - 0 %	0 r_chunk_head (66) - 0 %	466 r_similar (67) - 0.002 %
83 r_set>item (68) - 0 %	202 r_item>set (69) - 0.001 %	280 r_processus>agent (70) - 0.001 %	4479 r_variante (71) - 0.021 %
1674 r_syn_strict (72) - 0.008 %	1406 r_bigger_than (73) - 0.007 %	1546 r_smaller_than (74) - 0.007 %	554 r_accomp (75) - 0.003 %
435 r_der_morpho (99) - 0.002 %	1489 r_has_auteur (100) - 0.007 %	80 r_has_personnage (101) - 0 %	241 r_can_eat (102) - 0.001 %
6882 r_has_actors (103) - 0.032 %	52 r_deplac_mode (104) - 0 %	1 r_has_interpret (105) - 0 %	40742 r_color (106) - 0.19 %
215 r_cible (107) - 0.001 %	448 r_symptomes (108) - 0.002 %	744 r_predecesseur-time (109) - 0.003 %	466 r_diagnostique (110) - 0.002 %
112 r_predecesseur-space (111) - 0.001 %	37 r_successeur-space (112) - 0 %	7 r_social_tie (113) - 0 %	40 r_beneficiaire (150) - 0 %
28 r_descend_de (151) - 0 %	82108 r_aki (666) - 0.382 %	5777889 r_wiki (777) - 26.905 %	6 r_annotaion_exception (997) - 0 %
1005096 r_annotaion (998) - 4.68 %	48238 r_inhib (999) - 0.225 %	0 r_prev (1000) - 0 %	0 r_succ (1001) - 0 %
81 r_termgroup (1002) - 0 %	6 r_learning_model (2001) - 0 %		



action	→ objet	sujet	→ action	terme	→ générique	terme	→ couleur
terme	→ caractéristique	terme	→ lieu	terme	→ partie		



A typed directed and weighted lexical network: example