

Approche hybride - lexicale et thématique -
pour la modélisation, la détection
et l'exploitation
des fonctions lexicales
en vue d'une analyse sémantique de texte

Didier Schwab
LIRMM-INFO-TAL

Approche hybride - lexicale et thématique - pour la modélisation, la détection et l'exploitation des fonctions lexicales en vue d'une analyse sémantique de texte

GÉRARD SABAH
CHRISTIAN BOITET
CHRISTOPHE LECERF
ROLAND DUCOURNAU
VIOLAINE PRINCE
MATHIEU LAFOURCADE

Directeur de recherche au CNRS, LIMSI-CNRS, Orsay (rapporteur)
Professeur à l'Université Joseph Fourier, CLIPS-IMAG équipe GETA (rapporteur)
Professeur à l'EMA (examineur)
Professeur d'Informatique, UMII (examineur)
Professeur d'Informatique, UMII (directrice)
MCF Informatique, UMII (co-directeur)



Traitement Automatique du Langage Naturel

*Domaine d'étude des techniques automatiques
d'analyse (compréhension) et de génération (production)
d'énoncés oraux ou écrits.*



Traitement Automatique du Langage Naturel

*Domaine d'étude des techniques automatiques
d'analyse (compréhension) et de génération (production)
d'énoncés oraux ou écrits.*



- Identification/résolution d'un ensemble de phénomènes sémantiques
- Obtention de représentations calculables
- Par le biais des Fonctions Lexicales

« Jacques m'avait donné de *précieux* *conseils*. »

Bon

« Il *voit* la fille avec un *télescope*. »

Instr

« Jean a eu une *peur* *bleue*. »

Magn

« Le *chat* est monté sur la chaise. *L'animal* s'assoupit. »

Gener



FL formalisent les relations linguistiques entre les termes

[Mel'čuk]

- **FL paradigmatiques (relations sémantiques)**

synonymie **Syn**('avion') = 'aéronef', 'aéroplane', ...

antonymie **Anti**('certitude') = 'incertitude', 'doute', 'scepticisme'

générique **Gener**('truite') = 'poisson' **Gener**('tractopelle') = 'engin'

Gener('chat') = 'animal' **Gener**('avion') = 'appareil'
≠ 'véhicule'

- **FL syntagmatiques (collocations)**

intensificateur **Magn**('peur') = 'bleue', 'grande'

Magn('amour') = 'fou', 'sans borne'

laudatif **Bon**('conseil') = 'précieux', 'bon'

Bon('choix') = 'heureux', 'bon'

confirmateur **Ver**('argument') = 'valable', 'recevable'

Ver('peur') = 'justifiée'

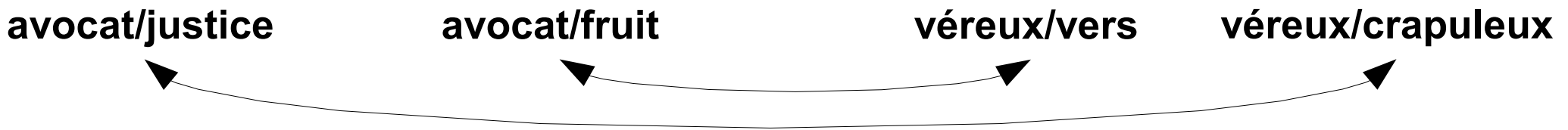


1) Ambiguïté lexicale

« *La **souris** mange le fromage.* »
souris/ordinateur ou *souris/animal*

2) Chemins interprétatifs

« *L'avocat est véreux.* » 2 interprétations probables et non 4





3) Référence

Résolution d'anaphore

« *L'homme* marcha sur la queue du *chien*, *il* aboya. »

Relations d'identité

« Le *chat* est monté sur la chaise. L'*animal* s'assoupit. »

4) Rattachement des groupes prépositionnels

« Il *voit* la fille *avec un télescope*. »



Recherche d'information

- Effets directs (égalité de valeurs)
 - « peur *bleue* » ≡ « *belle* peur »
 - « majorité *écrasante* » ≡ « *forte* majorité »
 - « Le *chat* est parti » ≡ « Le *matou* est parti »
 - « Ce nombre *n'est pas pair* » ≡ « Ce nombre est *impair* »
- Effets indirects (ambiguïté lexicale, rattach prep, références)
⇒ précision +, rappel +

Traduction Automatique

- Effets directs (transfert lexical)
 - « *grosse* fièvre » = « *high* fever »
 - « *grosse* pluie » = « *heavy* rain »
 - « L'*appareil* s'est posé. » ≡ « The *plane* has landed. »
- Effets indirects sur l'ensemble des phénomènes



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis

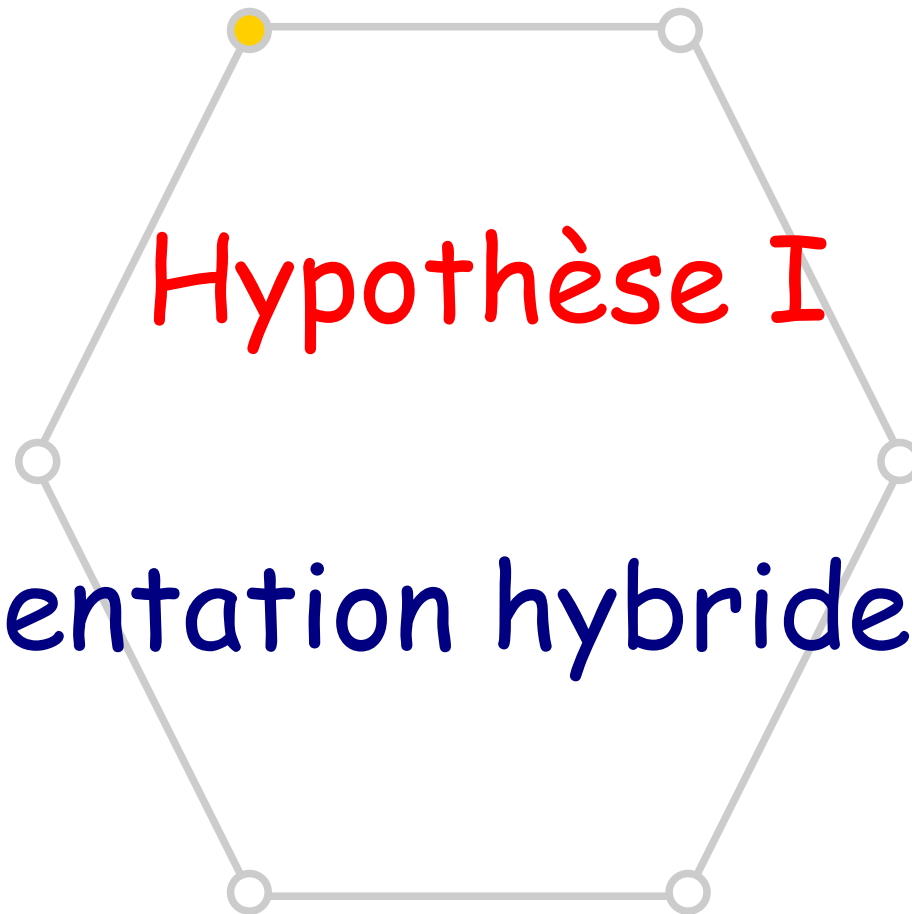


- Modélisation des fonctions lexicales

- Triple problème
 - Découverte d'un maximum d'items lexicaux
 - Acquisition des informations sur leur sens
 - Fabrication d'objets lexicaux représentant ces sens

- Triple questionnement
 - Comment représenter le sens ?
 - Comment le calculer ?
 - Comment obtenir un système générique et évolutif ?

- Quelles hypothèses avons-nous prises ?



Représentation hybride du sens



Pour les objets lexicaux

- Fonctions lexicales (discret, symboliques connexionnistes)
modélisation des relations entre OL
- Informations internes
 - symboliques
 - Morphologie (*nom, adj, verbe, masc, fem, ...*)
 - Informations étymologiques, niveau de langue, domaine, ...
 - numériques
 - Fréquence en usage
 - vectérielles
 - Informations thématiques (vecteurs conceptuels)



Représentation thématique [Chauché, Lafourcade]

Item lexical \Rightarrow Idées = Vecteur conceptuel

Par exemple, 873 composantes
(concepts issus du thésaurus Larousse)

(1) existence, (2) inexistence, (3) matérialité, ..., (516) liberté, ...,
(872) jeux, (873) jouets

Une composante du vecteur correspond à l'activation d'un concept.

Combinaison de vecteurs

Opérations \Rightarrow interprétations linguistiques raisonnables

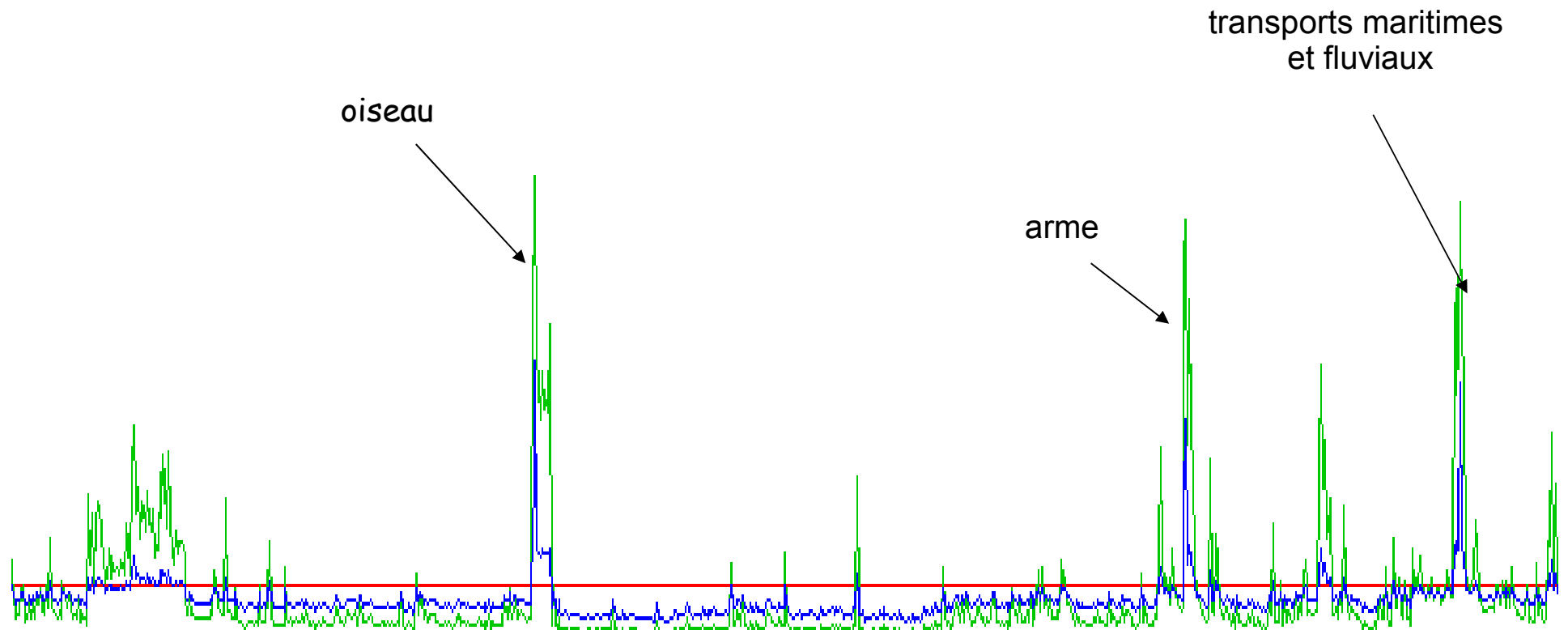
somme normée \oplus : union des idées

produit terme à terme \otimes : intersection des idées

contextualisation faible : $\gamma(A,B) = A \oplus (A \otimes B)$



Vecteur conceptuel de *frégate* (*polysémique*)





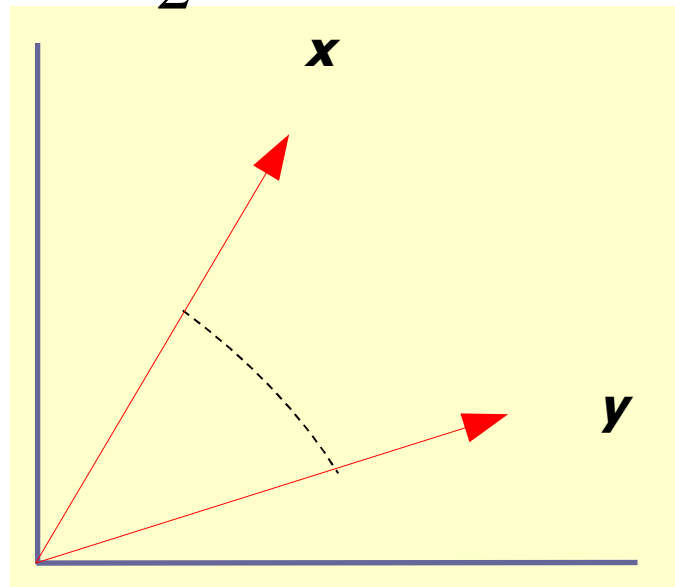
Distance thématique

$$D_A(x, y) = \text{angle}(x, y) = \arccos(\text{similarité}(x, y)) = \arccos\left(\frac{x \cdot y}{|x||y|}\right)$$

$$0 \leq D_A(x, y) \leq \frac{\pi}{2} \text{ (composantes positives)}$$

si 0 alors x et y sont colinéaires : même idée

si $\frac{\pi}{2}$: rien en commun





Distance thématique (exemples)

- $D_A('pélican', 'pélican') = 0 (0^\circ)$
- $D_A('pélican', 'grand-gosier') = 0,2 (11^\circ)$
- $D_A('pélican', 'train') = 1,22 (70^\circ)$
- $D_A('pélican', 'oiseau') = 0,46 (26^\circ)$
- $D_A('pélican', 'mouette') = 0,4 (23^\circ)$
- $D_A('pélican', 'poisson') = 0,35 (20^\circ)$



distance thématique \neq distance ontologique (de type *est-un*)
mais distance thématique \supset distance ontologique



- Limitation des VC dans la modélisation des fonctions lexicales
 - paradigmatiques
 - hyperonymie [Lafourcade et Prince, 2003]
 - synonymies (relative, subjective) [Lafourcade et Prince, 2001]
 - antonymies (complémentaire, scalaire, duale)
[COLING'2002, JADT'2002, TALN'2002]
 - syntagmatiques
 - collocations
- Allier le fort rappel des VC à la forte précision des relations
- Adéquation avec le modèle cognitif
 - 3 zones du cerveau
 - zone 1 : fabrication et classement des concepts
 - zone 2 : gestion de la "surface" du langage (syntaxe, associations lexicales)
 - zone 3 : combinaison des informations des 2 premières

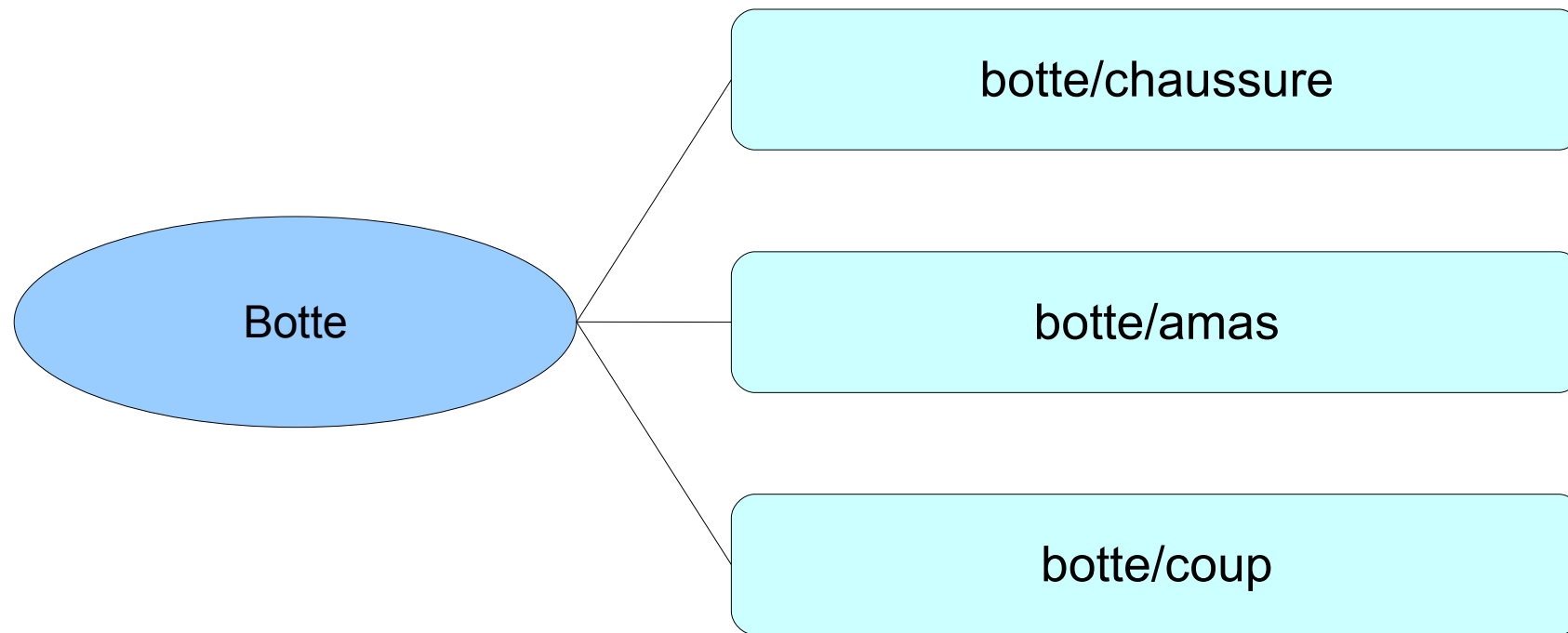


Hypothèse II

Utilisation conjointe
d'objets lexicaux de type
ACCEPTION et ITEM LEXICAL

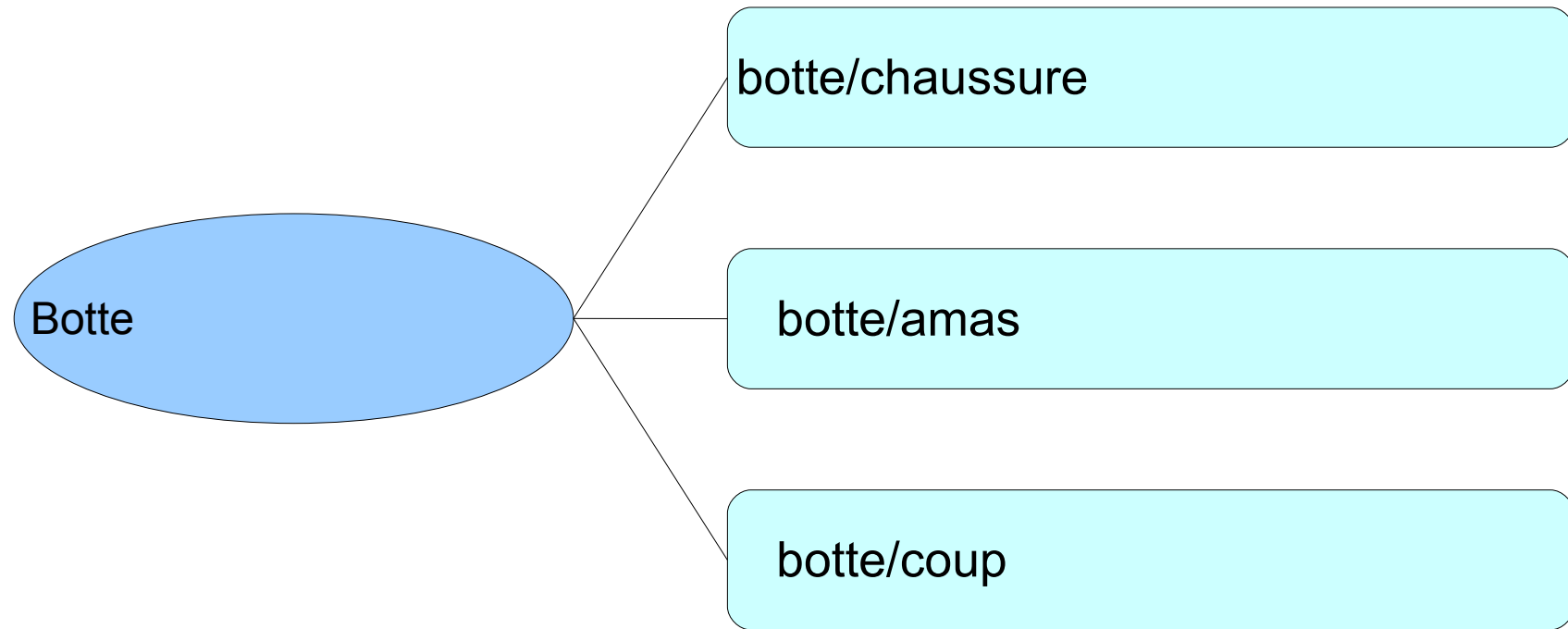


- Item lexical point d'entrée vers le sens
- Les termes sont **monosémiques** ou **polysémiques**
'cajou', 'neuroleptique', 'daucus carota', 'botte', 'lapin', 'carotte'
- **Acception** : sens particulier d'un item admis et reconnu par l'usage
- La compréhension du sens n'est pas que sélectionner l'acception mais aussi d'établir des relations entre structure de surface et structure profonde



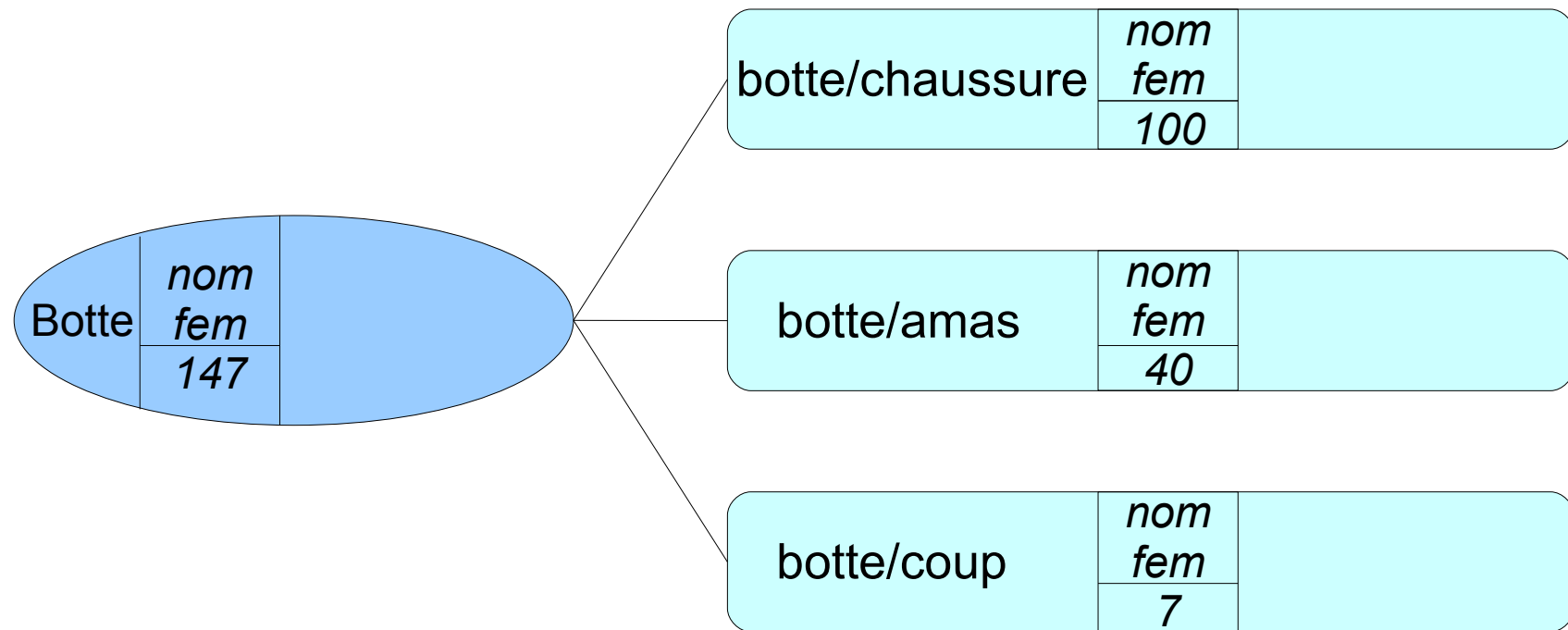
ITEM LEXICAL

ACCEPTIONS



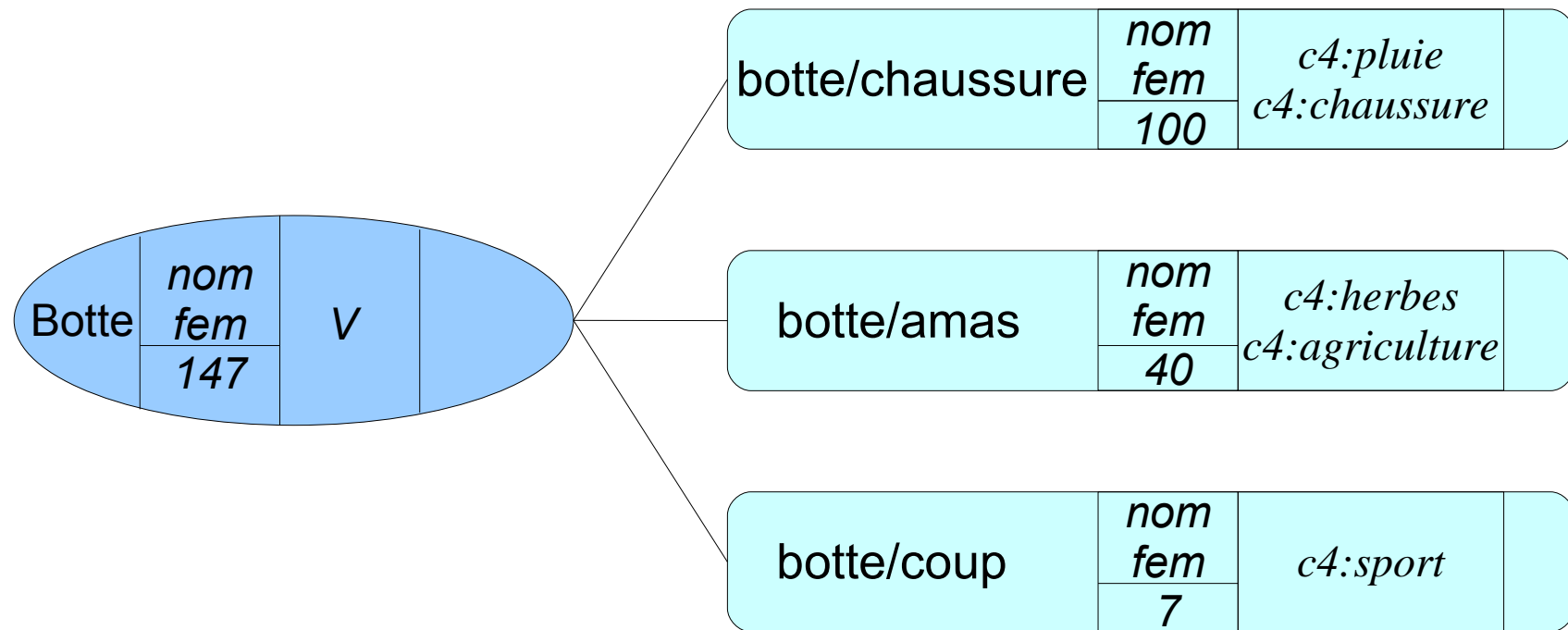
ITEM LEXICAL

ACCEPTIONS



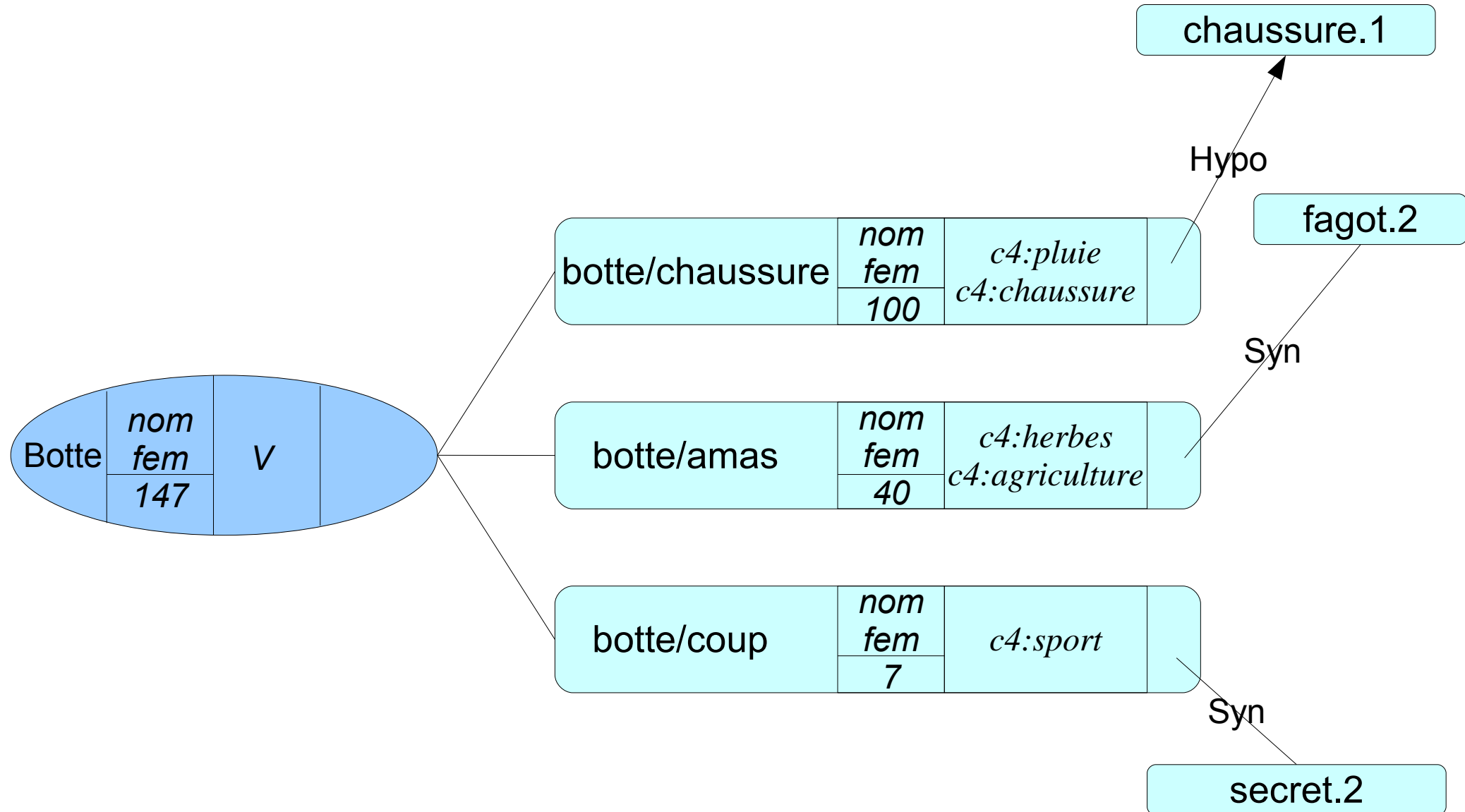
ITEM LEXICAL

ACCEPTIONS



ITEM LEXICAL

ACCEPTIONS





Hypothèse III

Génération automatique
des objets lexicaux



Objectif : construire une base de stockage d'objets
lexicaux ACCEPTIONS et ITEMS LEXICAUX

Pour le français, plus de 100 000 entrées, taux de
polysémie de 61%

5 définitions en moyenne, 400 000 objets lexicaux
Impossible à indexer manuellement



Comment ?

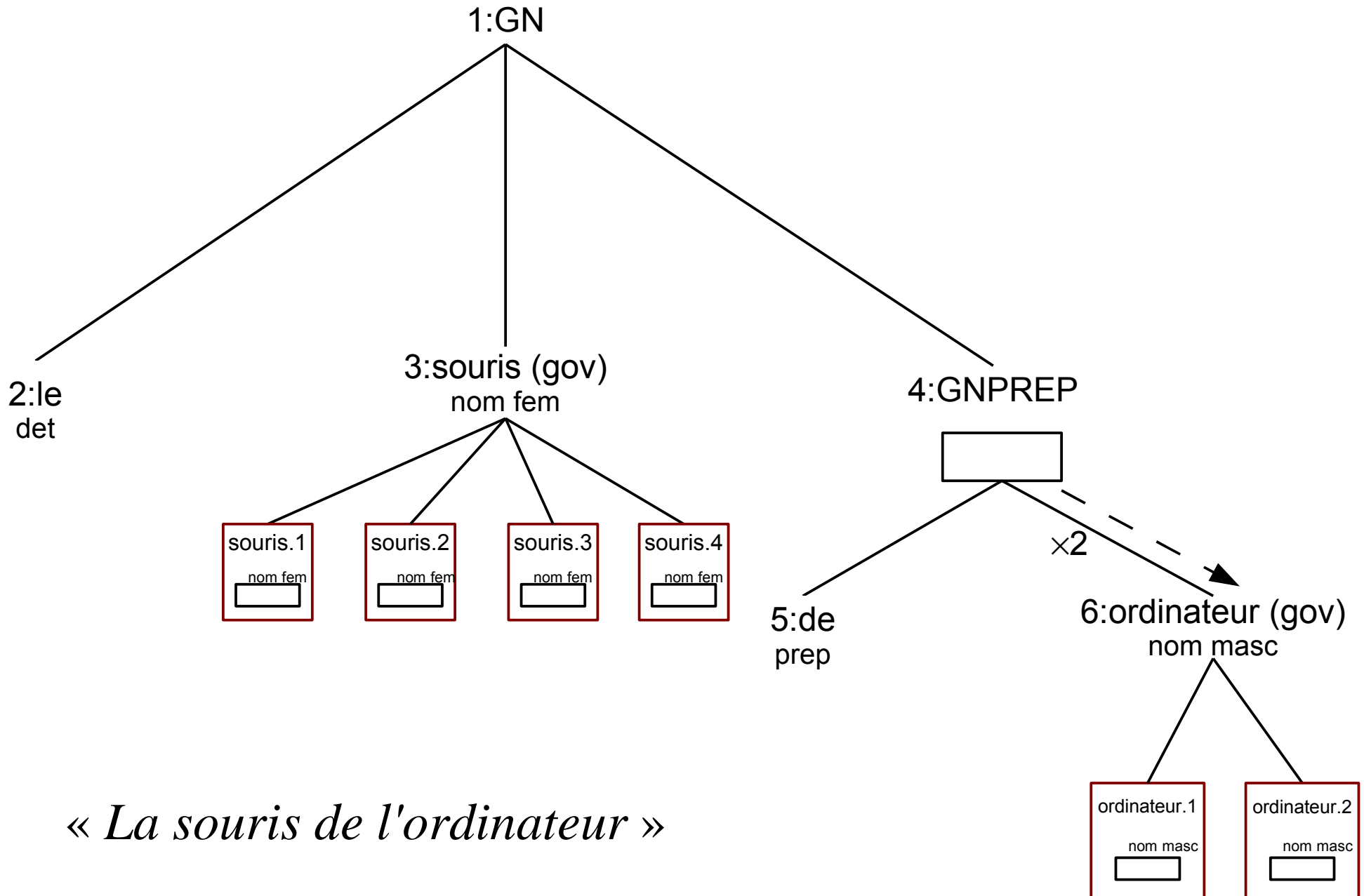
- à partir d'un noyau réduit de termes pertinents (1000-2000) indexés à la main
- indexation automatique des autres termes

Utilisation d'informations extraites de diverses sources

- dictionnaires (analyse sémantique)
- listes de synonymes (vecteurs + morphologie)
- listes d'antonymes (vecteur (fct d'antonymie) + morphologie)
- Web (sites d'informations, Google, ...)
- Corpus, ...



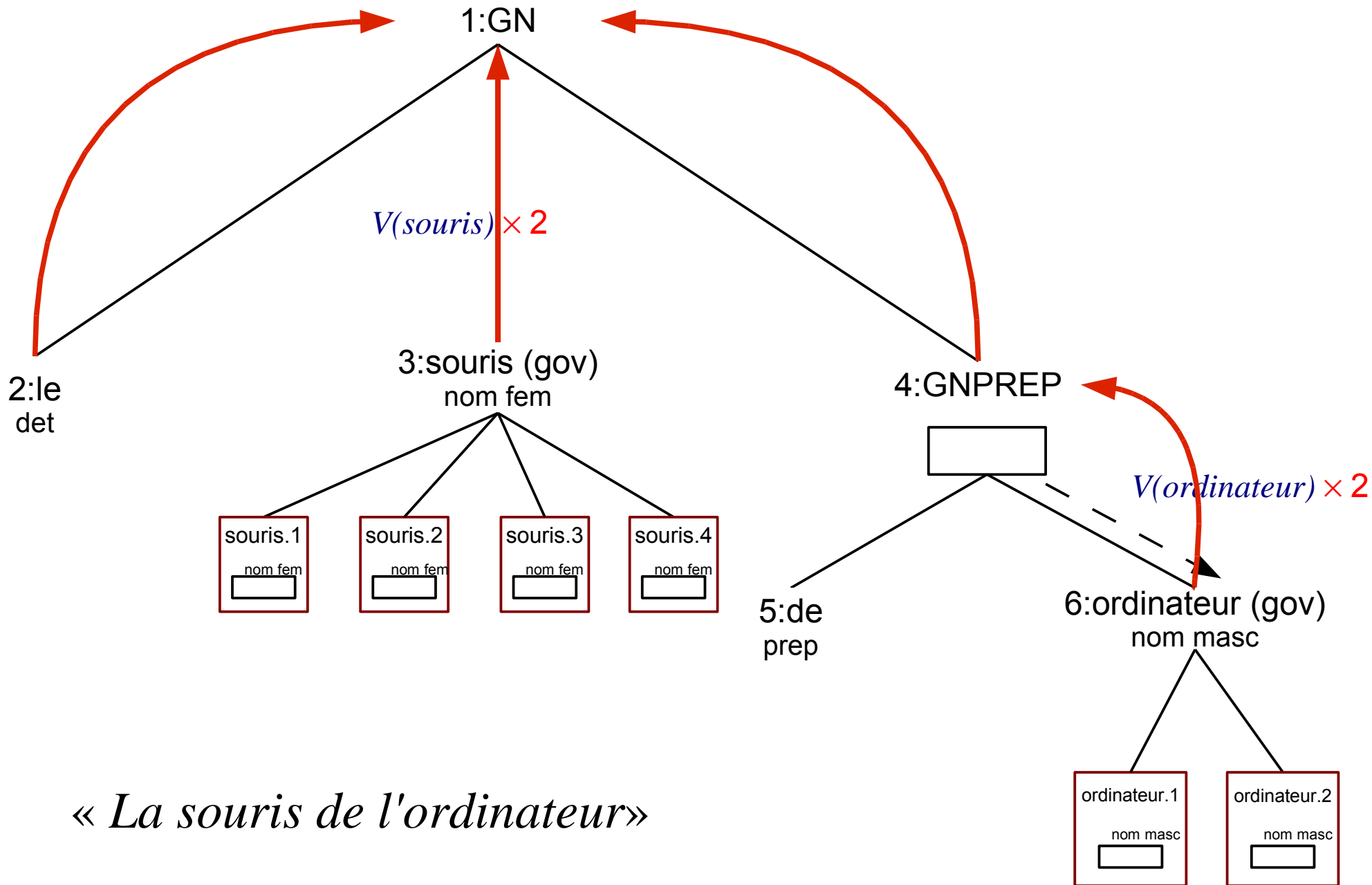
Analyse en remontée-descente



« *La souris de l'ordinateur* »



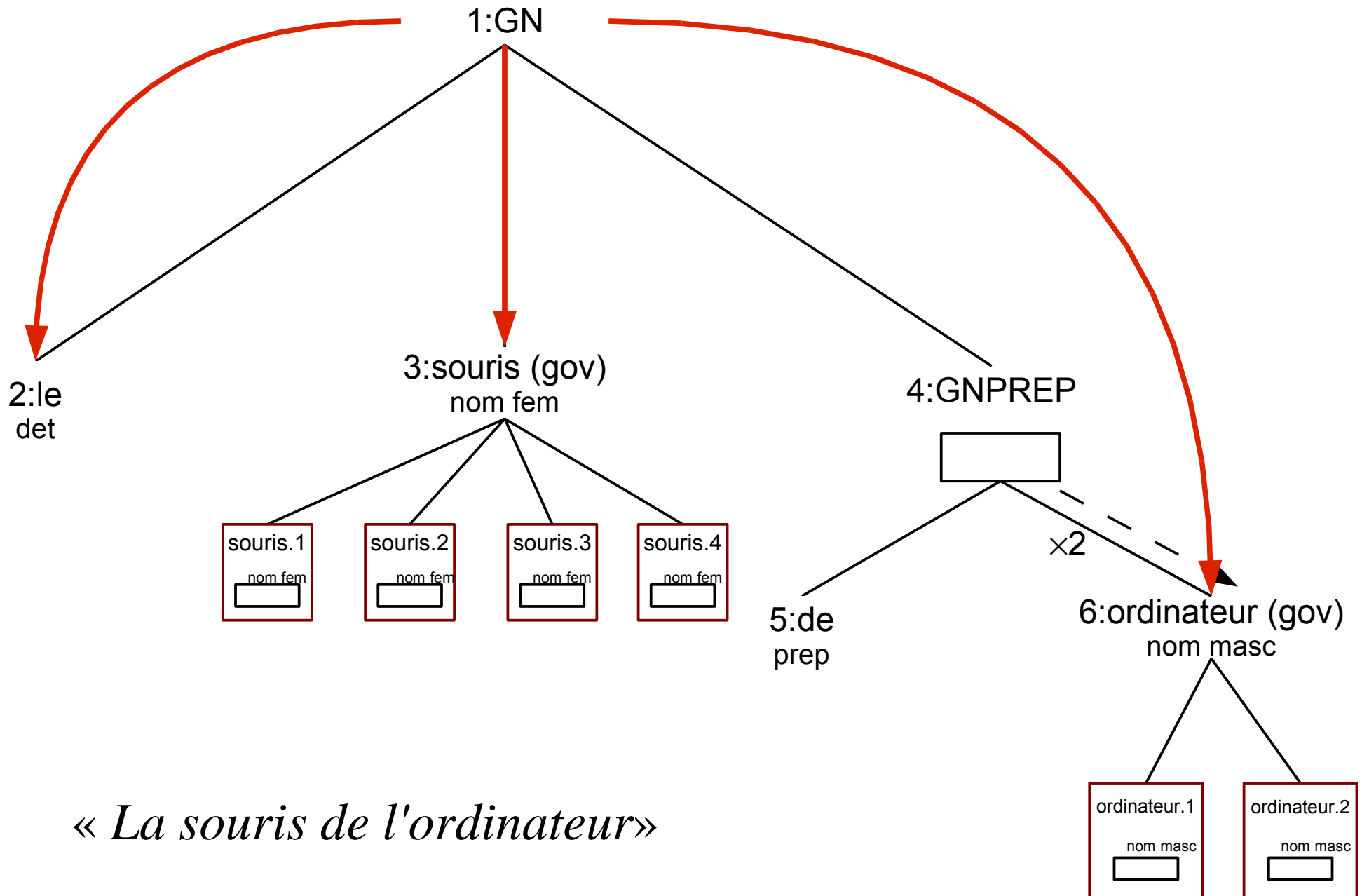
Analyse en remontée-descente



« *La souris de l'ordinateur* »

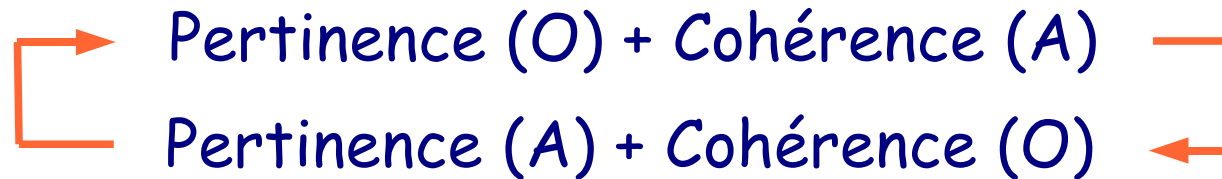


Analyse en remontée-descente

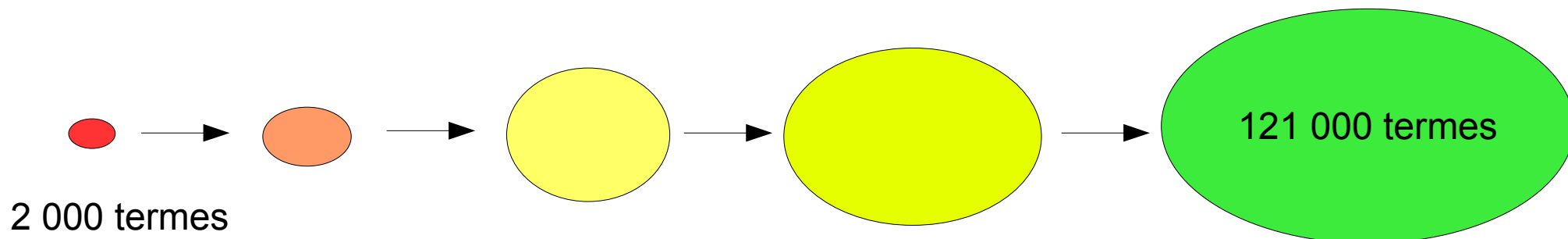


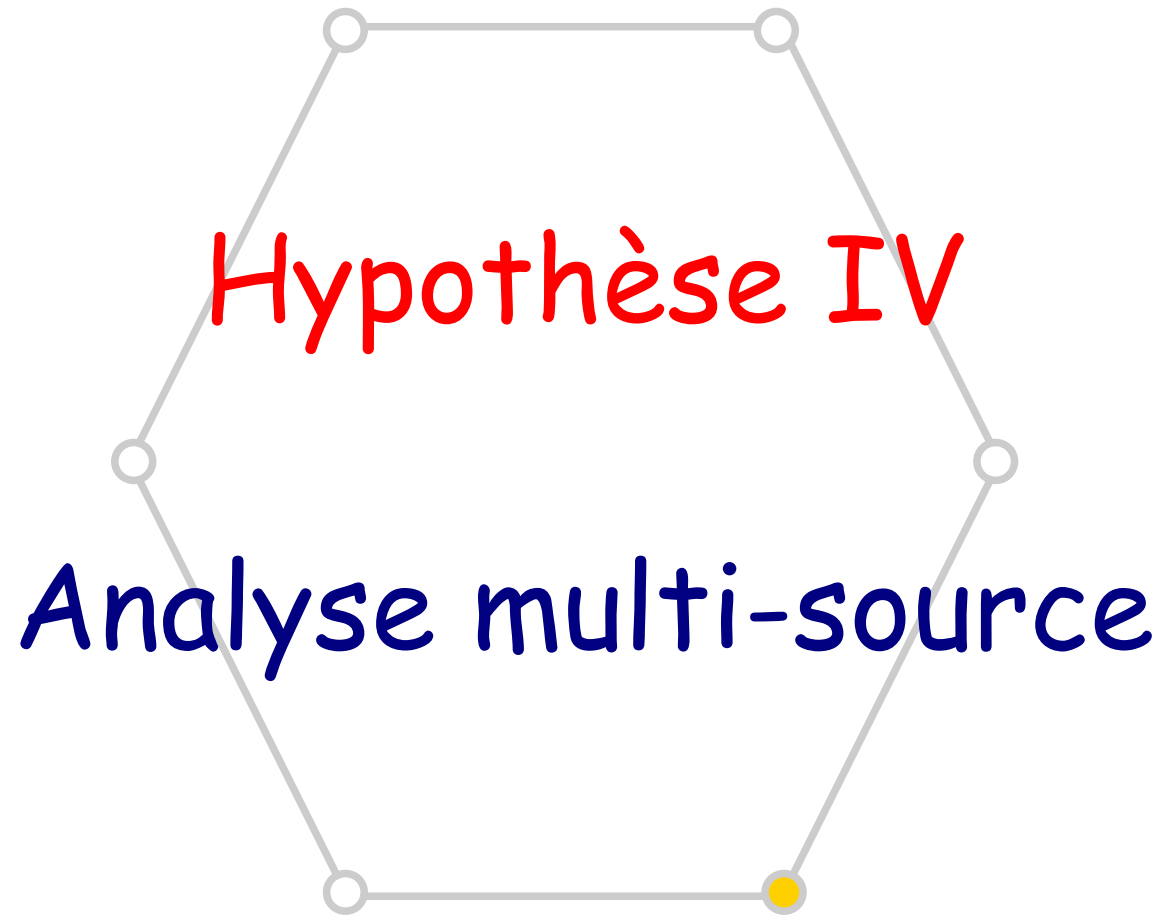


Le noyau d'objets lexicaux O est pertinent
l'apprentissage A doit être cohérent



Aujourd'hui : 121 000 termes indexés automatiquement







- **Métalangage** : *se dit de, terme qualifiant, ...*
 - *aboyer* : « Crier en parlant du chien. »
 - *anthropophage* : « Qui mange de la chair humaine en parlant de l'humain. »
- **Couverture du lexique**
 - évolution constante
 - « non-complétude » des dictionnaires
 - 'liturgiste' ∈ Robert
 - ∉ Larousse
- **Solution**
 - Construction d'une LEXIE par définition
 - LEXIE = atome de notre base



botte-1 : #nf# Réunion de végétaux de même nature liés ensemble. (Une botte de paille, de radis, de fleurs) . [Hach]

botte-2 : #nf# En escrime, coup porté à l'adversaire avec un fleuret ou une épée. (Pousser, porter, parer une botte) (Botte secrète.) . [Hach]

botte-3 : #nf# Chaussure de cuir, de caoutchouc ou de plastique qui enferme le pied et la jambe, parfois la cuisse. (Des bottes de cavalier) Chaussure d'extérieur basse. (Botte d'hiver, de ski, de marche) . [Hach]

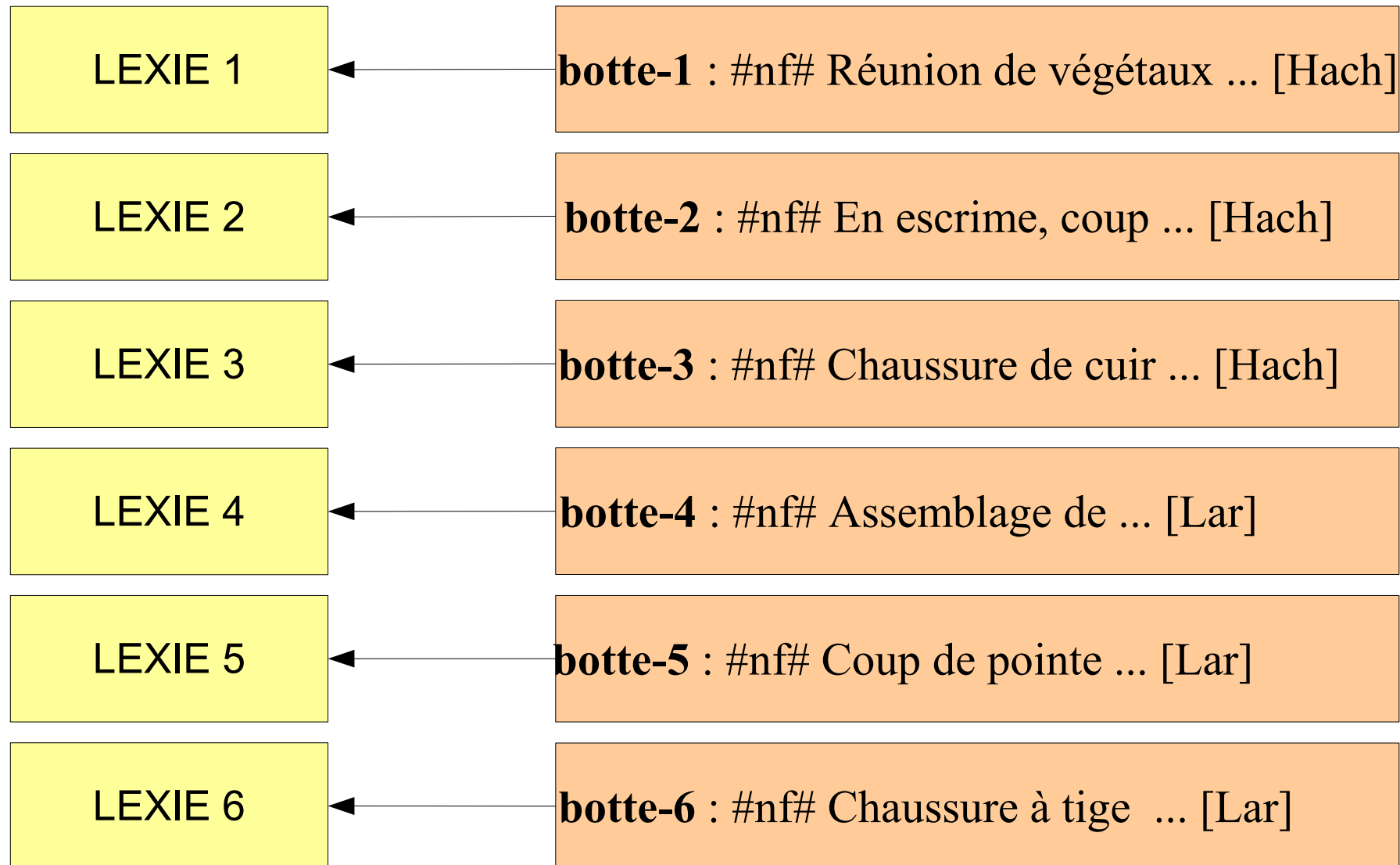
botte-4 : #nf# (néerl. bote, touffe de lin) . Assemblage de végétaux de même nature liées ensemble : (Botte de paille. Botte de radis.) . [Lar]

botte-5 : #nf# (#ethym-it# botta, coup) . Coup de pointe donné avec le fleuret ou l'épée . [Lar]

botte-6 : #nf# (p.-ê. de bot) . Chaussure à tige montante qui enferme le pied et la jambe généralement jusqu'au genou : (Bottes de cuir) . [Lar]



Recueil des informations lexicales et calcul des vecteurs conceptuels





Catégorisation [Jalabert, Lafourcade] [Schwab]

des sens en fonction

des informations

- morphologiques
- étymologiques
- lexicales
- vectorielles

botte.1

botte.2

botte.3

LEXIE 1

LEXIE 2

LEXIE 3

LEXIE 4

LEXIE 5

LEXIE 6

#nf# Réunion de végétaux ... [Hach]

#nf# En escrime, coup ... [Hach]

#nf# Chaussure de cuir ... [Hach]

#nf# Assemblage de ... [Lar]

#nf# Coup de pointe ... [Lar]

#nf# Chaussure à tige ... [Lar]



Catégorisation [Jalabert, Lafourcade] [Schwab]

des sens en fonction

des informations

- morphologiques
- étymologiques
- lexicales
- vectorielles

botte.1

botte.2

botte.3

LEXIE 1

LEXIE 2

LEXIE 3

LEXIE 4

LEXIE 5

LEXIE 6

#nf# Réunion de végétaux ... [Hach]

#nf# En escrime, coup ... [Hach]

#nf# Chaussure de cuir ... [Hach]

#nf# Assemblage de ... [Lar]

#nf# Coup de pointe ... [Lar]

#nf# Chaussure à tige ... [Lar]

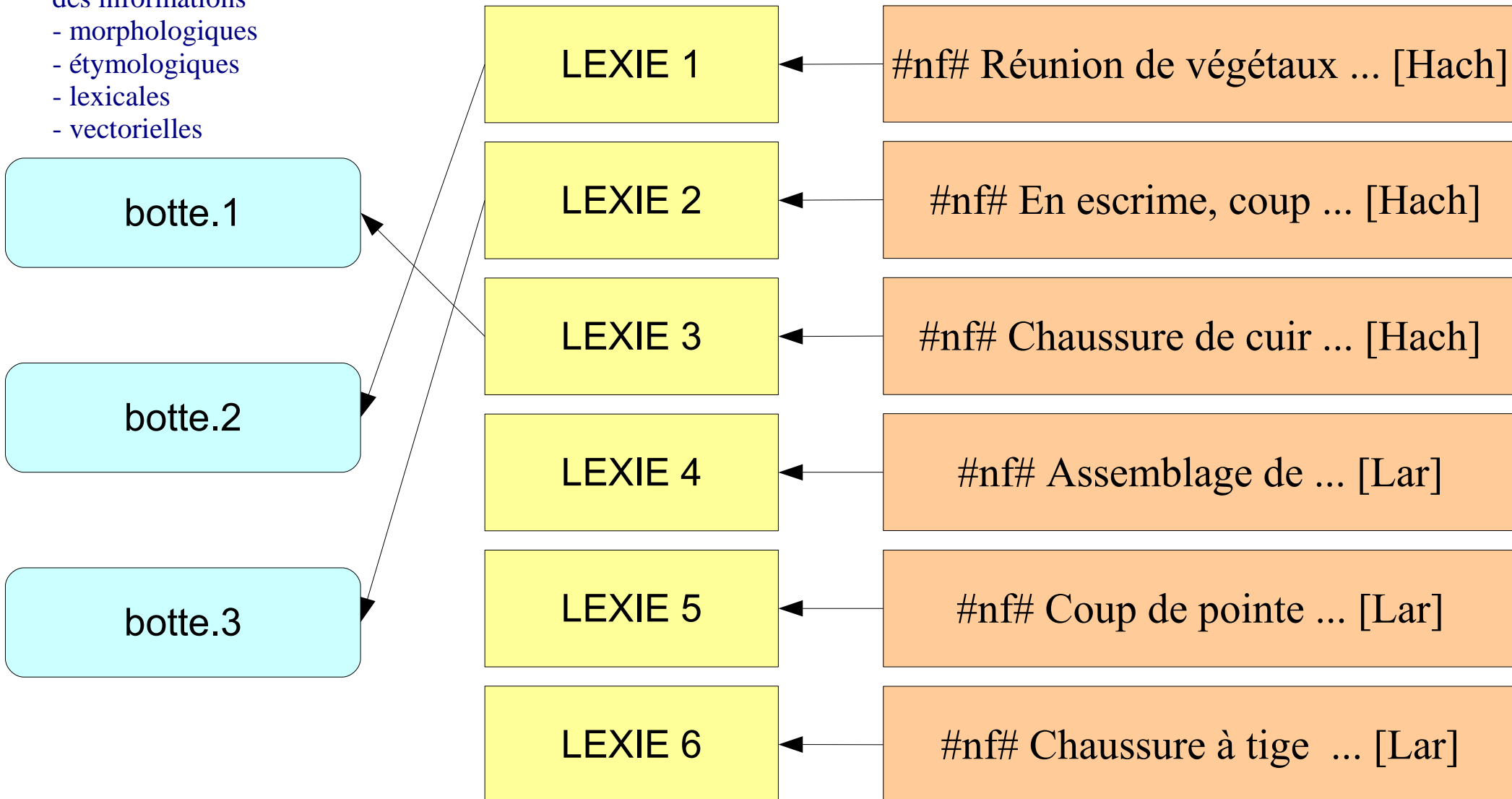


Catégorisation [Jalabert, Lafourcade] [Schwab]

des sens en fonction

des informations

- morphologiques
- étymologiques
- lexicales
- vectorielles



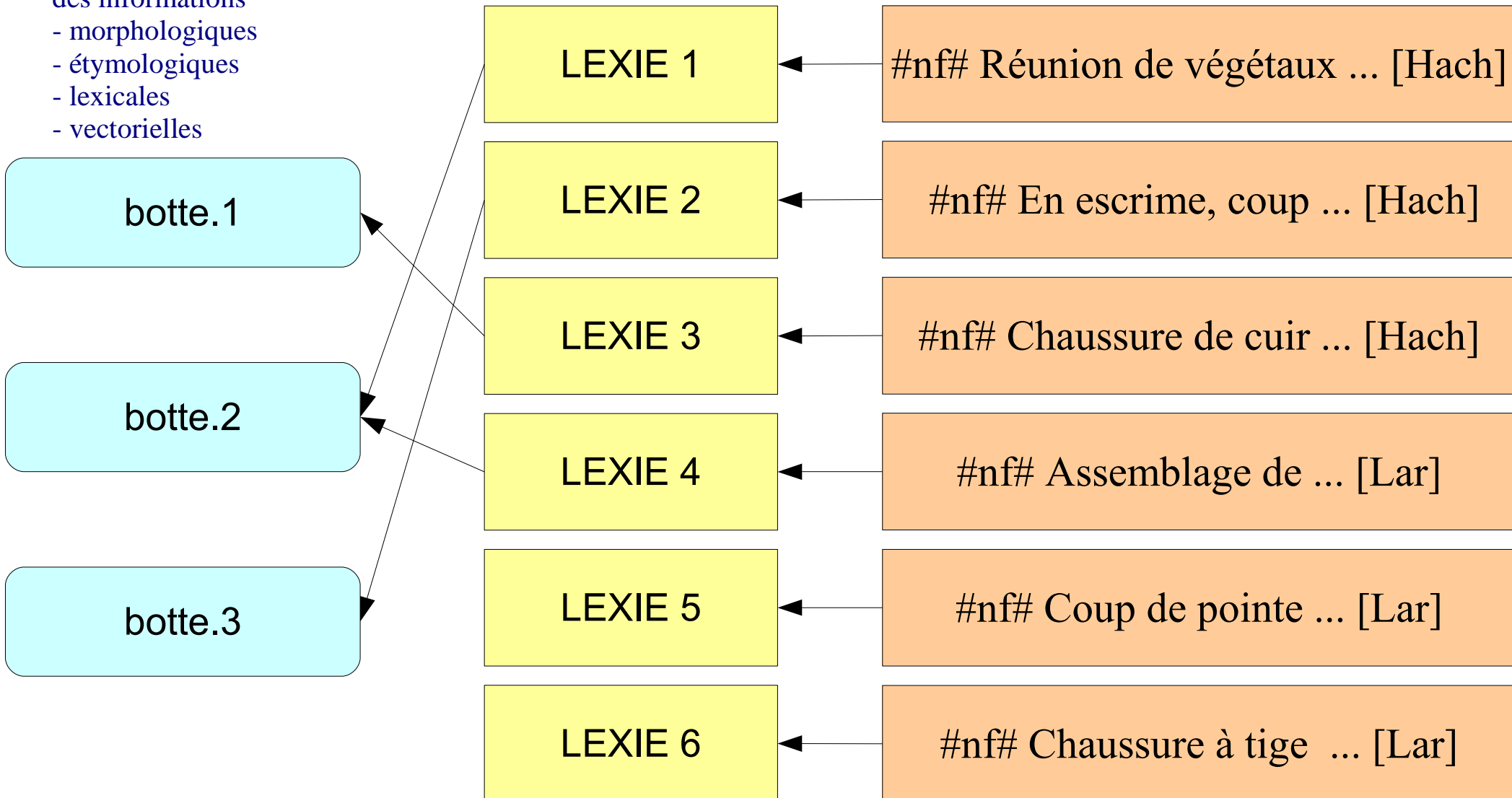


Catégorisation [Jalabert, Lafourcade] [Schwab]

des sens en fonction

des informations

- morphologiques
- étymologiques
- lexicales
- vectorielles



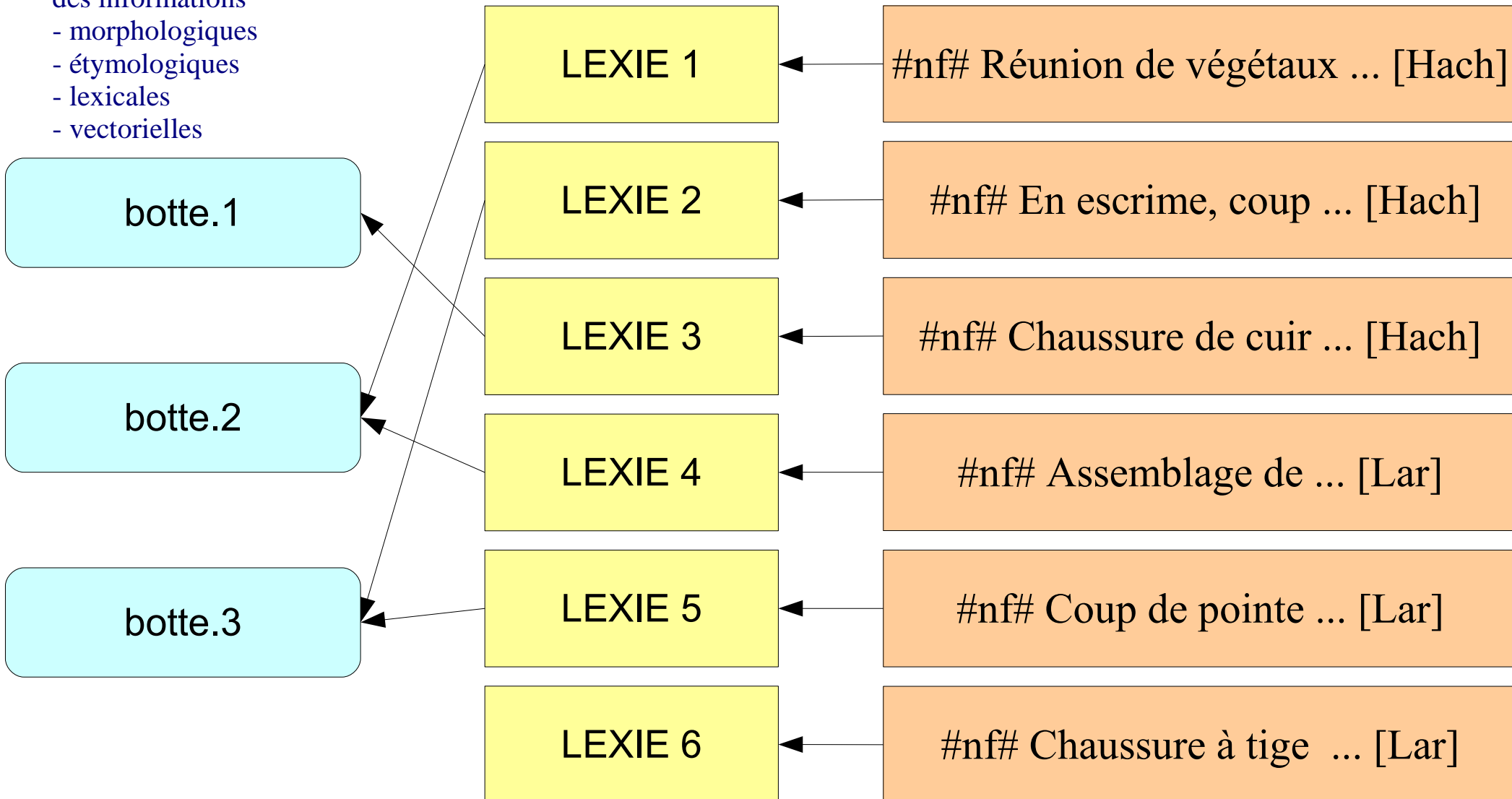


Catégorisation [Jalabert, Lafourcade] [Schwab]

des sens en fonction

des informations

- morphologiques
- étymologiques
- lexicales
- vectorielles



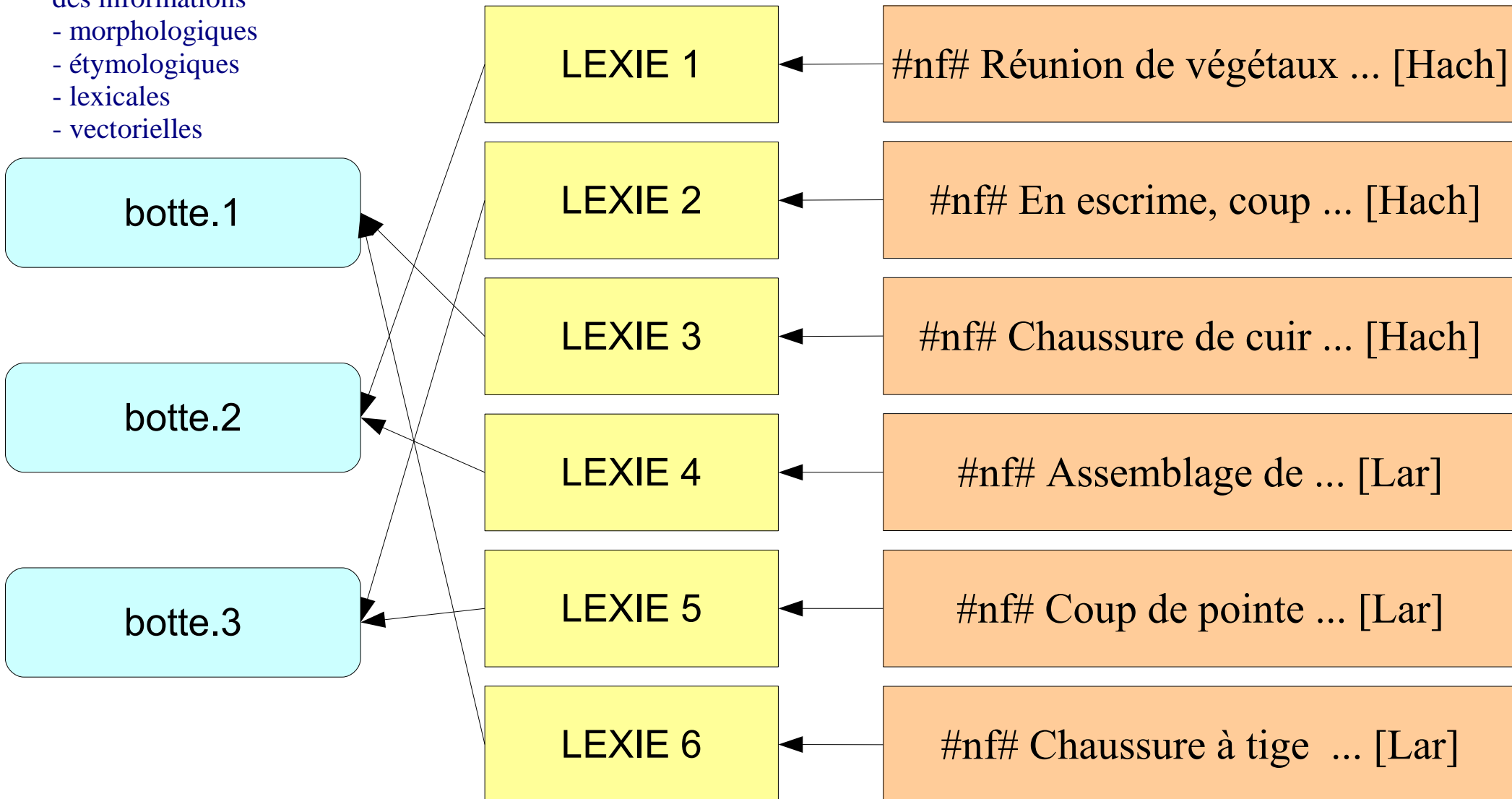


Catégorisation [Jalabert, Lafourcade] [Schwab]

des sens en fonction

des informations

- morphologiques
- étymologiques
- lexicales
- vectorielles

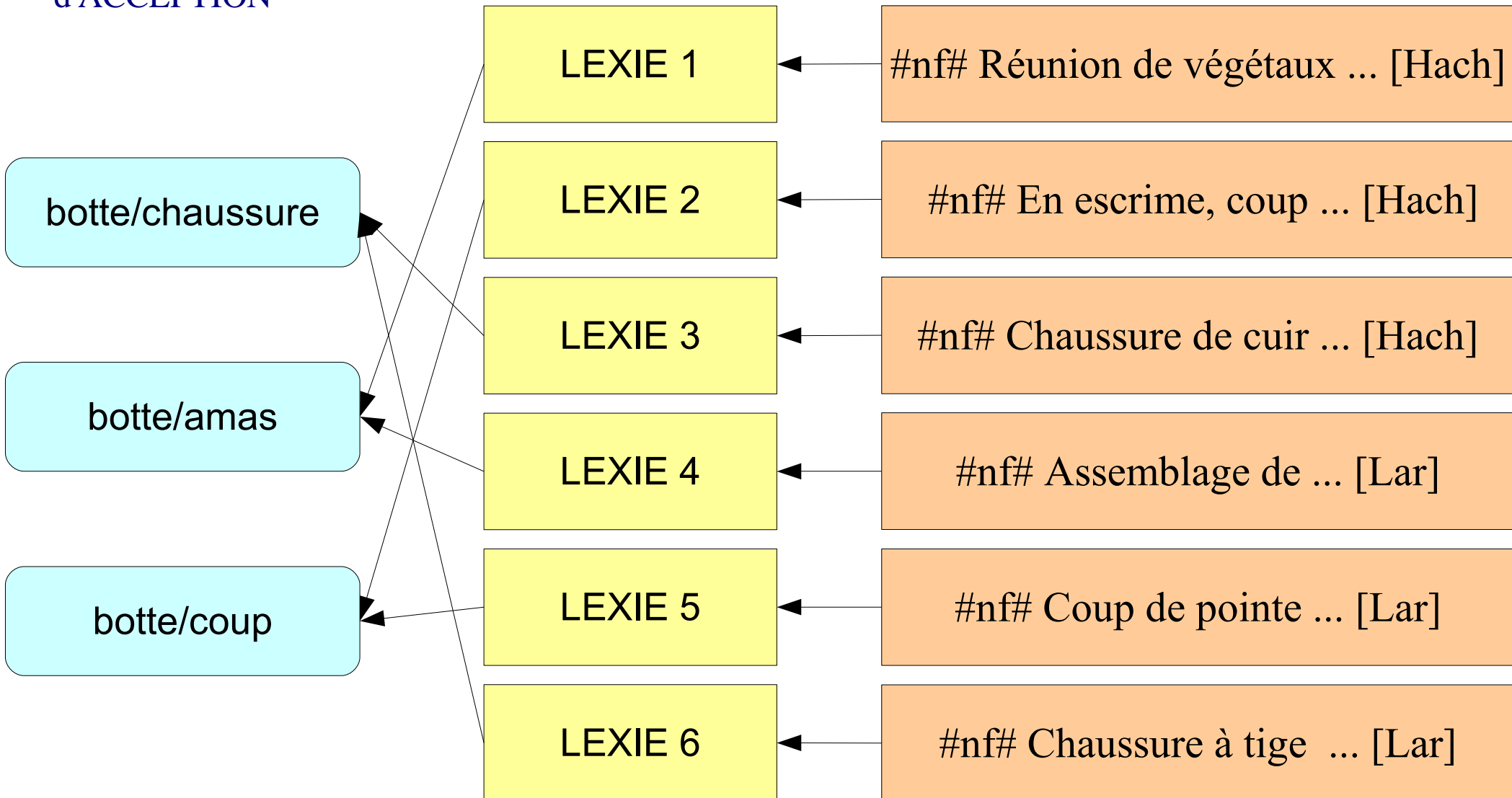




[Jalabert, Lafourcade]

[Schwab]

Nommage
d'ACCEPTION





Hypothèse V

Apprentissage permanent

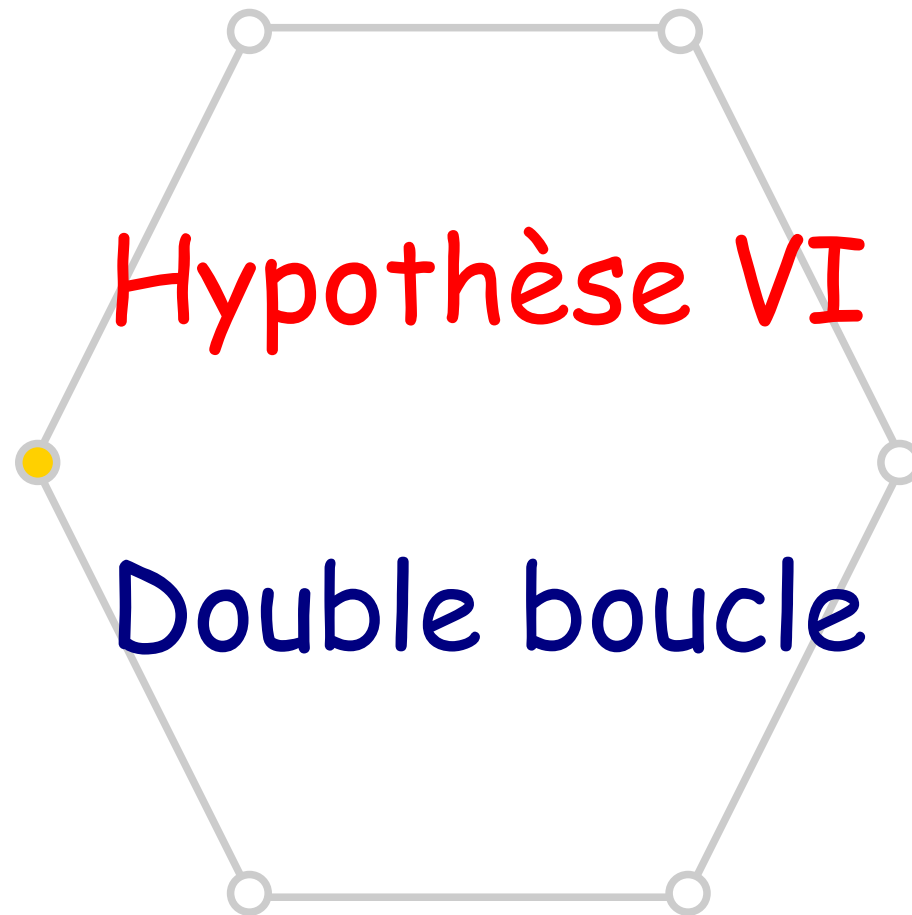


Pour l'analyse de journaux

- Néologismes
 - Entités nommées
 - Personnes (Paris Hilton, ...)
 - Entreprises (Arcelor, ...)
- utilisation du Web et des serveurs d'informations

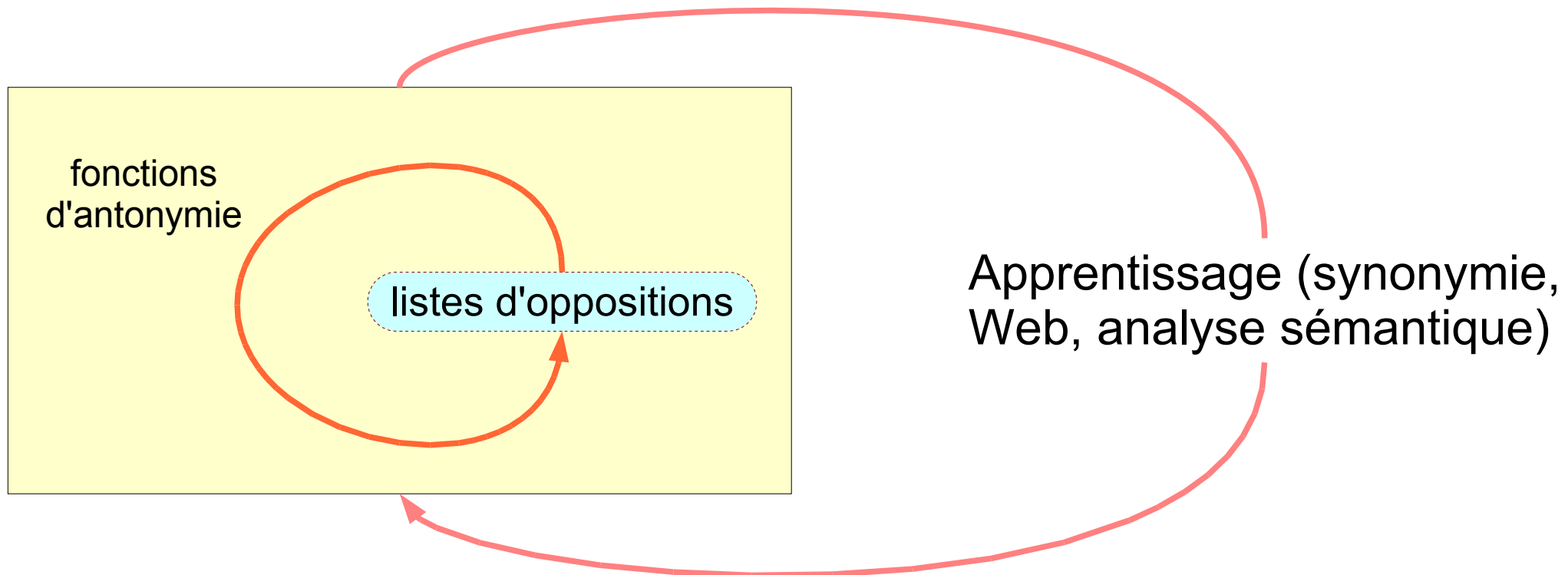
Pour la cohérence de la base

- Base non cohérente lors des premiers cycles
- Convergence des vecteurs vers une position quasi-stable dans un nombre de cycles difficile à pré-calculer (expérimentalement au moins 10)
- Cycle fonction de l'ordre d'apprentissage des items et des définitions.





- Issue de la biologie [Lecerf]
- Élément structurel invariant de l'organisme
- Permet l'action sur son environnement et en est le produit
- Exemple : fonction d'antonymie [COLING'2002, JADT'2002, TALN'2002]





Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



- Objectif : création d'un système permettant **simultanément** l'apprentissage des objets lexicaux et leur exploitation.
- Apprentissage multiple : définitions, relations sémantiques, ...
- Analyse multi-critère : vecteurs conceptuels, fonctions lexicales, fréquences, ...
- Applications multiples : transfert lexical, RI, ...

⇒ Architecture multi-agent ?



Entité physique ou virtuelle [Ferber]

- *qui peut communiquer directement avec les autres agents*
- *capable d'agir sur son environnement*
- *qui possède ses ressources propres*
- *qui est capable de percevoir son environnement*
- *qui possède des compétences*
- *qui offre des services*



On distingue deux types d'agents [Ferber]

- *Agents réactifs (par exemple, les fourmis)*
 - Agents sans base de connaissances
 - Communication par diffusion de signaux
- *Agents cognitifs*
 - Chaque agent possède sa base de connaissances
 - Communications par envoi de messages

à l'échelle **macroscopique**, agents cognitifs

à l'échelle **microscopique**, agents réactifs pour analyse sémantique à l'aide de fourmis (agents récursifs)

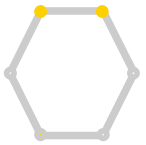


Pourquoi adopter un SMA distribué ?

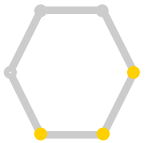


Raisons dues aux hypothèses

Hypothèses I et II sur la représentation du sens
Indépendantes de l'architecture



Hypothèses III (*génération automatique*), IV (*multi-source*)
et V (*apprentissage permanent*)
Utilisation d'agents scrutant dictionnaires, listes d'antonymes,
de synonymes, Web, ...



Hypothèse VI (*double boucle*)

- Facilite la conception et l'implémentation
- Les agents modifient leur base de connaissances en fonction des informations lexicales rencontrées + effet retour





Raisons dues aux applications visées

Applications variées et hétérogènes (RI, TA, RA, ...)

Possibilité de rajouter des modules

Raisons techniques

Problèmes de ressources

lexique d'au moins 1 000 000 d'entrées

stockage, mémoire vive

⇒ distribution sur plusieurs machines

Génie Logiciel

facilité pour connecter plusieurs BLS (ex : français anglais)

développement (agents longs à charger)



Blexisma (Base LEXIcale Sémantique Multi-Agent)

implémente un noyau (agents, communications)

développée en Java (Jdk 1.3 à 1.5)

Agents implémentés

14 types différents

-agents FL (modélisation, extraction)

-analyse sémantique (remontée-descente, à fourmis)

-base

-catégoriseur

-contextualiseur

-apprentissage définitions

...

Accessible via le Web (<http://www.lirmm.fr/~schwab>)



Exemple : apprentissage d'une LEXIE

LIRMM > Blexisma

Extracteur
de
définitions

apprentissage

base

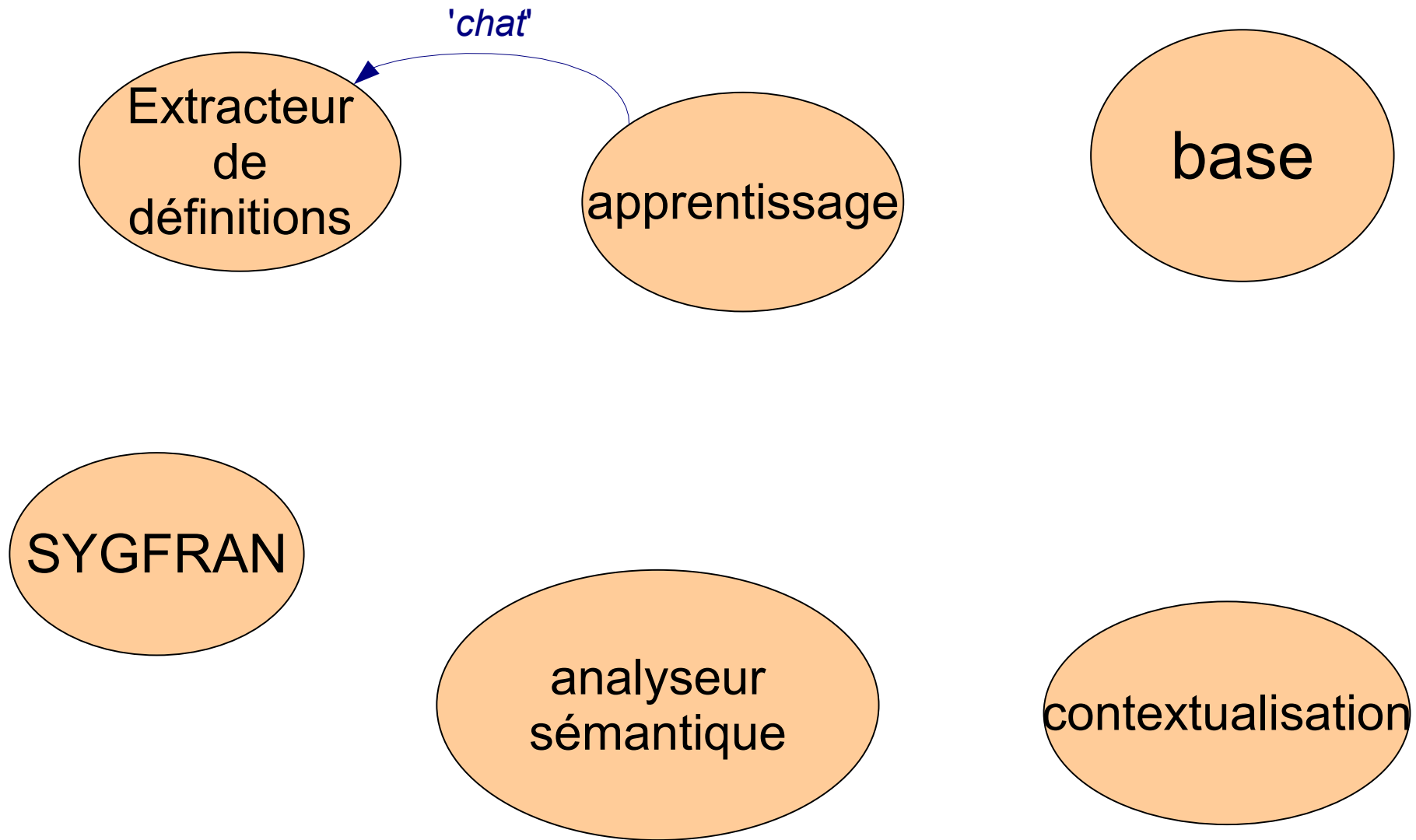
SYGFRAN

analyseur
sémantique

contextualisation

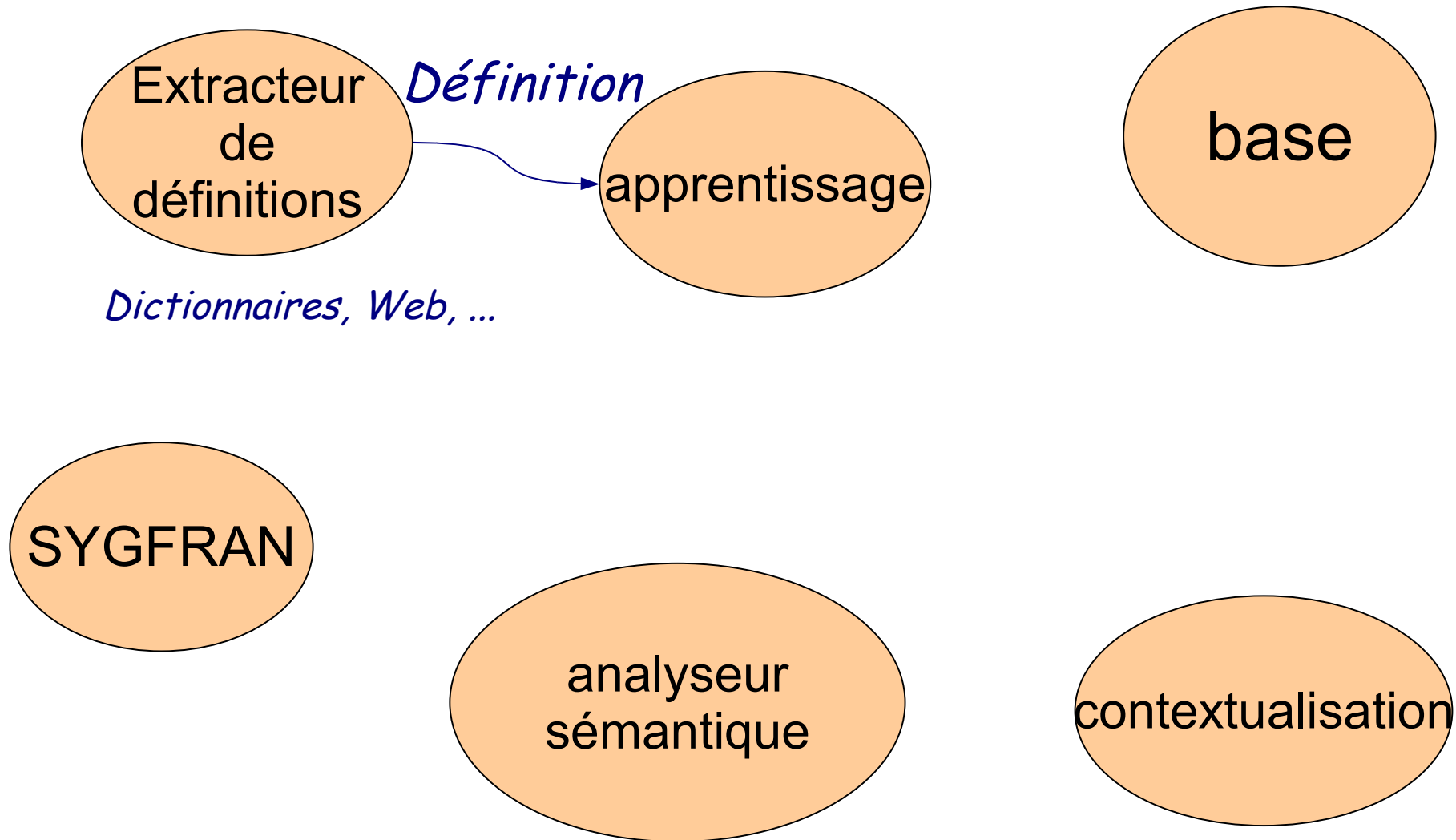


Exemple : apprentissage d'une LEXIE



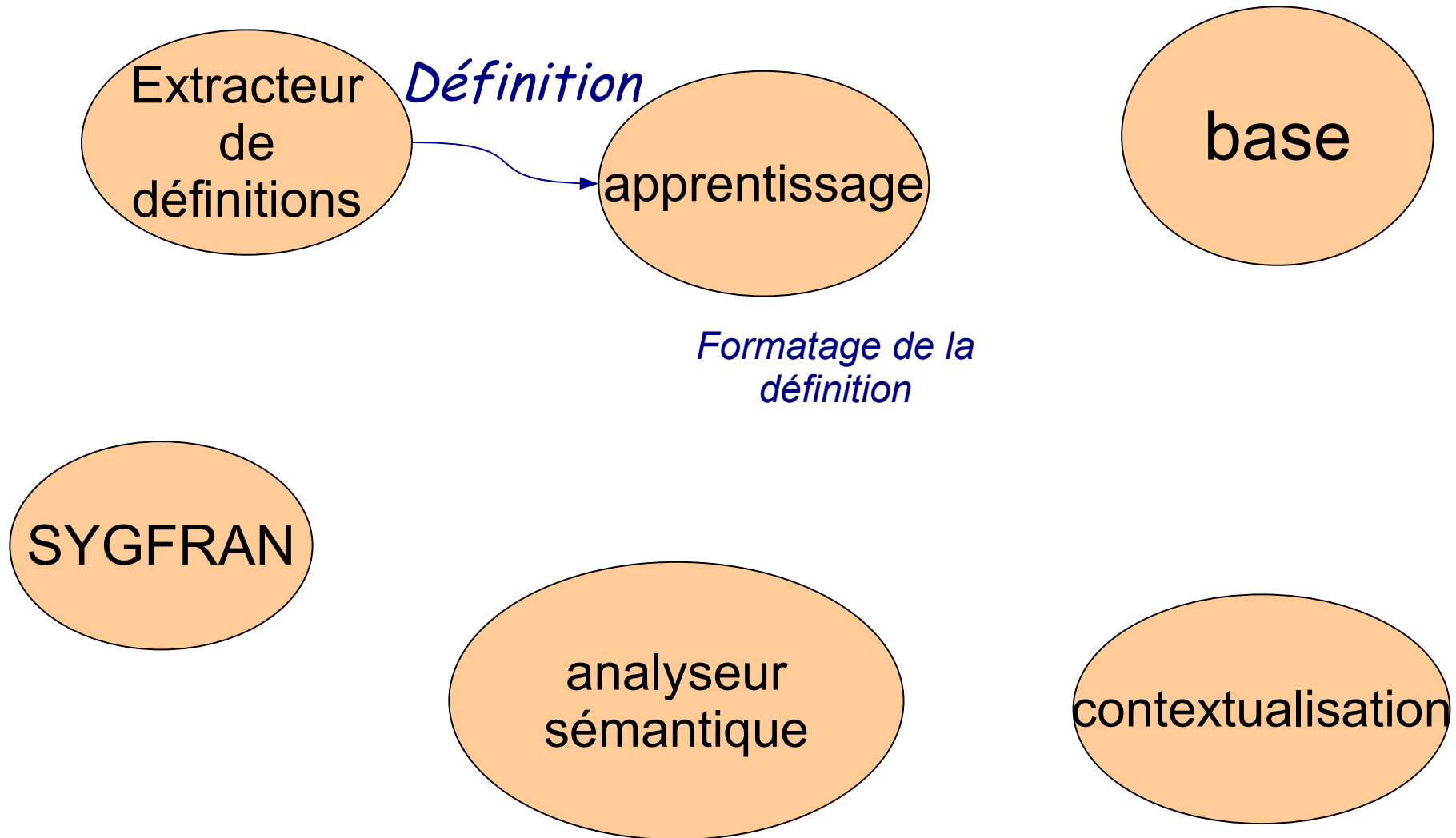


Exemple : apprentissage d'une LEXIE



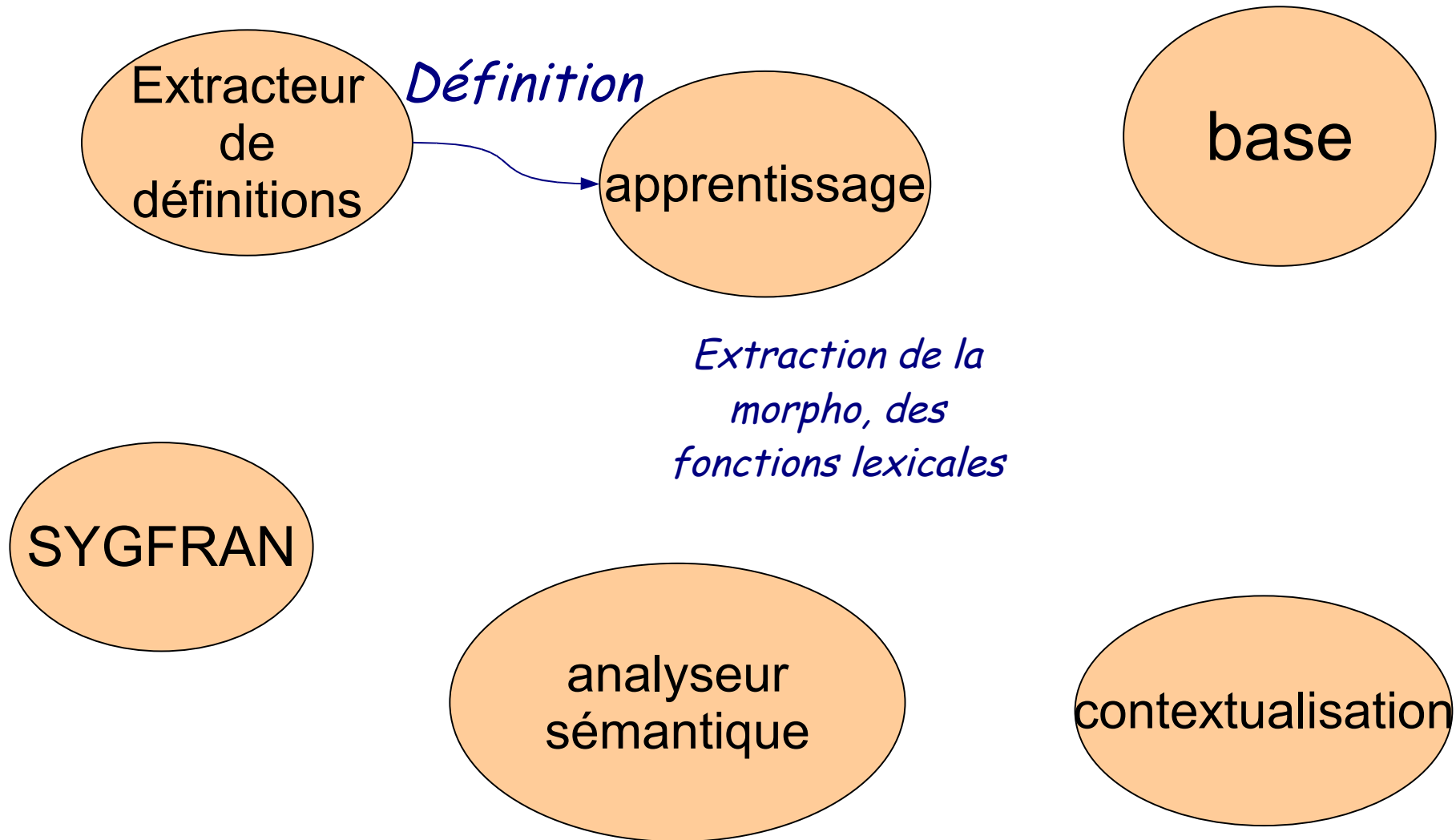


Exemple : apprentissage d'une LEXIE



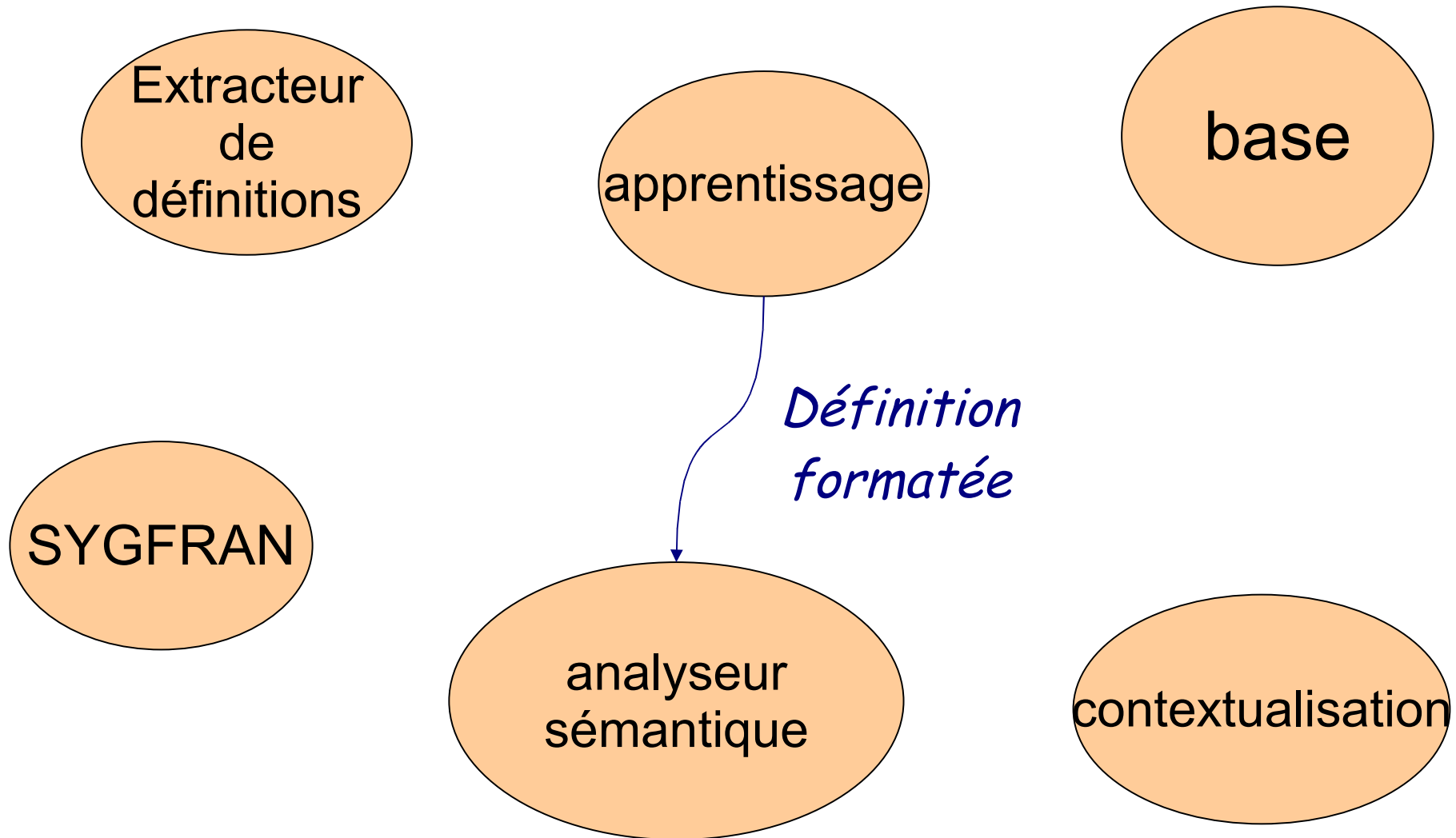


Exemple : apprentissage d'une LEXIE



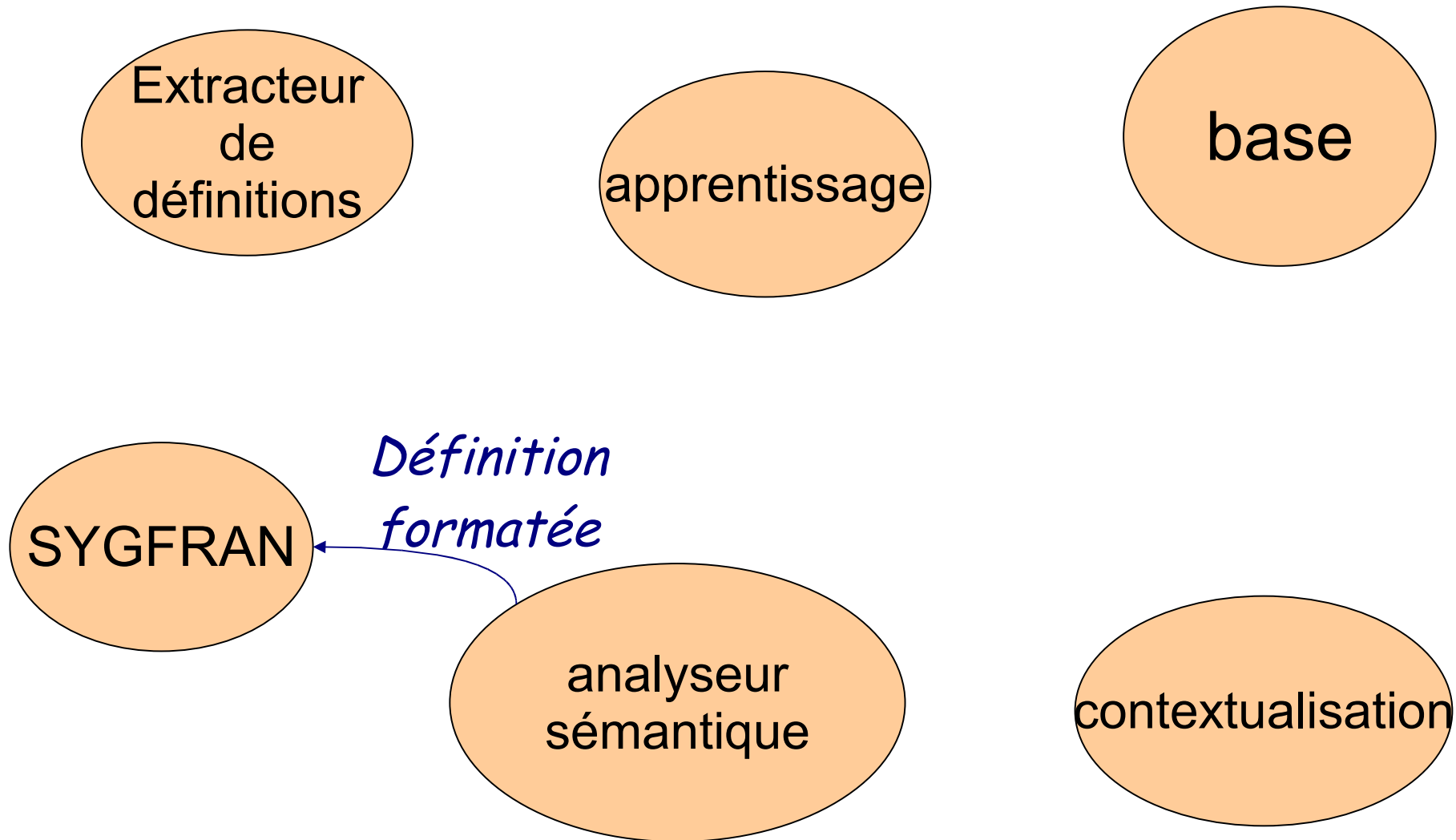


Exemple : apprentissage d'une LEXIE



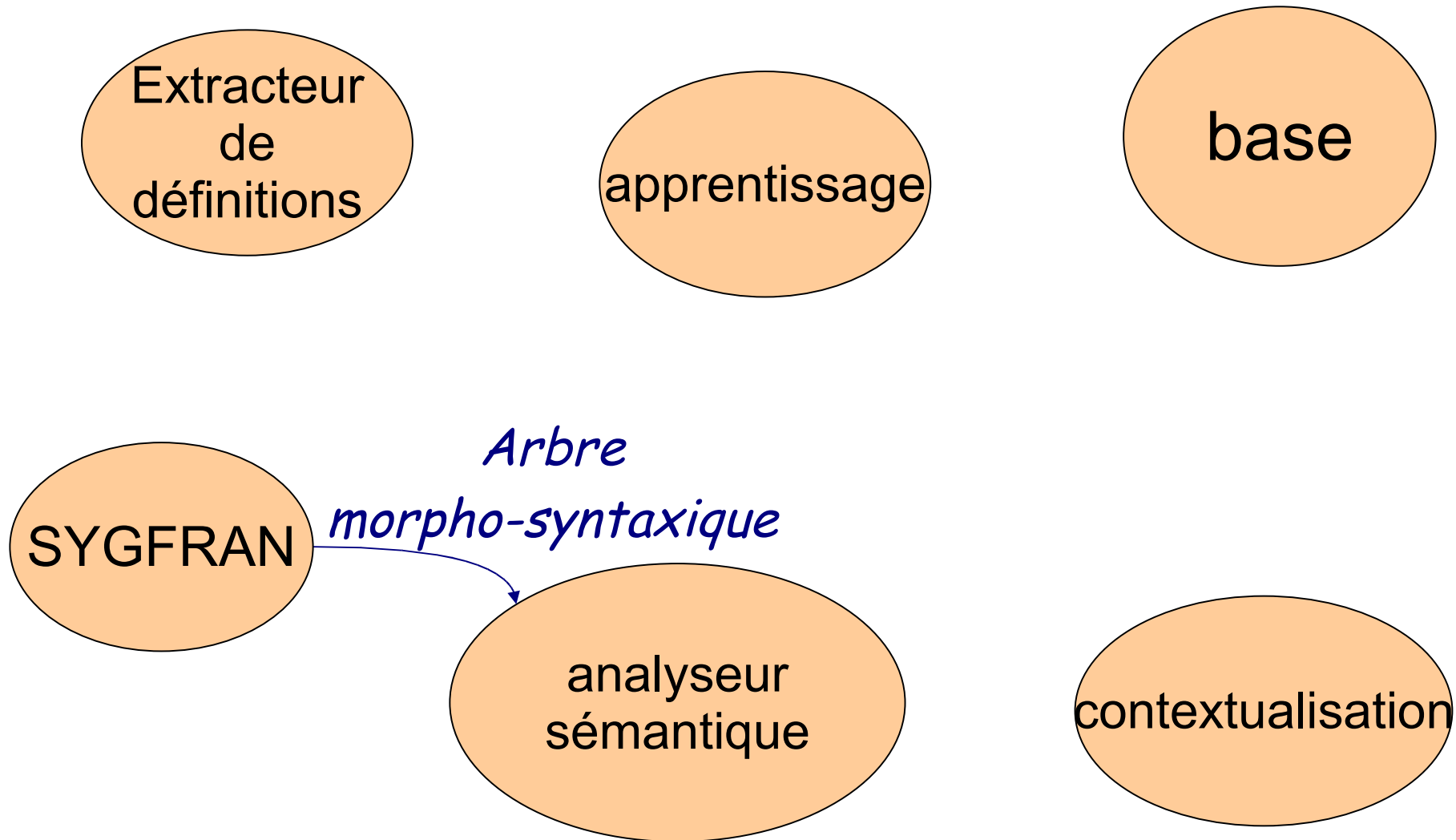


Exemple : apprentissage d'une LEXIE



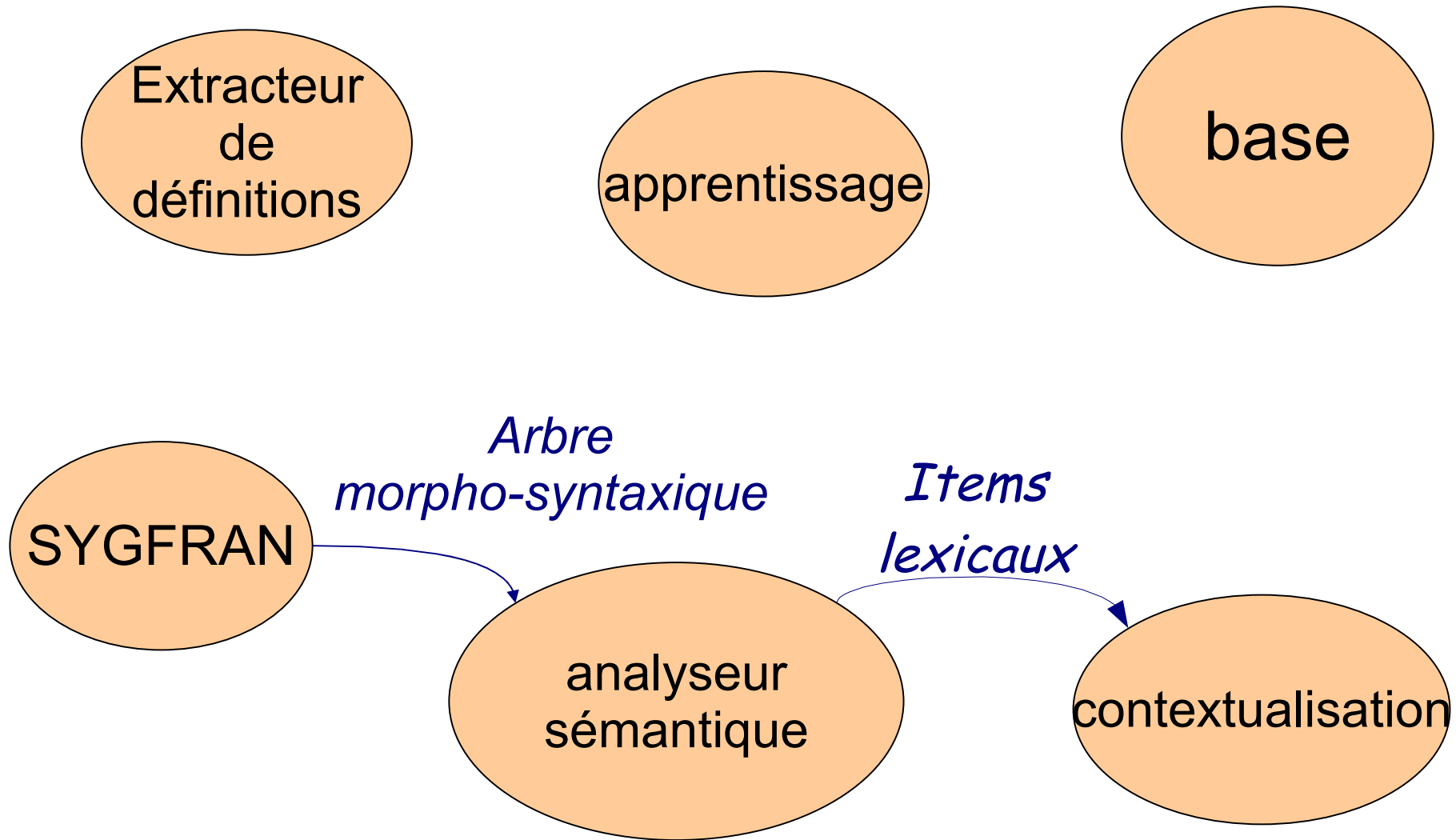


Exemple : apprentissage d'une LEXIE



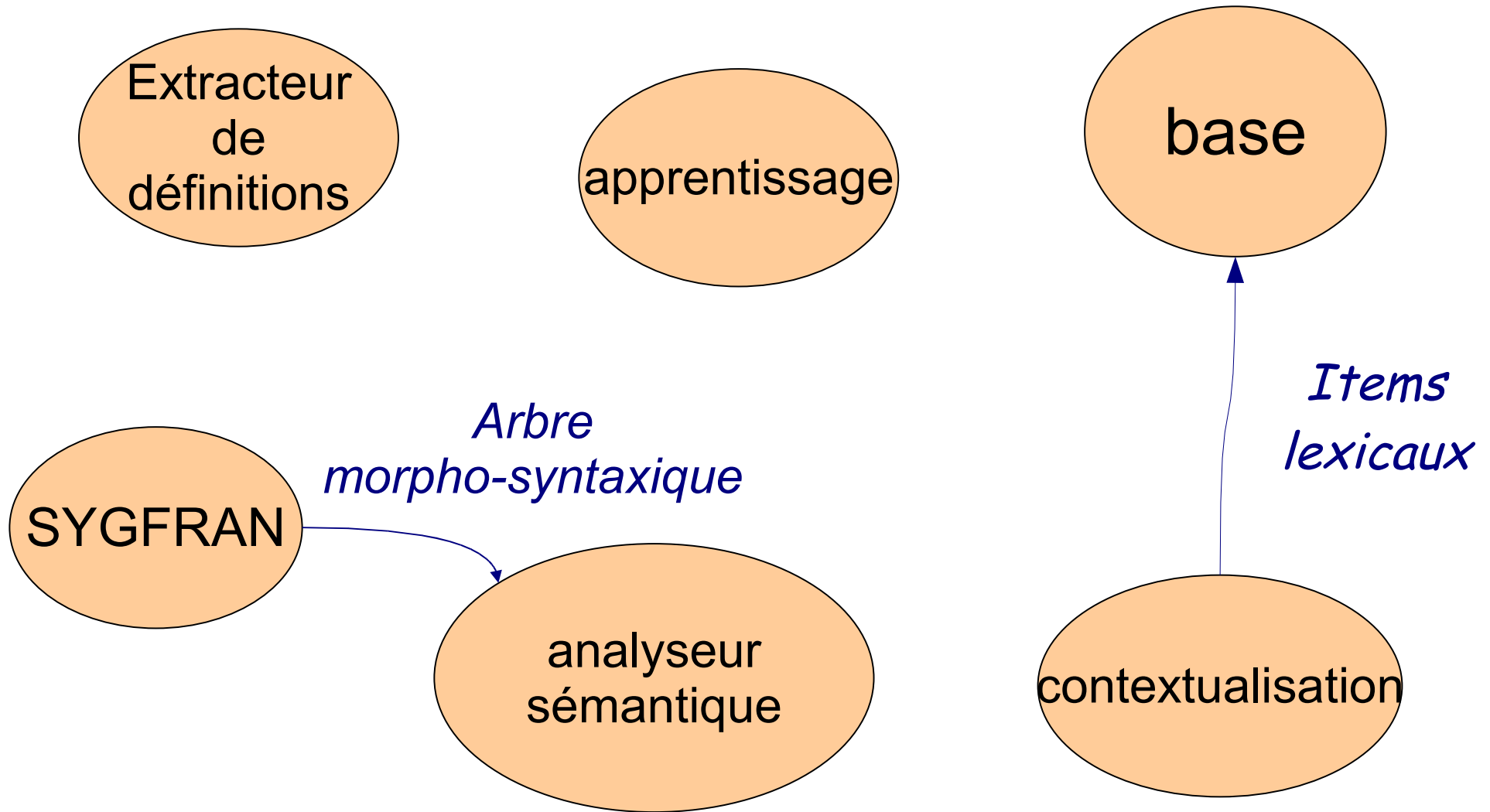


Exemple : apprentissage d'une LEXIE



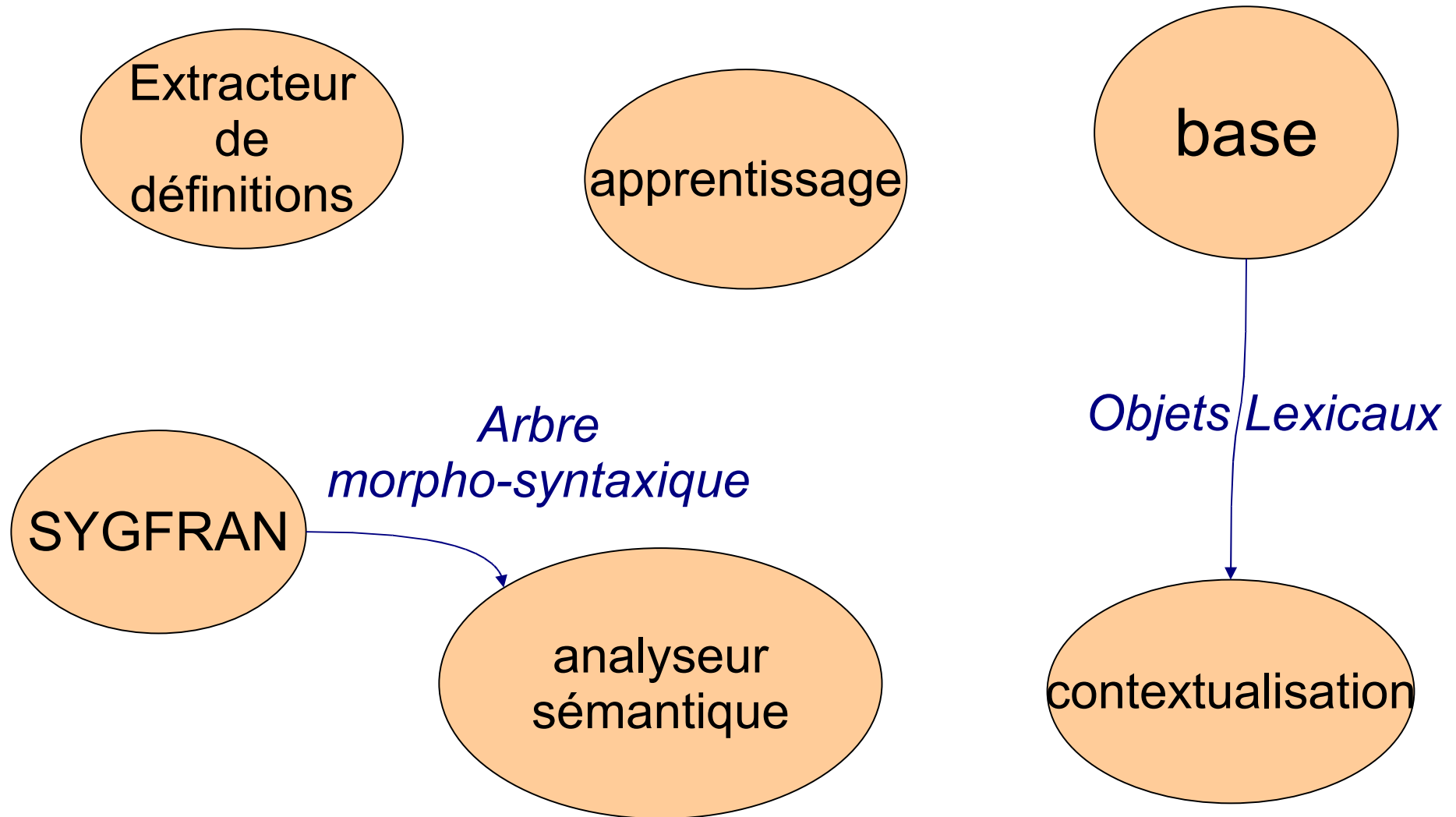


Exemple : apprentissage d'une LEXIE



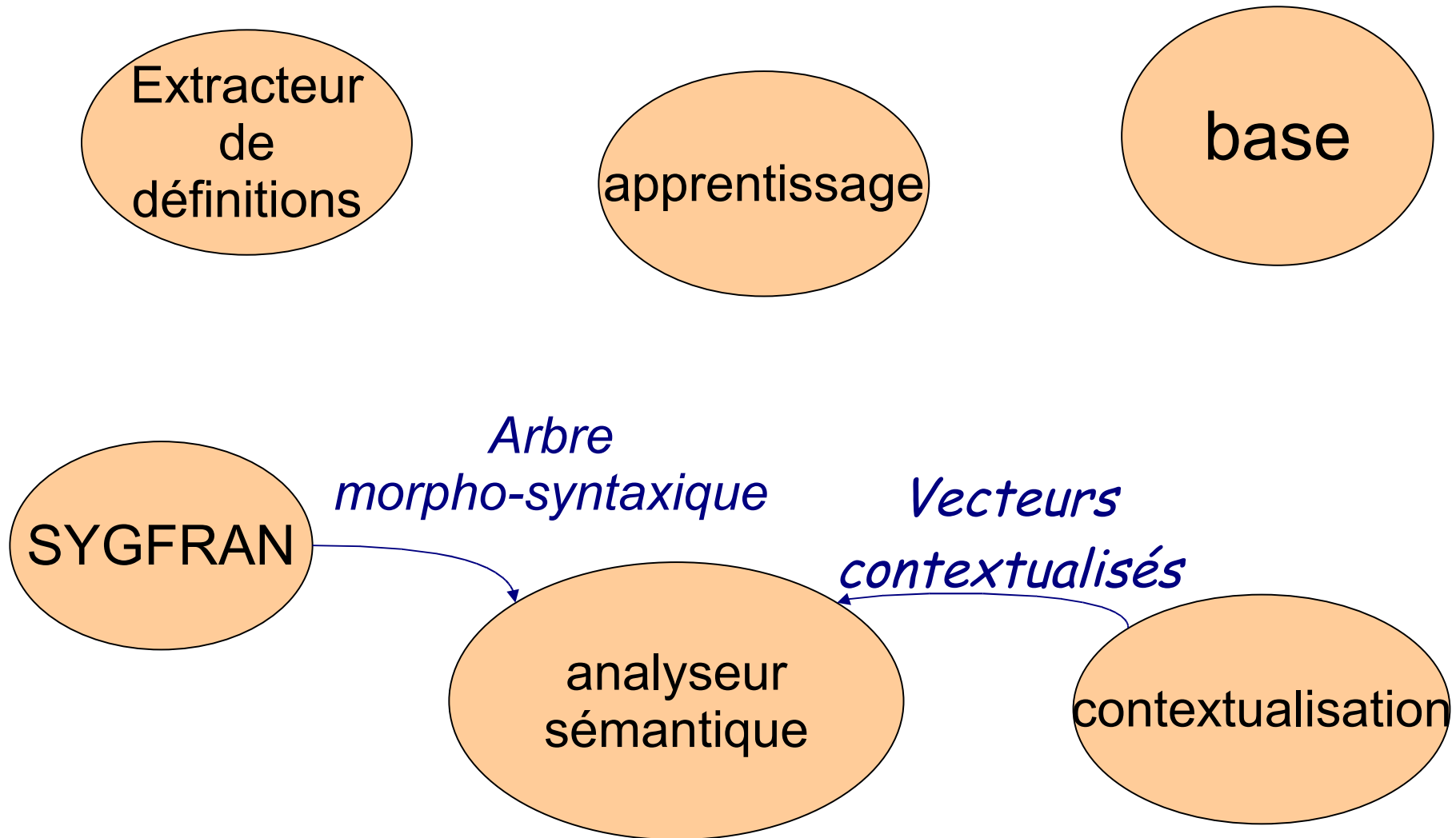


Exemple : apprentissage d'une LEXIE



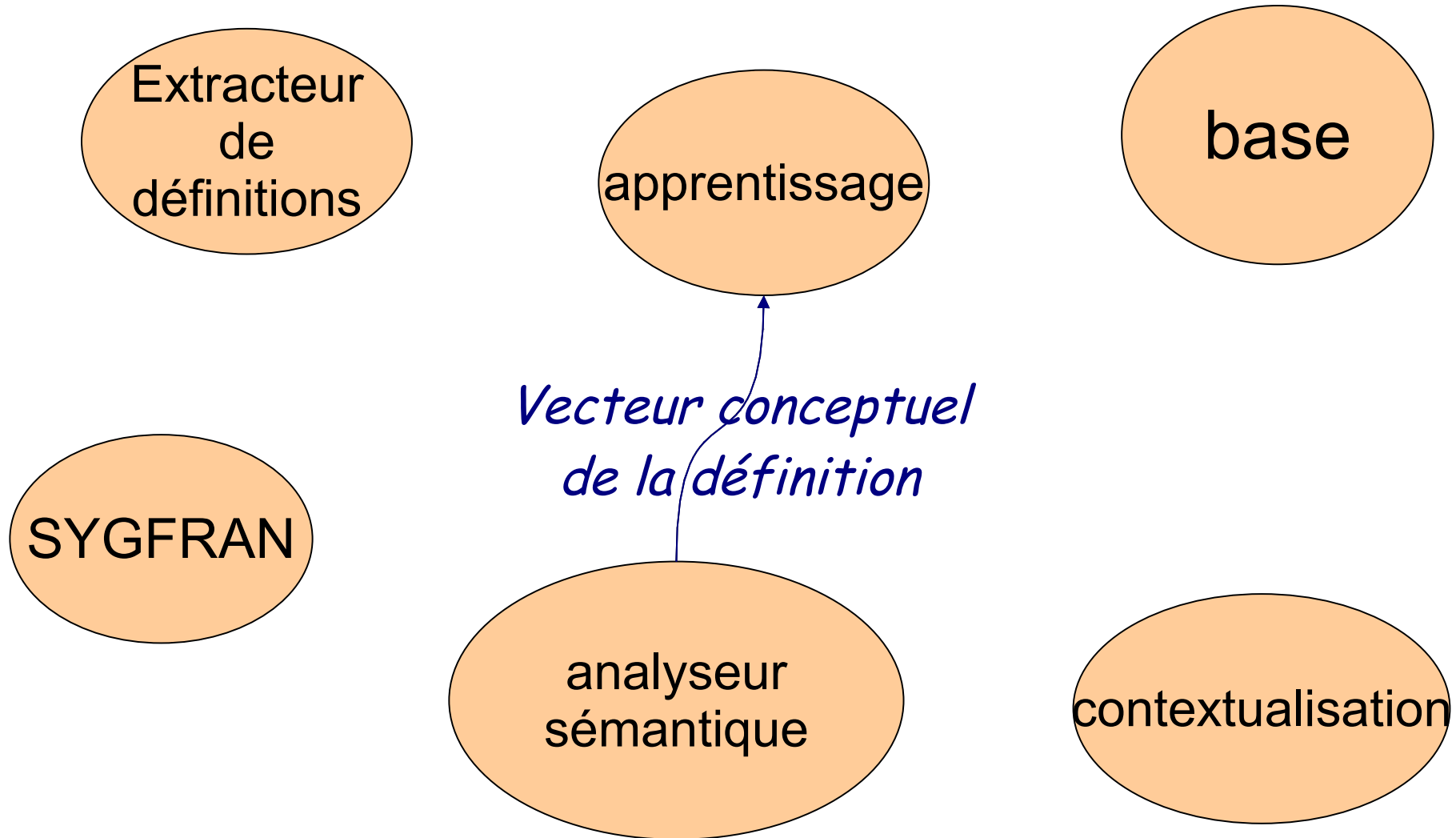


Exemple : apprentissage d'une LEXIE



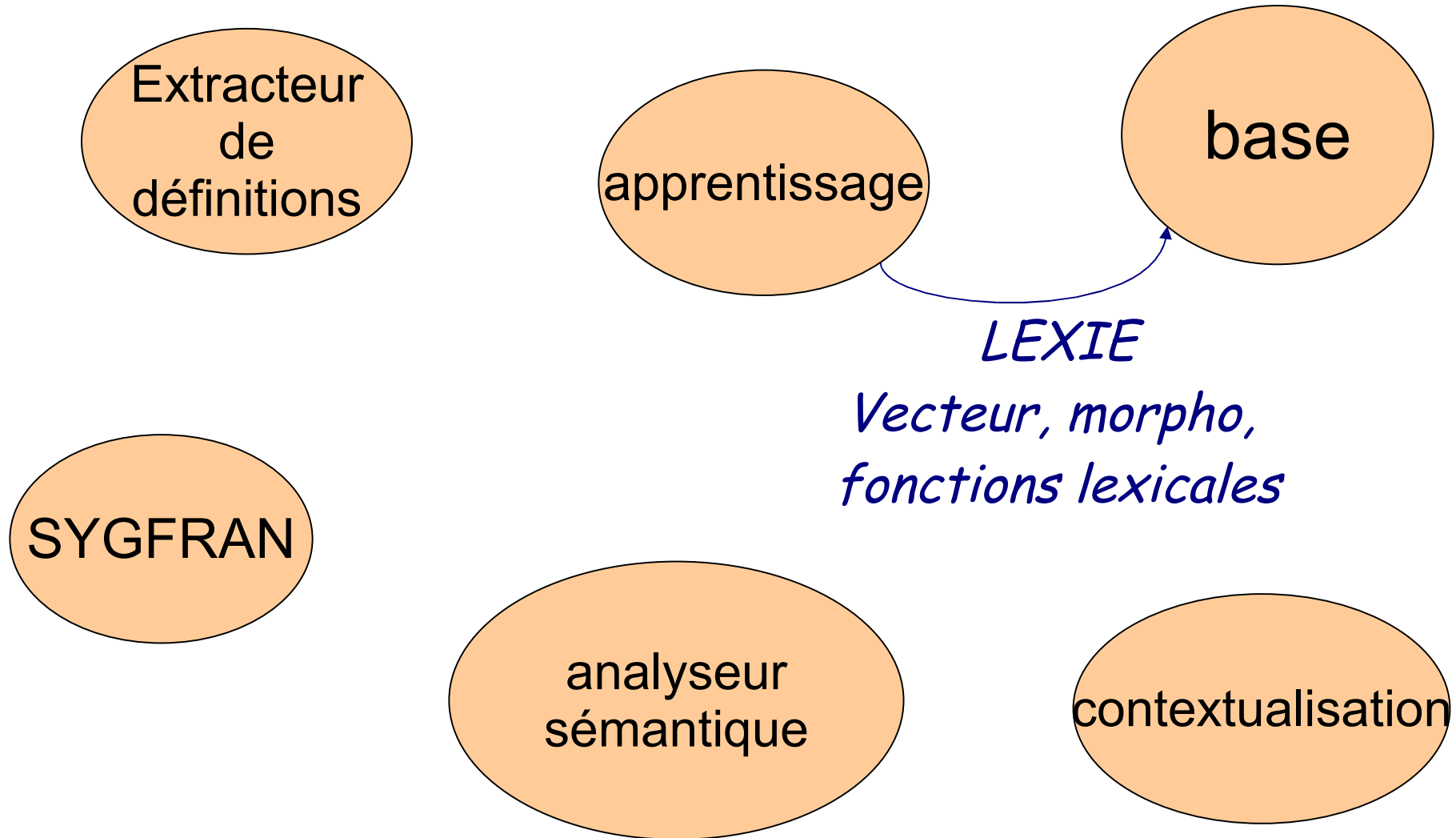


Exemple : apprentissage d'une LEXIE





Exemple : apprentissage d'une LEXIE





115 agents (1 base, jusqu'à 10 de chaque autre type)

5 machines (PC Linux, Sun Unix)

5 sources (Larousse, Robert, thésaurus Larousse,
dictionnaire synonymes, antonymes de Caen)

Base du Français

121 000 ITEMS LEXICAUX

276 000 ACCEPTIONS

842 000 LEXIES

Cycle (environ 4 jours)



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



2 types de connaissances

- connaissances lexicales
- connaissances du monde

Nécessité de connaître les deux pour l'analyse d'un texte

« Jean a eu une *peur bleue*. » (lexical)

Magn

« Il a *creusé* un trou avec une *pelle*. » (monde)

Instr

« *Zidane* est un *footballeur*. » (monde)

Ins

« Un *chat* est un *animal*. » (monde et lexical)

Hyper

Gener



FLA pour les connaissances linguistiques (inspirée des FL de Mel'čuk)

- synonymie, antonymie, générique
- FLA adjectivales (intensification, laudatif, ...)
- nom du centre, nom de l'équipe

...

FLA pour les connaissances du monde (inspirées de UNL)

- hyperonymie de classe
- instance
- méronymie (partie de)
- instrument

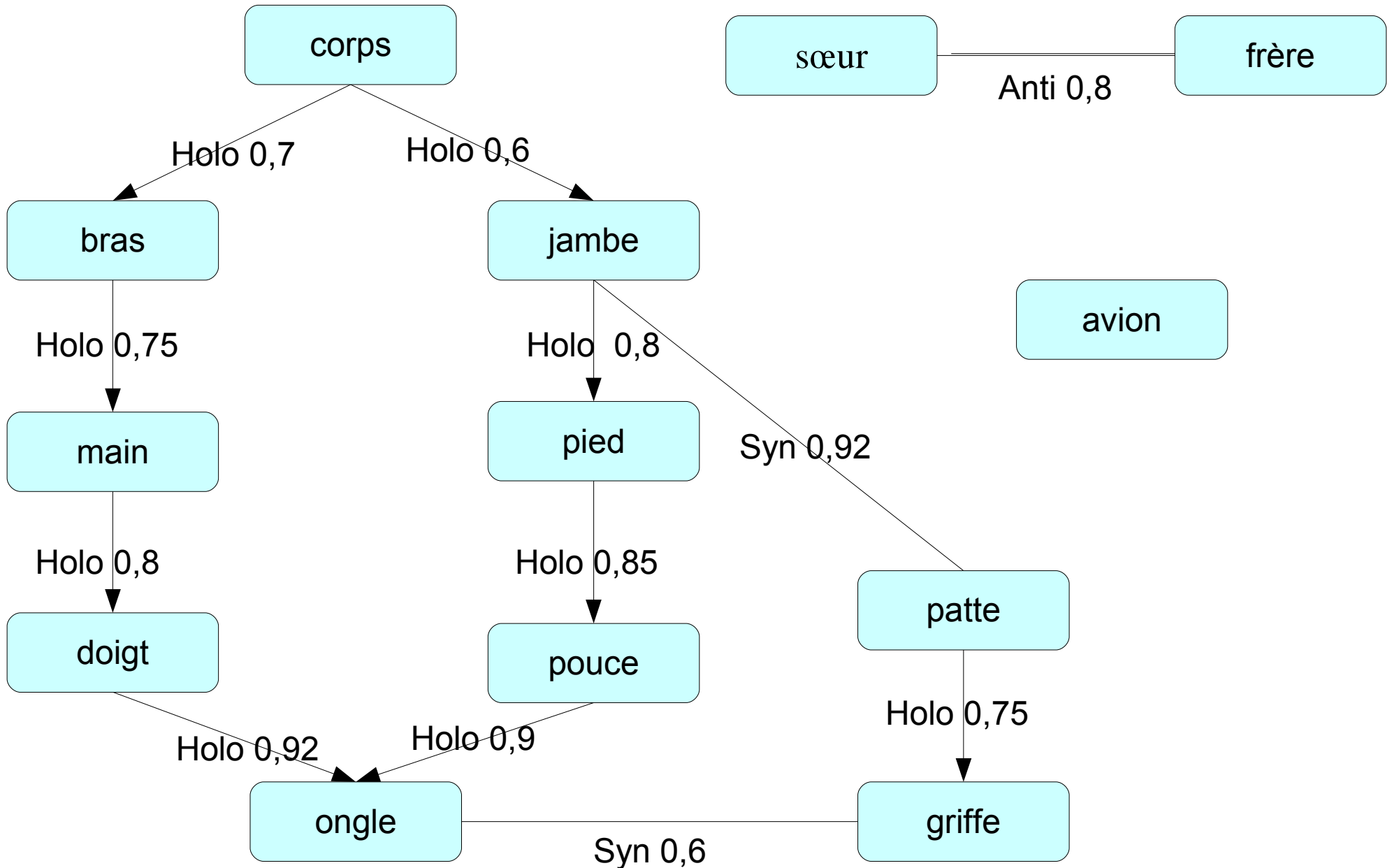
...

La Base Lexicale Sémantique doit contenir ces fonctions

Sous quelle forme ?



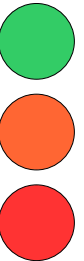
Relations Sémantiques Valuées.





Hypothèse IV (analyse multi-source)

qualité inégale : dictionnaires synonymes, antonymes
dictionnaires classiques
Web, corpus de textes

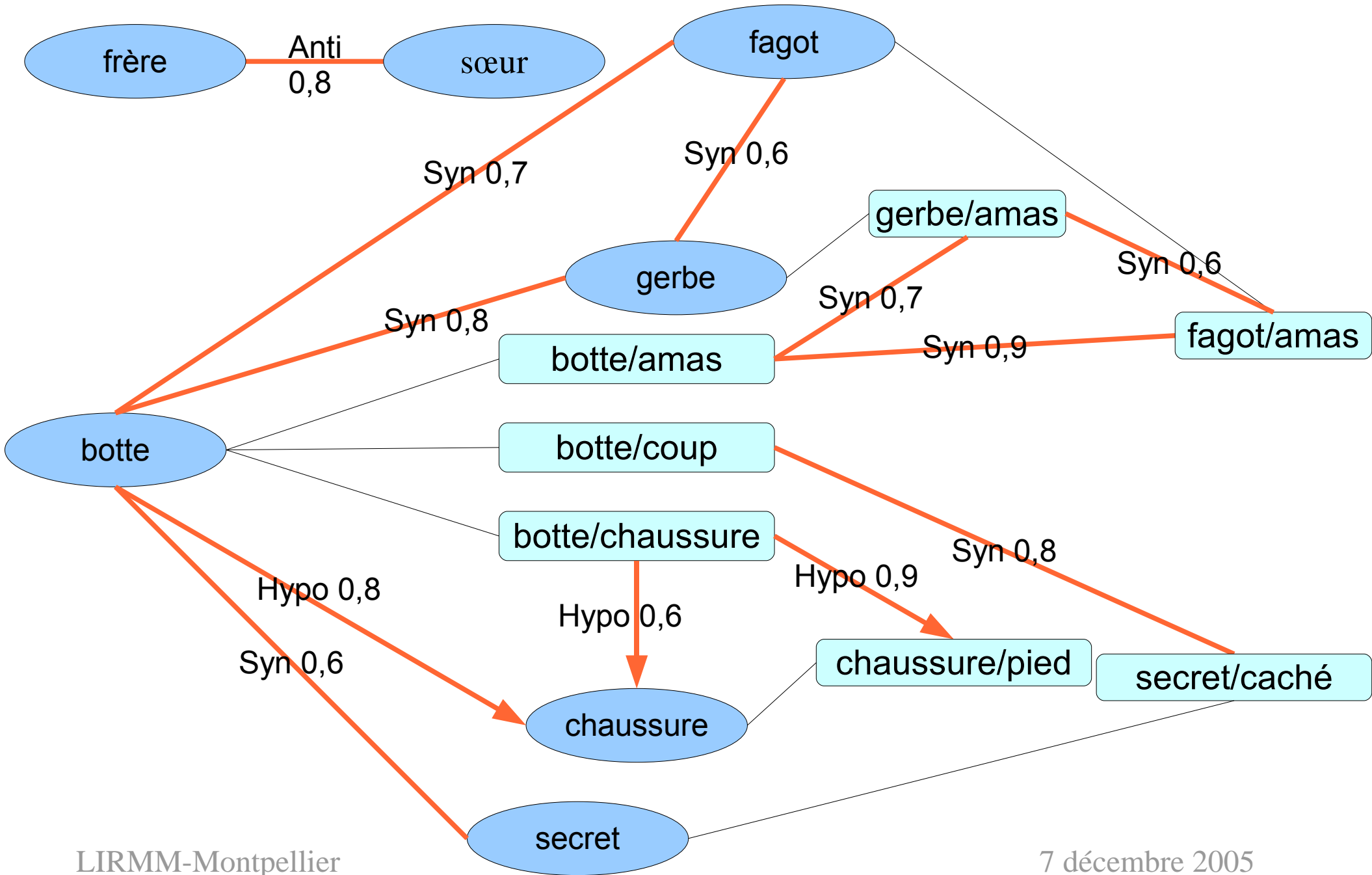


⇒ Pondération des relations repérées/identifiées
entre items lexicaux
entre acceptions

Hypothèse V (apprentissage permanent) ⇒ pondération entre
objets lexicaux différents



Relations Sémantiques Valuées..





2 types de Fonctions Lexicales d'Analyse

- *FLA de construction*
permettent de construire des vecteurs conceptuels
(apprentissage)
- *FLA d'évaluation*
mesurent la pertinence d'une relation entre deux objets
lexicaux (évaluation de la pertinence de la base)



	lexicale	lexicales et thématiques
connaissances du monde	méronymie, instrument, ...	hyponymie, instanciation (AS remonté descente)
connaissances linguistiques	collocations	Syn, Anto (paradigmatiques de Mel'čuk)



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Hypothèses
pour la construction d'une Base Lexicale Sémantique

Société d'agents apprenants - Blexisma

Fonctions Lexicales d'Analyse

Exploitation dans l'Analyse Sémantique
algorithmes à fourmis



Système Multi-Agent réactifs

Intelligence en essaim [Bonabeau, Théraulaz]

Auto-organisé, interactions simples

Meta-heuristique pour les problèmes combinatoires

TSP - recherche opérationnelle - réseau - bioinfo [Dorigo, Stützle]

Principe

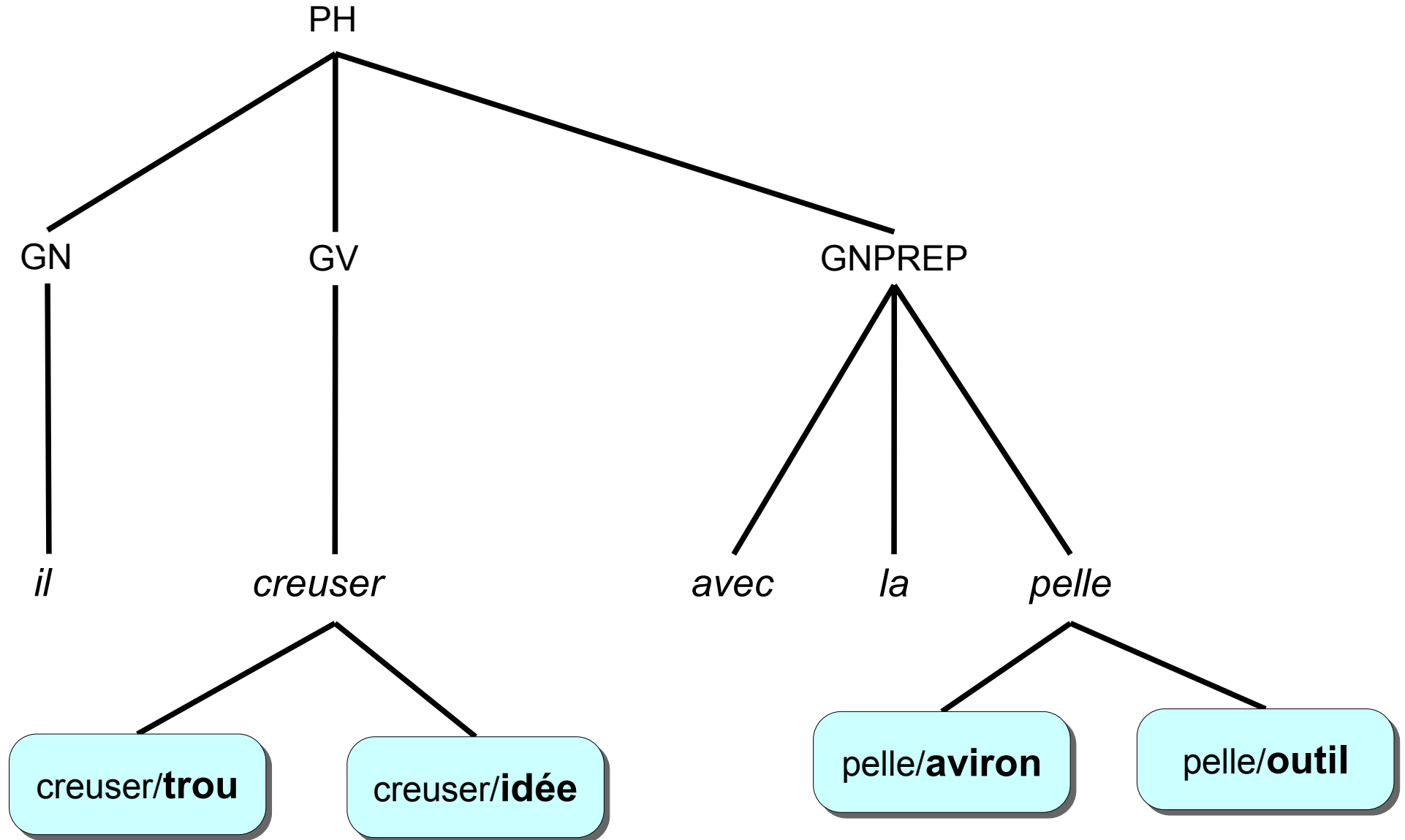
Stygmergie = communication indirecte
par modification de l'environnement

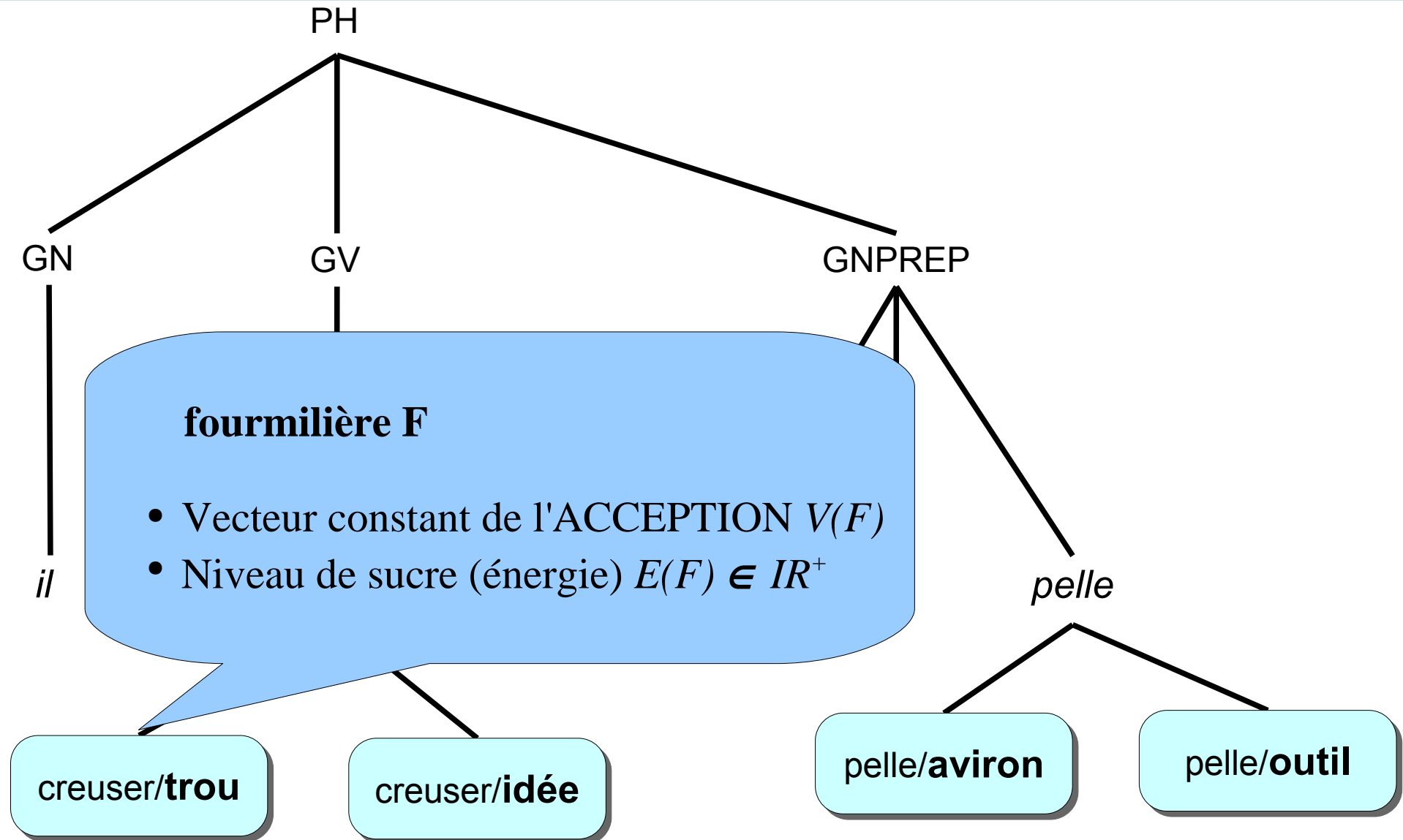
Dépot de phéromones - évaporation lente

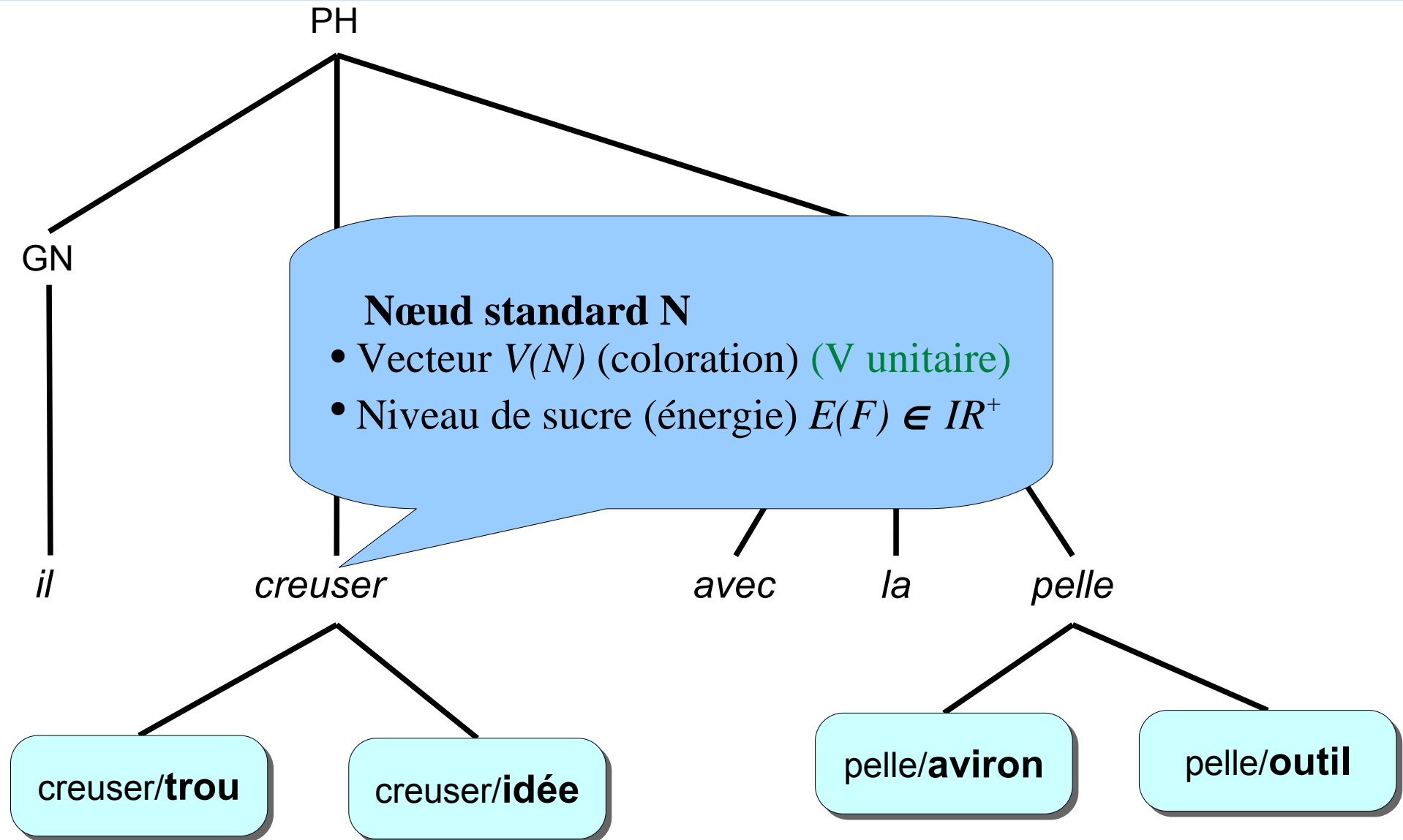
Premiers modèles pour l'analyse sémantique de texte

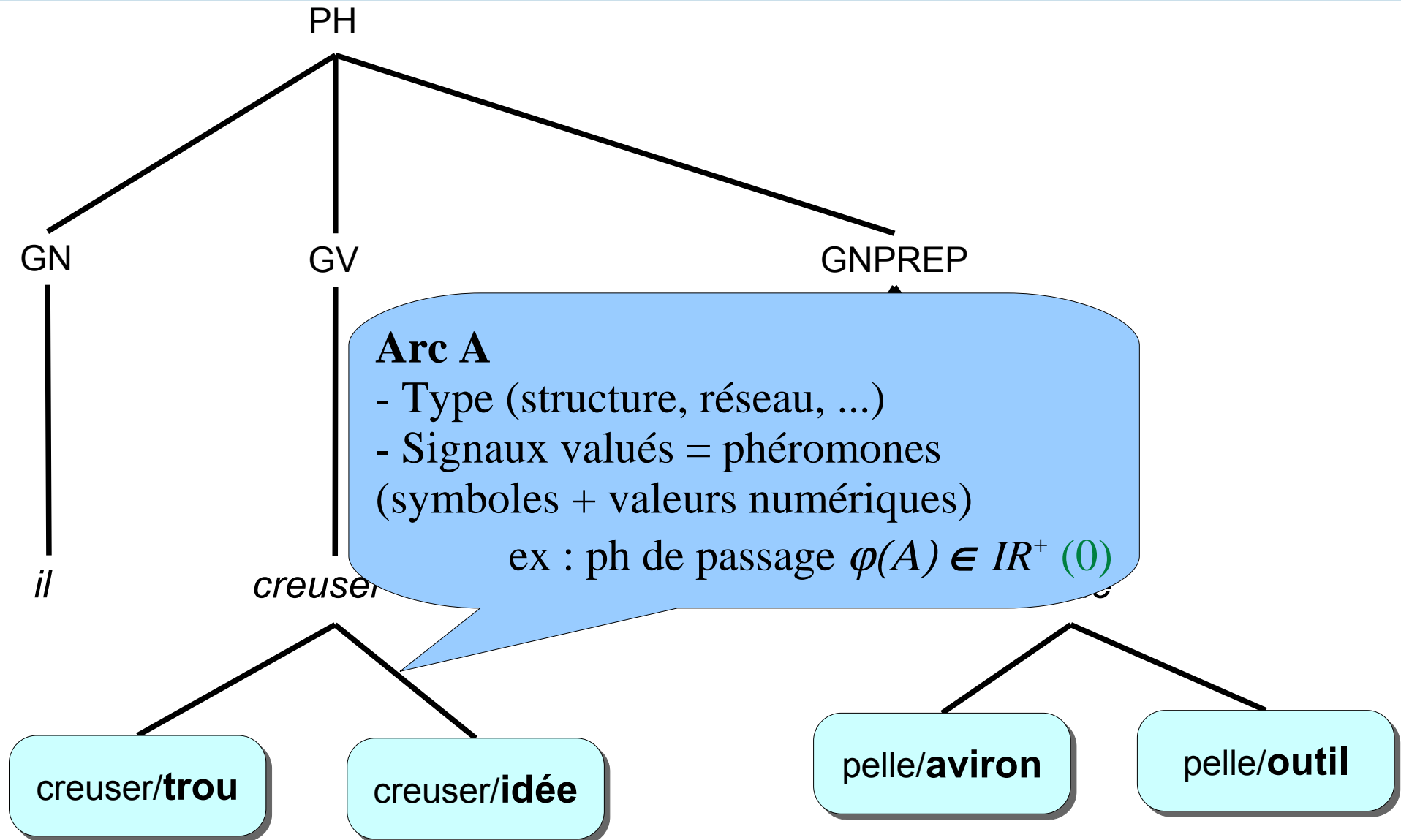
Mono-caste mono-environnement [Lafourcade, Guinand]

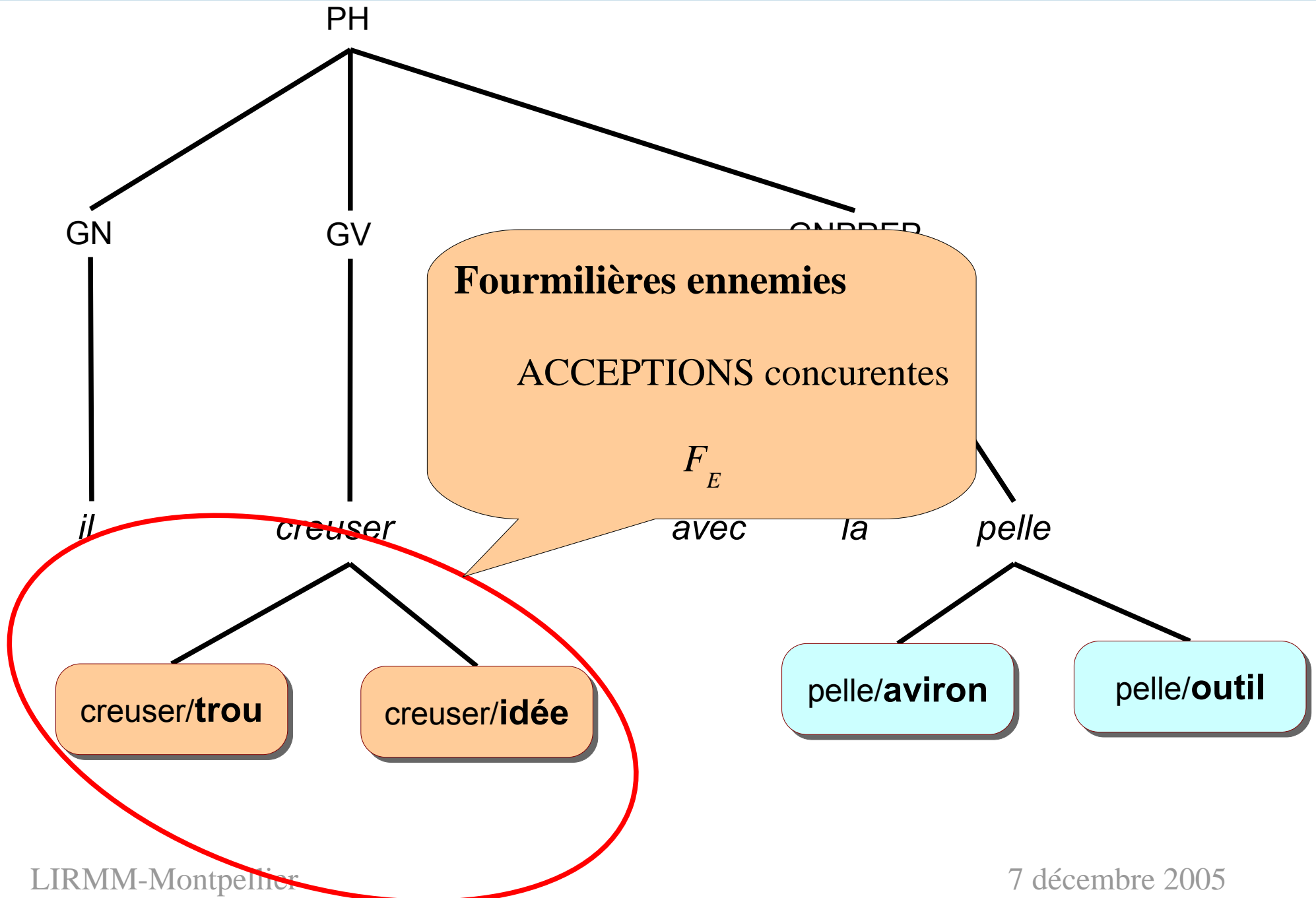
Multi-caste environnement séparé [Zamora, Lafourcade]

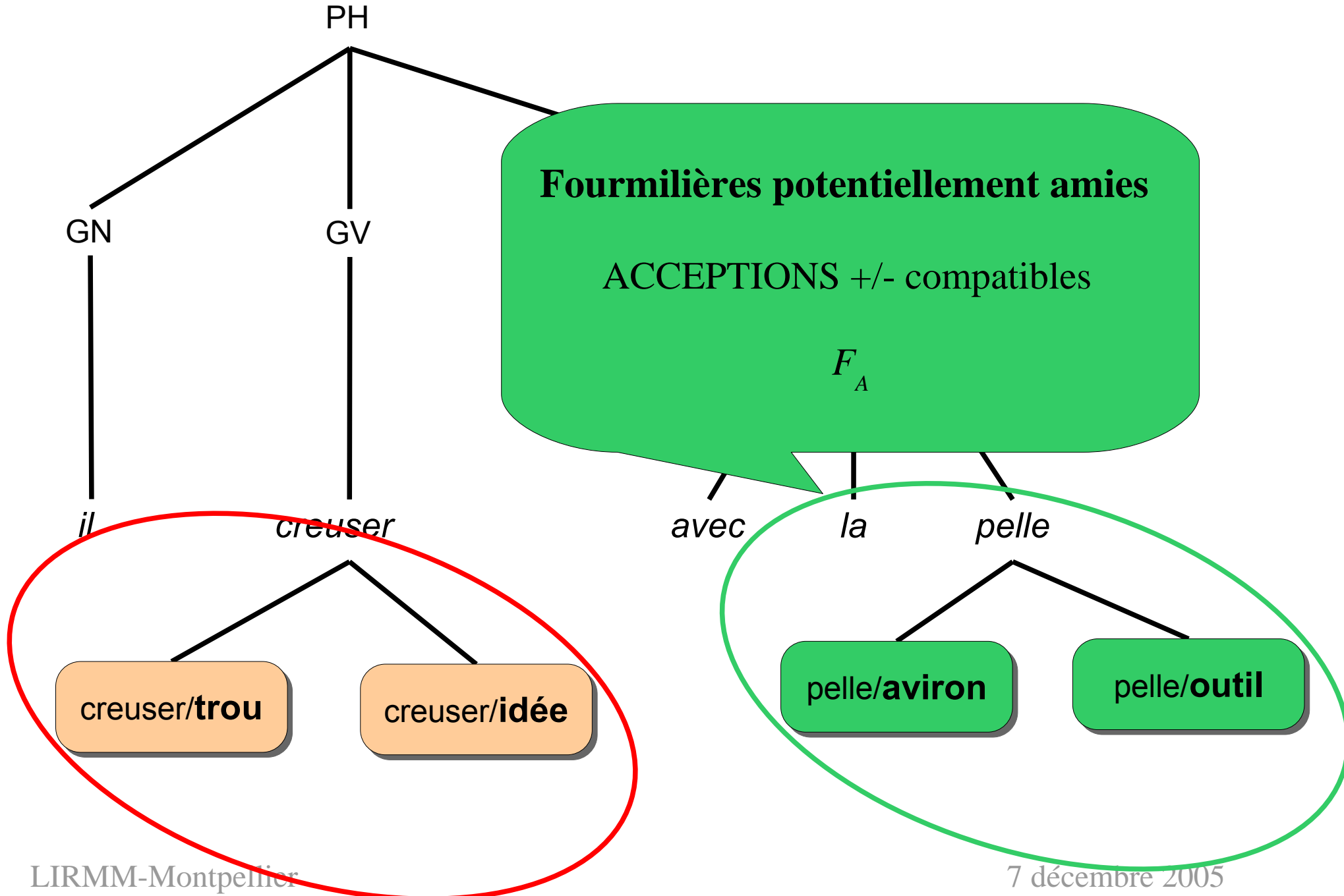


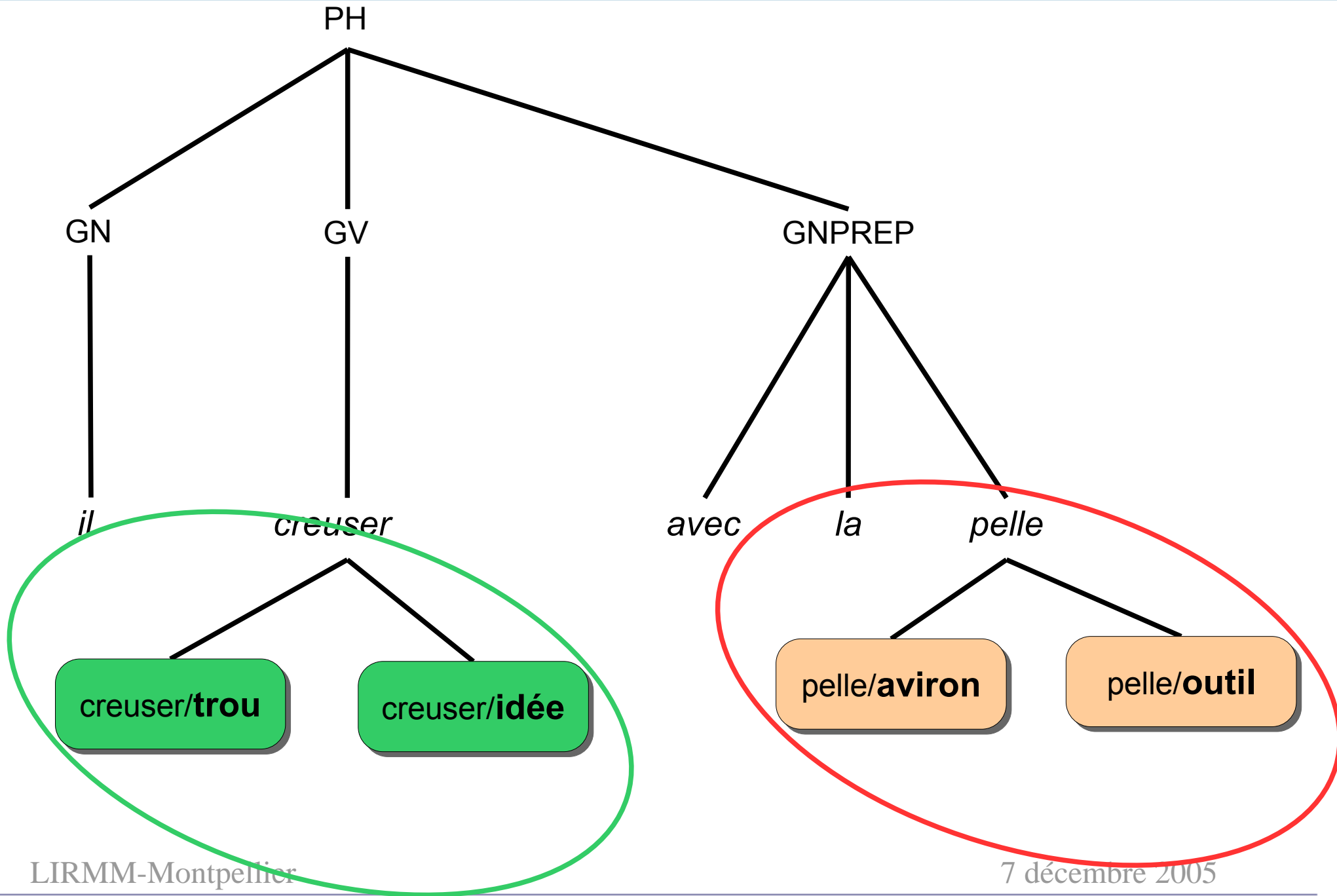














Attributs d'une fourmi f

Durée de vie : expérimentalement 20 cycles

Quantité de "sucre" transporté : $E(f)$

Référence à la fourmilière d'origine F : $V(f) = V(F)$

Mode de déplacement courant

$$P(\text{retour}) = \frac{E(f)}{E_{\max}(f)}$$

Restitution de la ressource à la mort

Deux modes de déplacement

Recherche de "sucre"

Retour vers la fourmillère



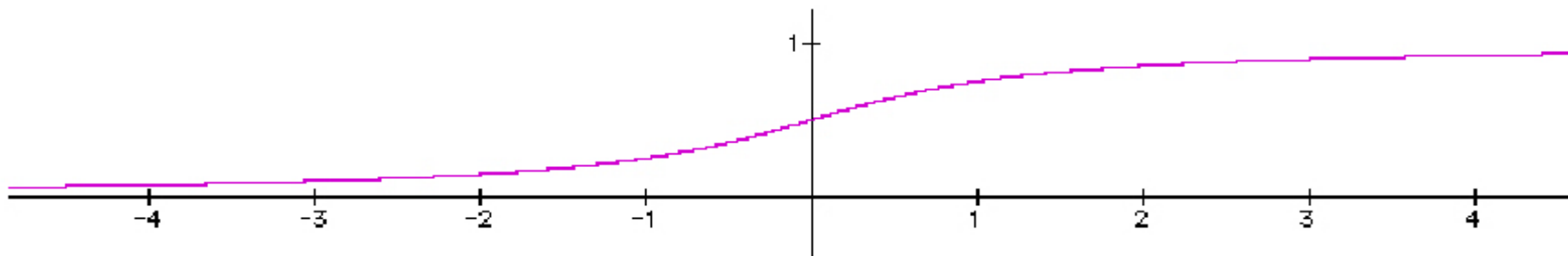
Production de fourmis

Une (ou plusieurs) fourmi à chaque cycle de la simulation

Selon une probabilité liée au niveau de sucre sur la fourmilière $E(F)$

fonction "sigmoïde"

coût de production (expérimentalement 1/10)





Pourquoi ?

Phénomènes linguistiques difficiles à résoudre
(potentiellement fonction des autres)

Critères difficiles à combiner
(parfois contradictoires, pondérations fonction des autres)

Comment ?

Déplacements

fourmi **cherche_hyper** favorisera un arc *hyperonymie*

fourmi **cherche_instr** favorisera un arc *instrument*

Phéromone de caste $\varphi_c(A)$



Déplacements pseudo-aléatoires

Evaluation des destinations possibles + liste tabou (k = 1)

Modification de l'environnement

Dépot de phéromone sur l'arc traversé : $\varphi_{t+1}(A) = \varphi_t(A) + \theta$

Coloration du vecteur du nœud atteint : $V_{t+1}(N) = V_t(N) \oplus \alpha V(f)$

Pont (interprétation)

Création à partir d'une fourmillière $F_i \in F_A$ vers F

Disparition du pont si la phéromone est entièrement évaporée

creuser/**trou**

pelle/**outil**

$$\varphi_{t+1}(A) = \varphi_t(A) - D\delta$$

où D est la distance entre F_A et F

$$\varphi_t(A) = \theta$$
$$\varphi_c(A) = \theta$$



Principe général

Position courante = un nœud N avec k voisins N_i via des arcs A_i

Fonction d'évaluation de chaque $N_i = \text{Eval}(N_i)$

Fonction d'évaluation de chaque $A_i = \text{Eval}(A_i)$

Fonction d'évaluation globale de la destination $dest_i$

$$\text{Eval}(dest_i) = \max \left\{ \begin{array}{c} \text{Eval}(N_i) + \text{Eval}(A_i) \\ \epsilon \end{array} \right\}$$

$$P(N_i, A_j) = \frac{\text{Eval}(dest_i)}{\sum_{k=1}^n \text{Eval}(dest_k)}$$



Mode recherche

- attraction des nœuds qui ont beaucoup de sucre
- attraction des arcs qui ont peu de phéromone de passage

$$Eval_E(N_i, A_j) = E_f(N_i)$$

$$Eval_\varphi(N_i, A_j) = 1 - \varphi_f(A_j)$$

Mode retour

- attraction des nœuds dont l'odeur est proche de la leur
- attraction des arcs excités

$$Eval_{odeur}(N_i, A_j) = 1 - \frac{\pi}{2} D_A(V(f), V(N))$$

$$Eval_\varphi(N_i, A_j) = \varphi_f(A_j)$$



Caste

- attraction de la phéromone de leur caste
- attraction de la relation correspondant à leur caste

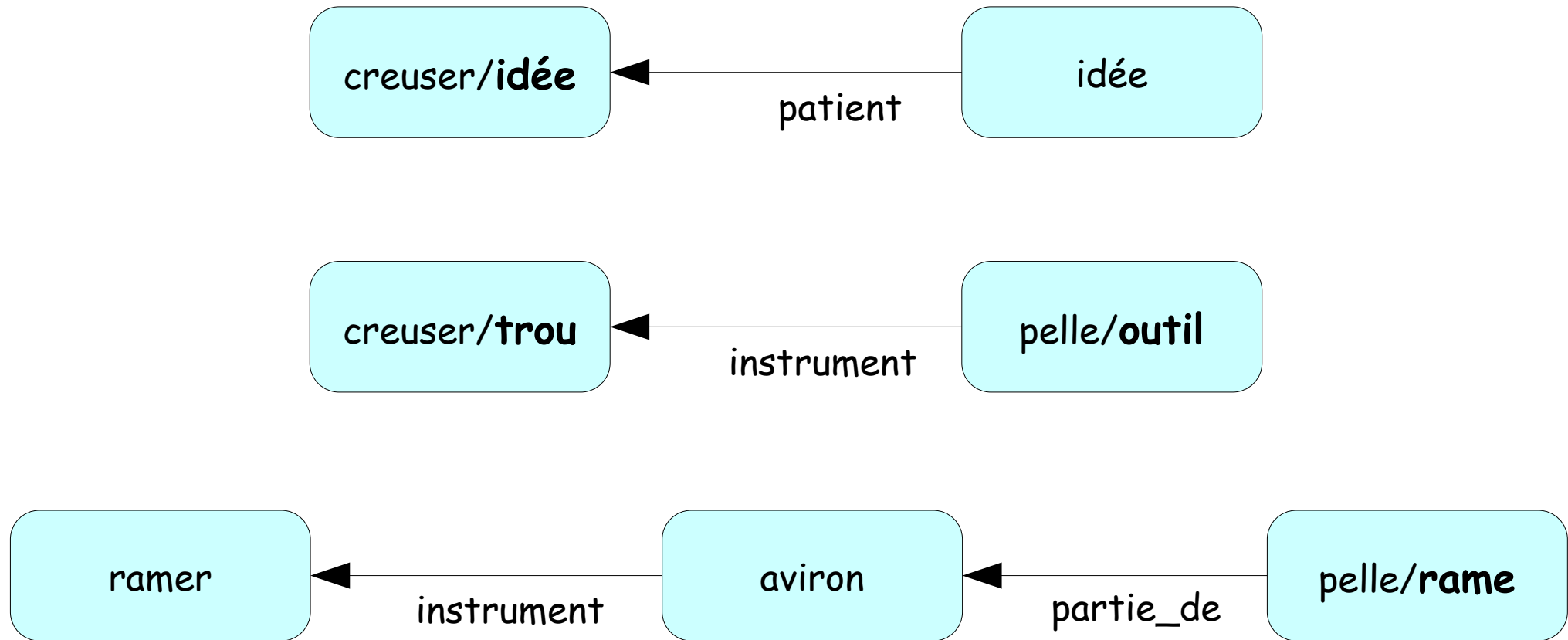
$$Eval_{\varphi_c}(N_i, A_i) = \varphi_{c_f}(A_j)$$

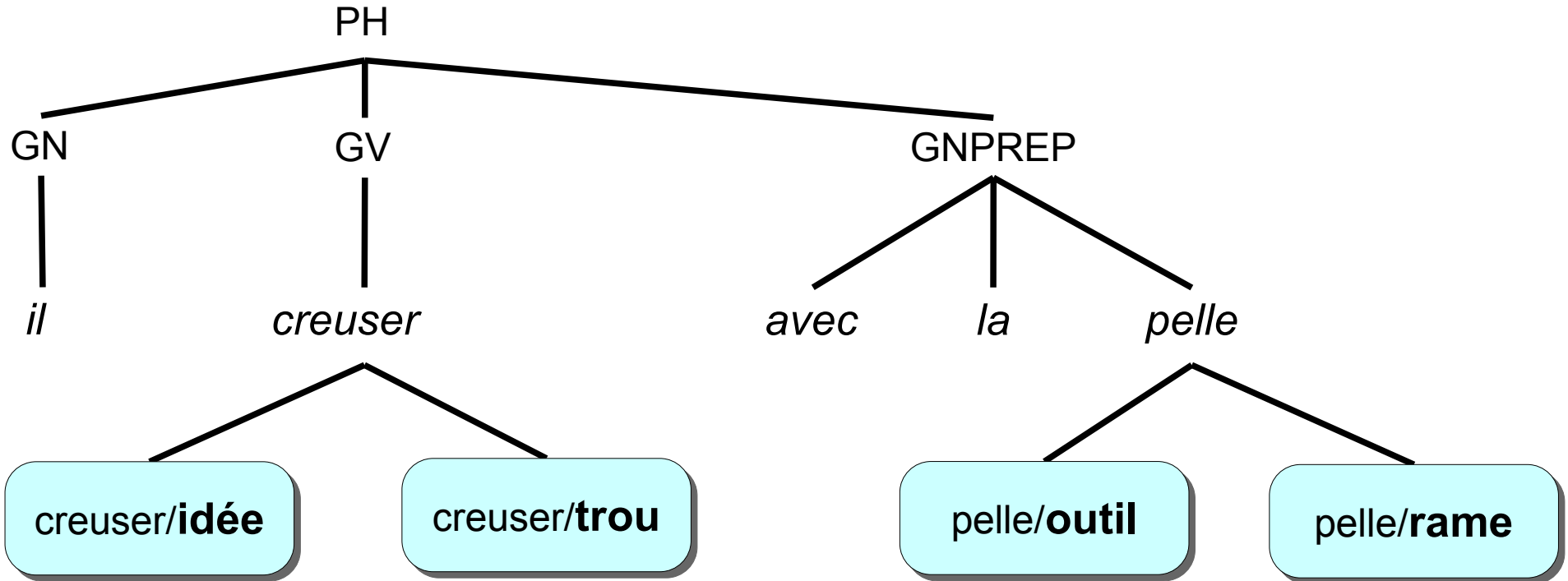
$$Eval_{caste}(N_i, A_i) = \begin{cases} 3 & \text{si } caste(f) = relation(A_j) \\ 0 & \text{sinon} \end{cases}$$

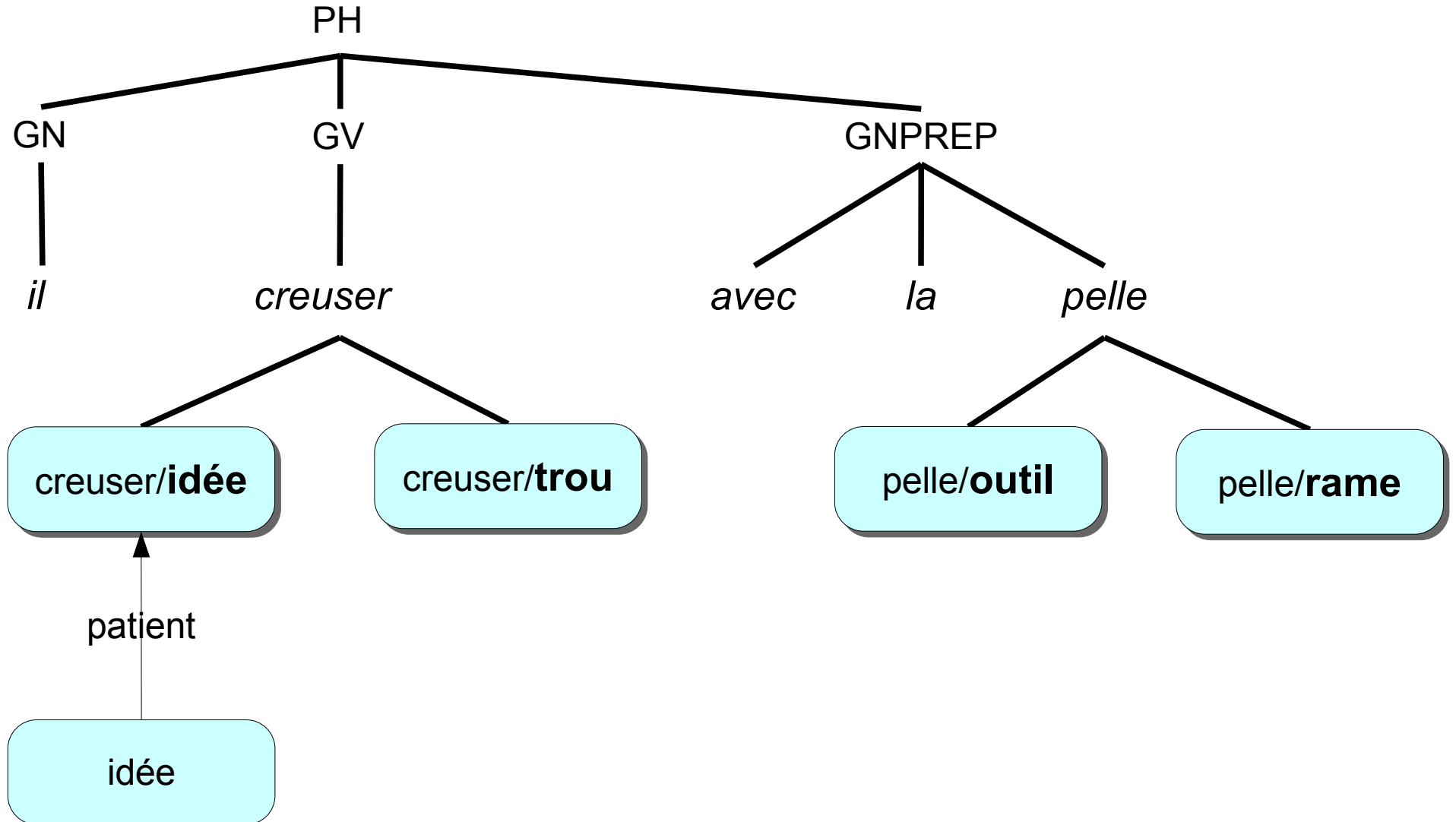


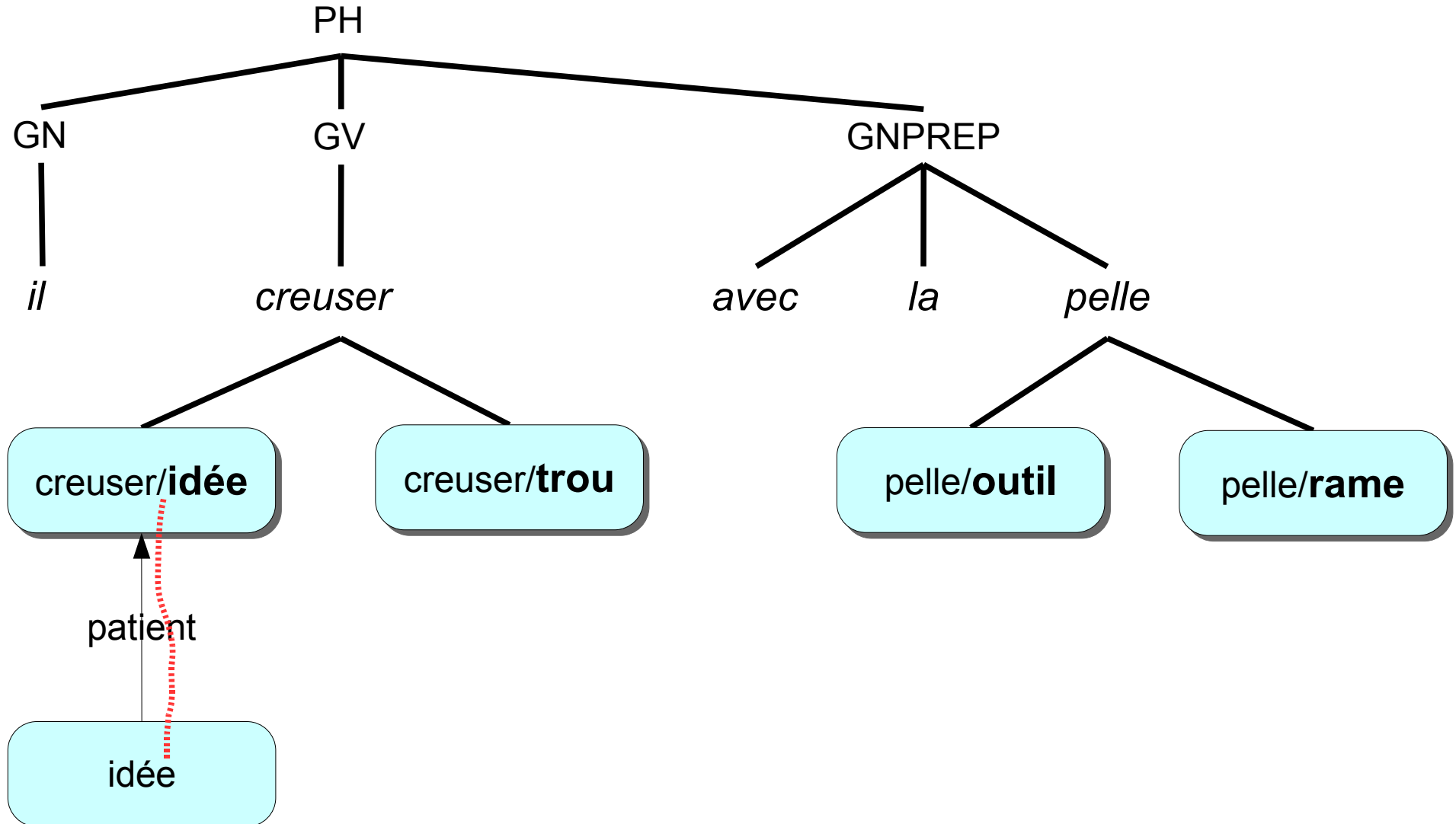
Exemple : réseau lexical

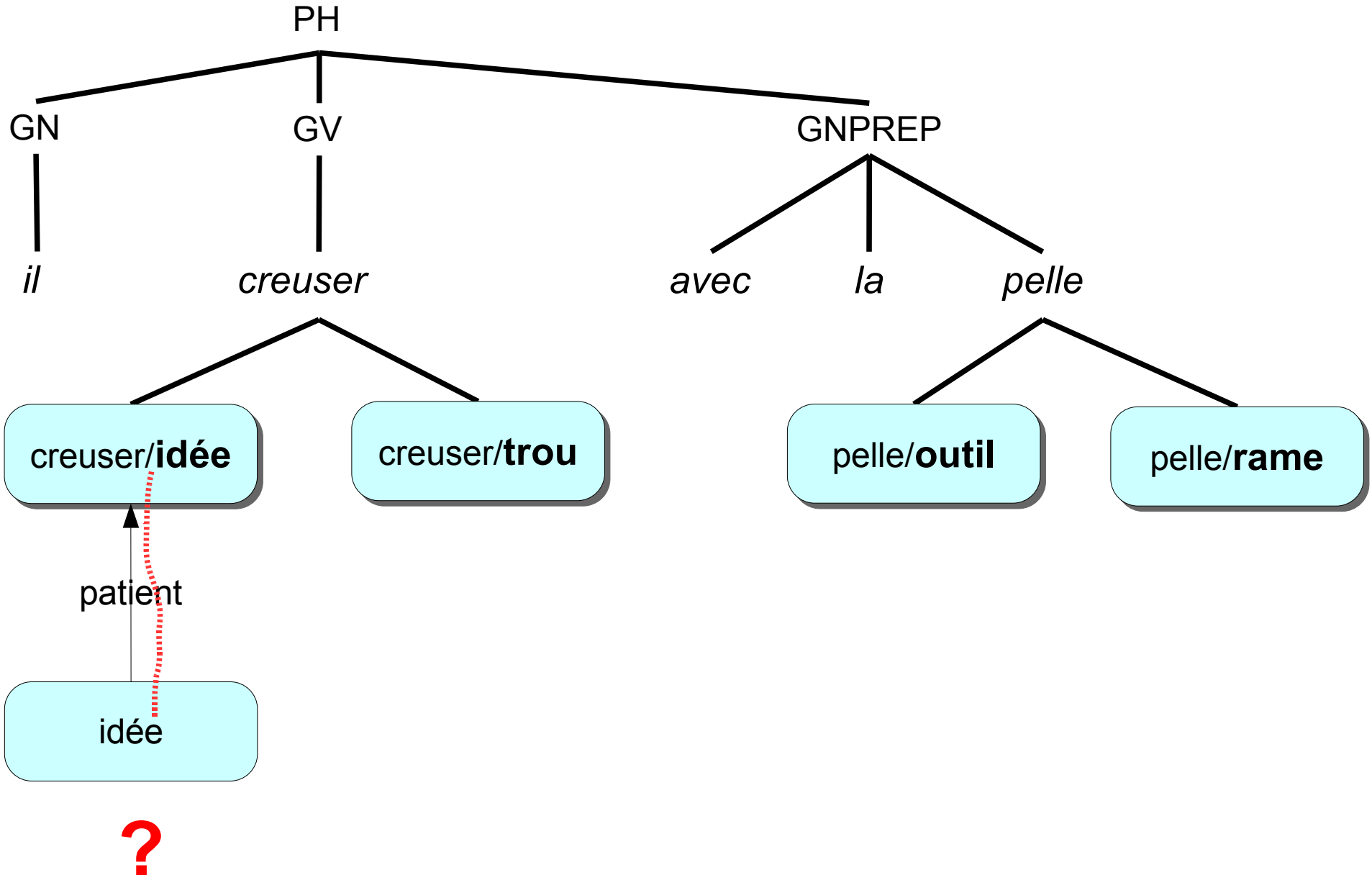
LIRMM > analyse sémantique > exemple

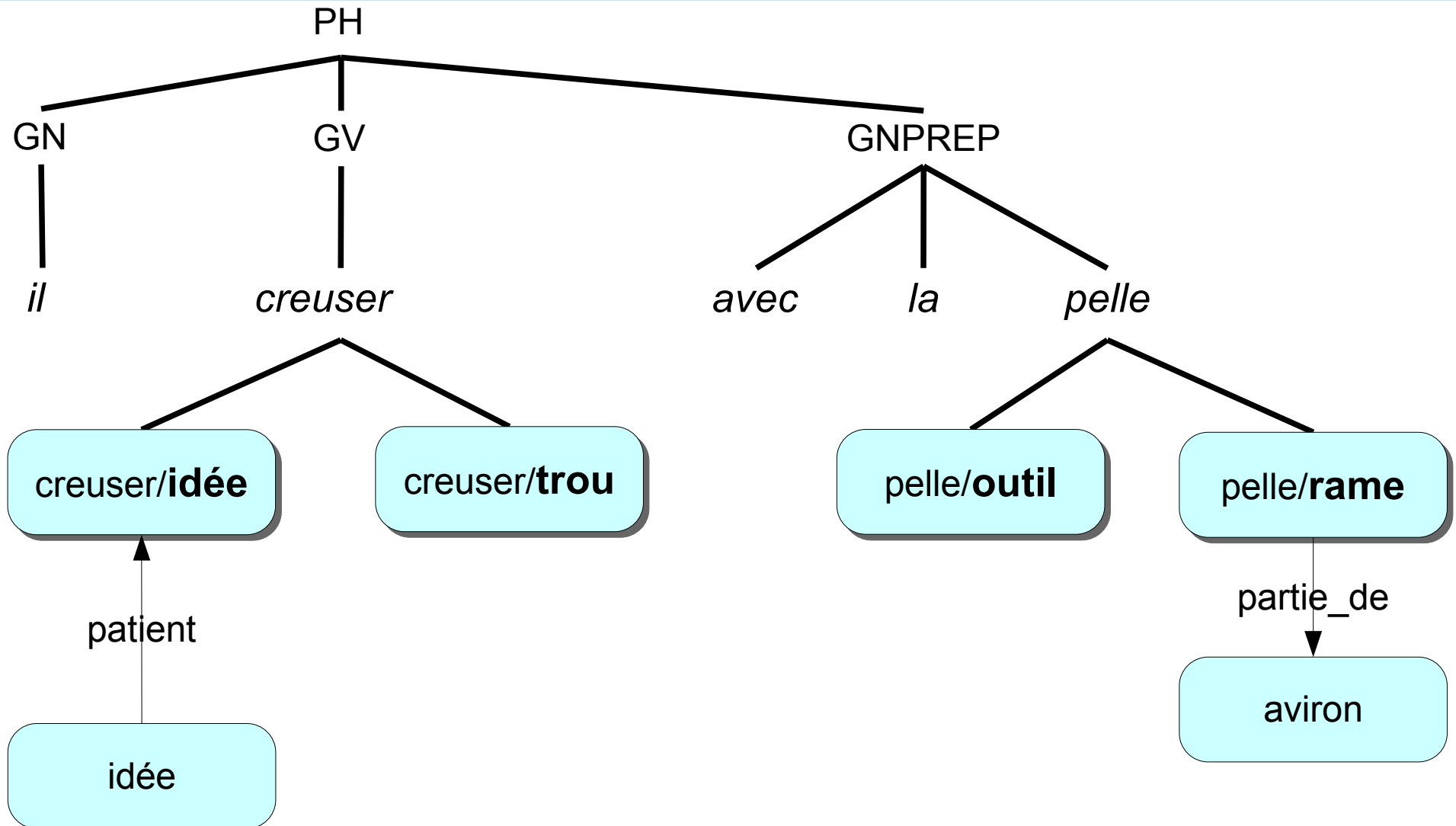






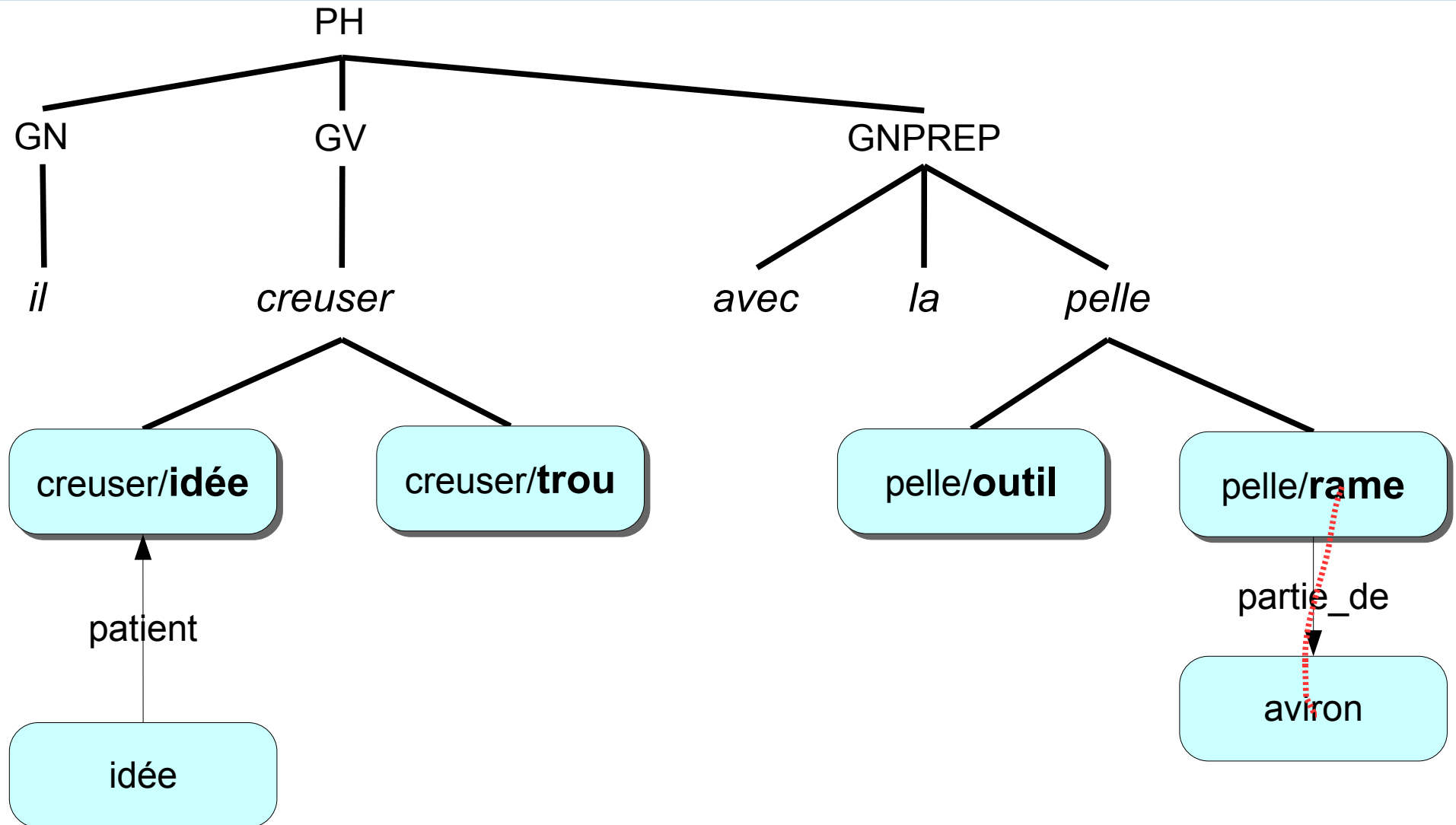






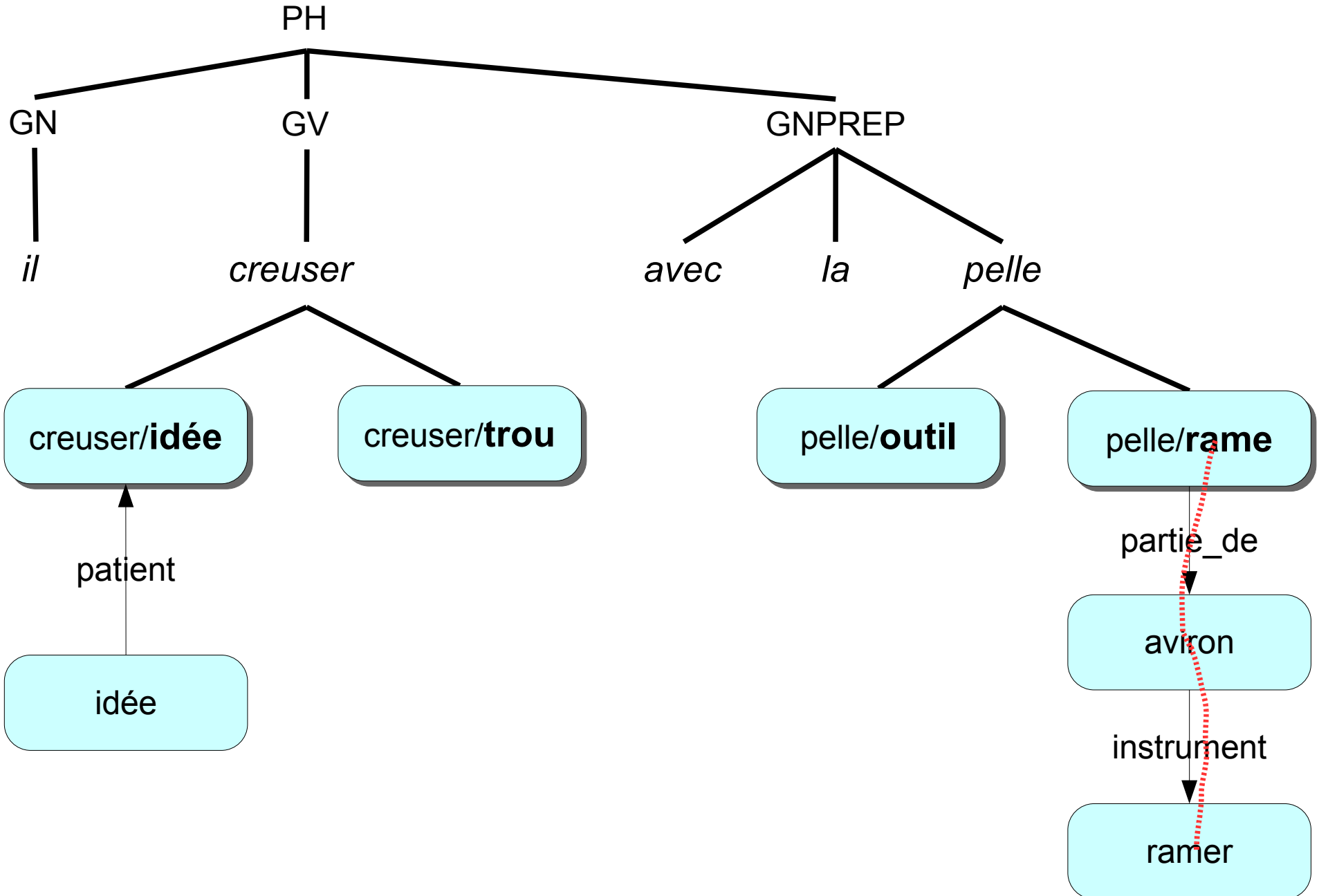


Exemple



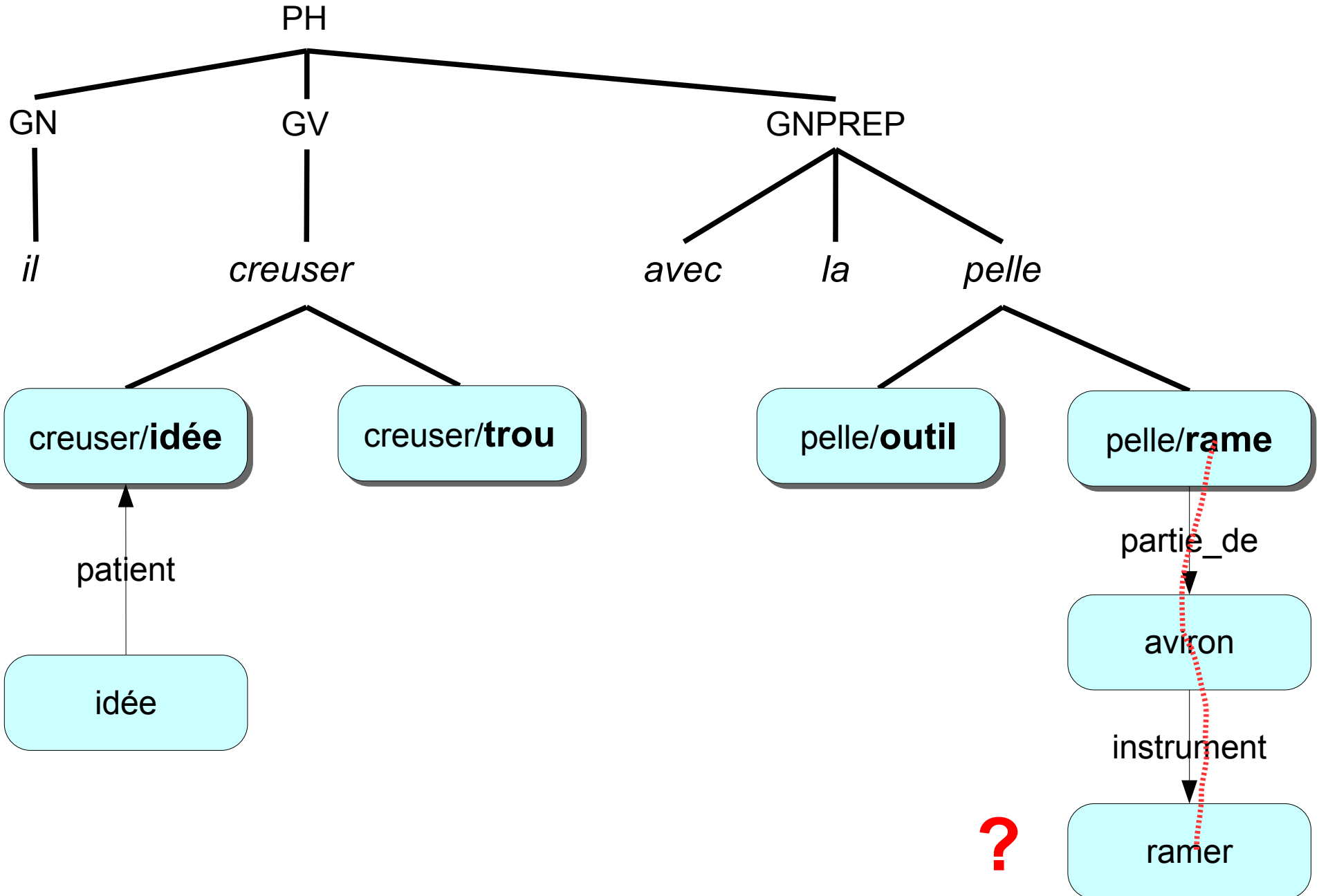


Exemple



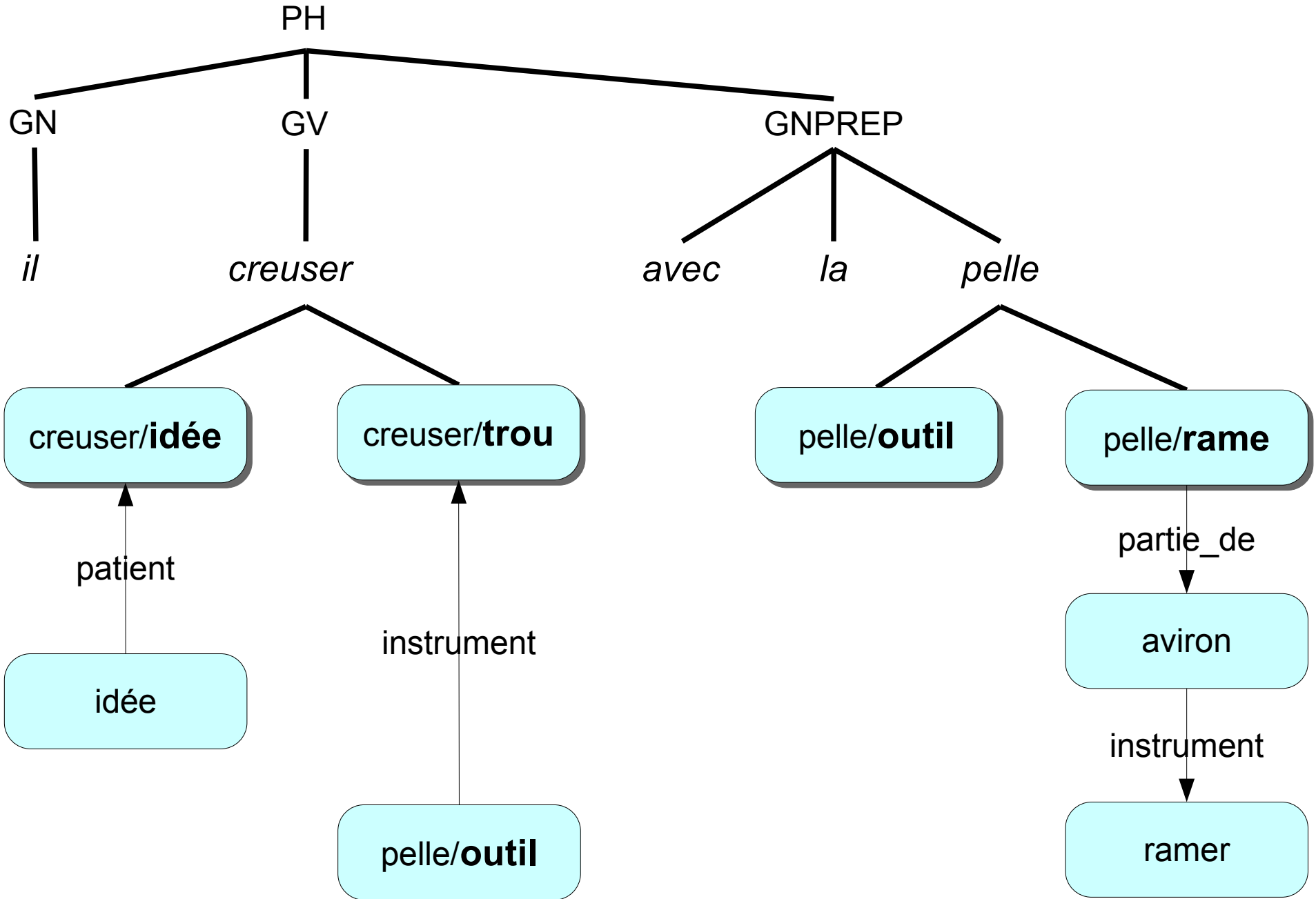


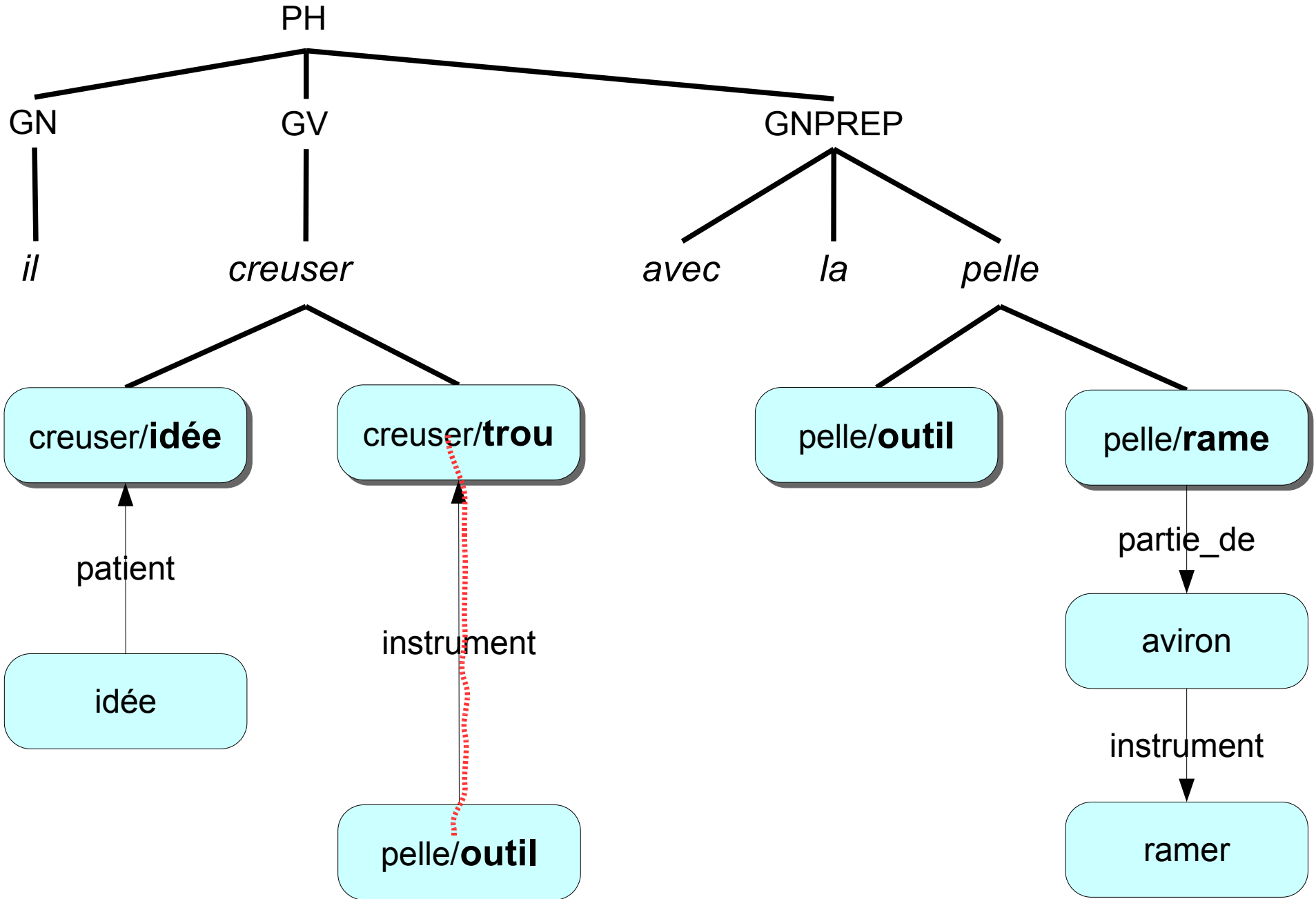
Exemple

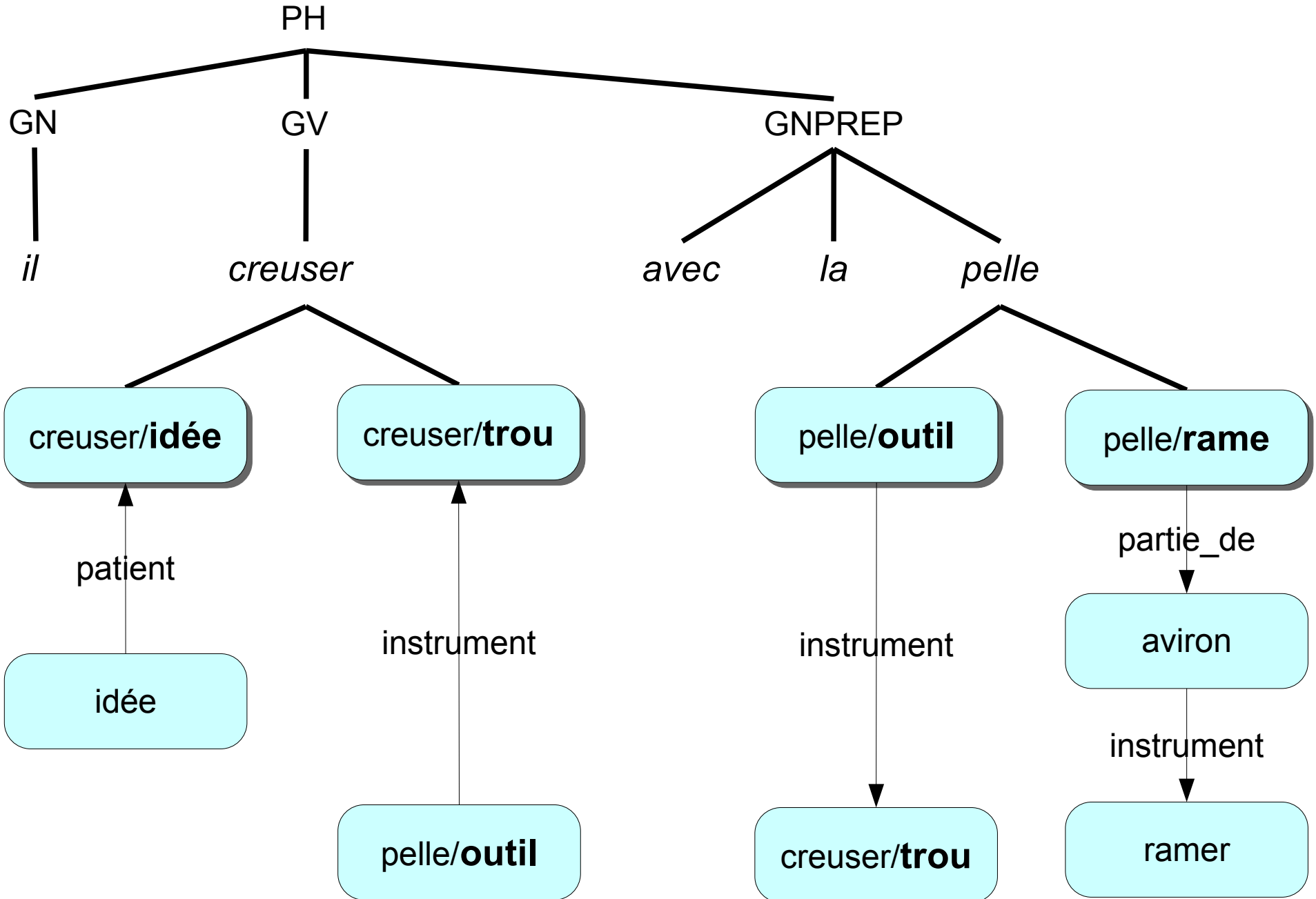




Exemple

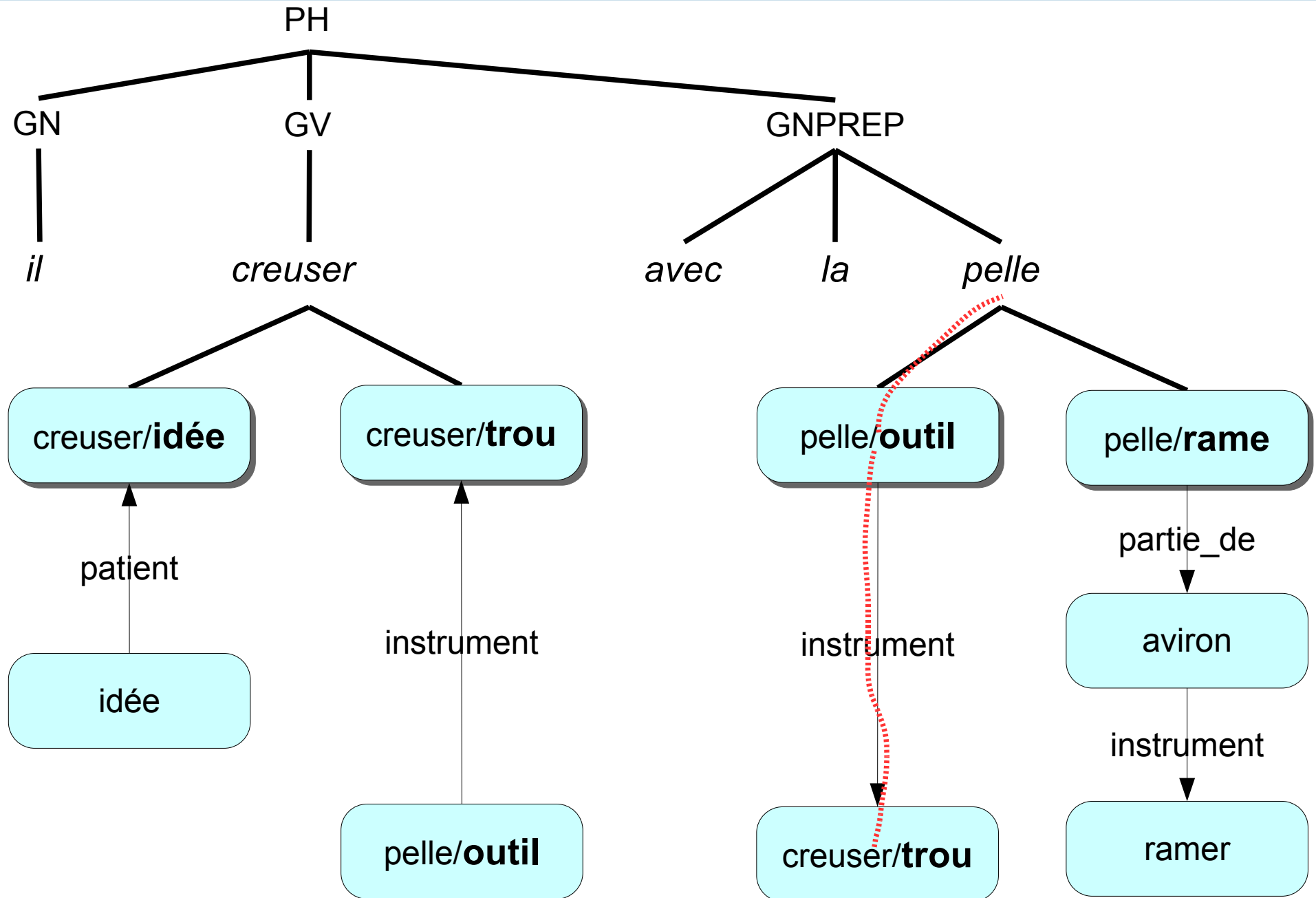






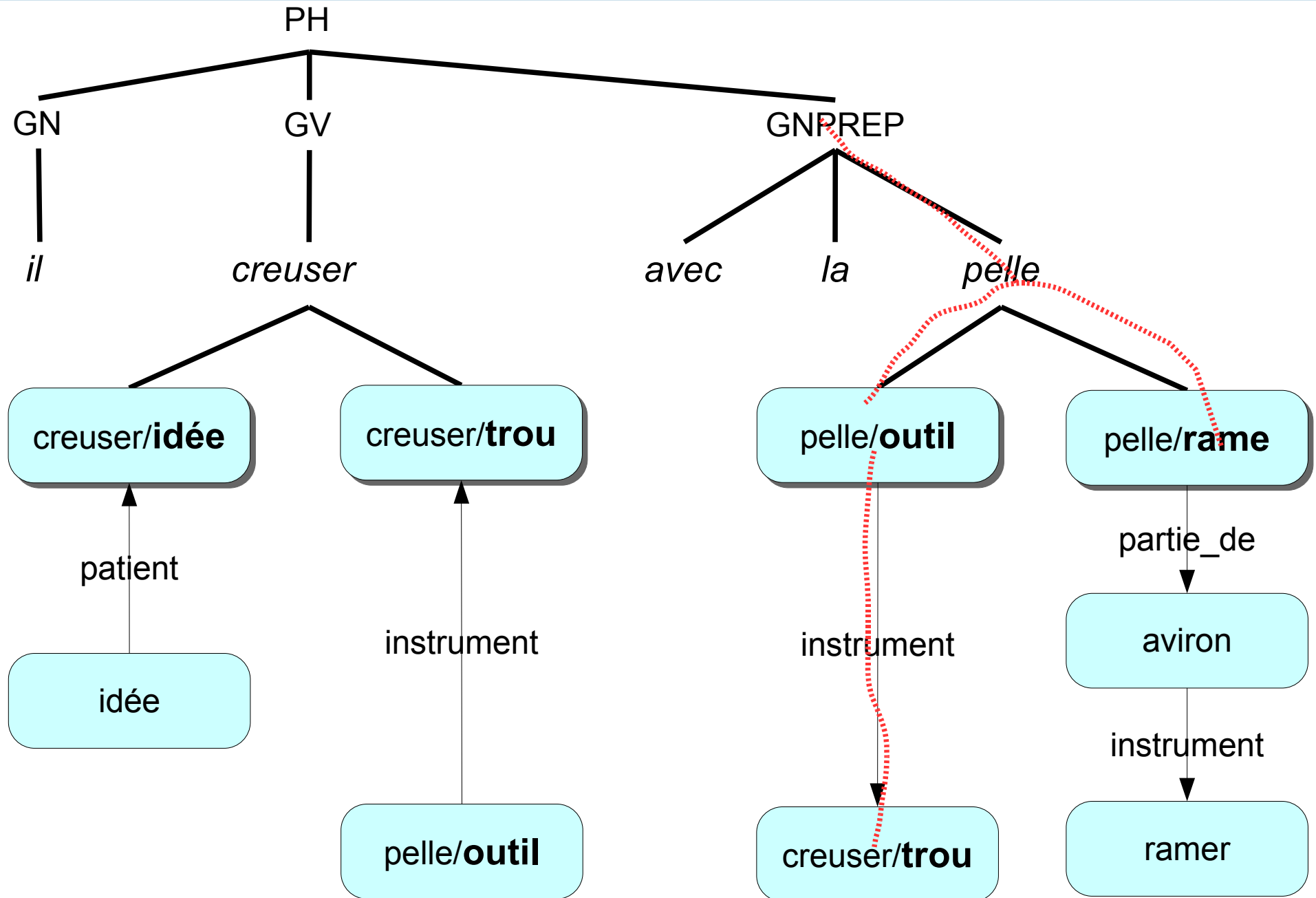


Exemple



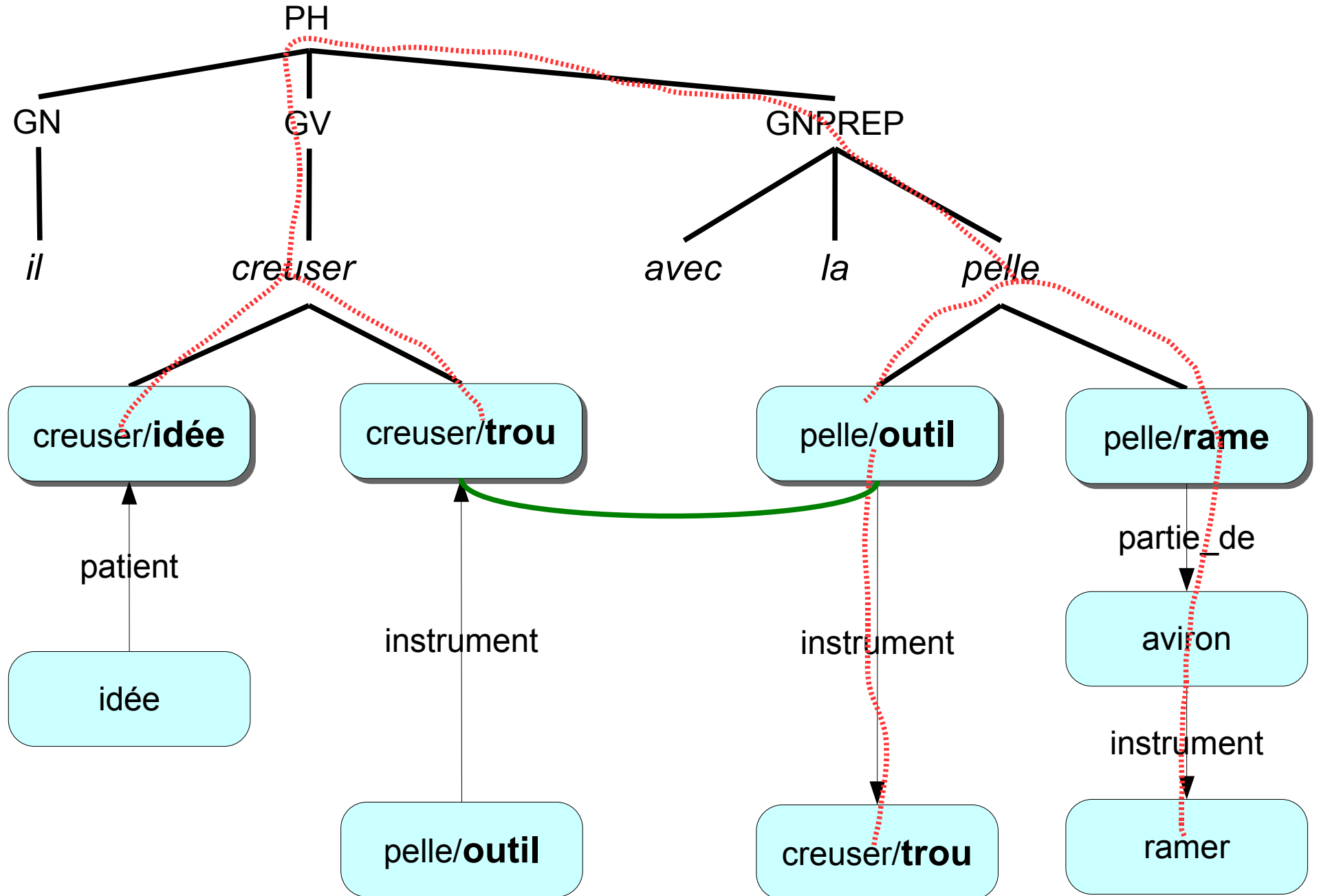


Exemple



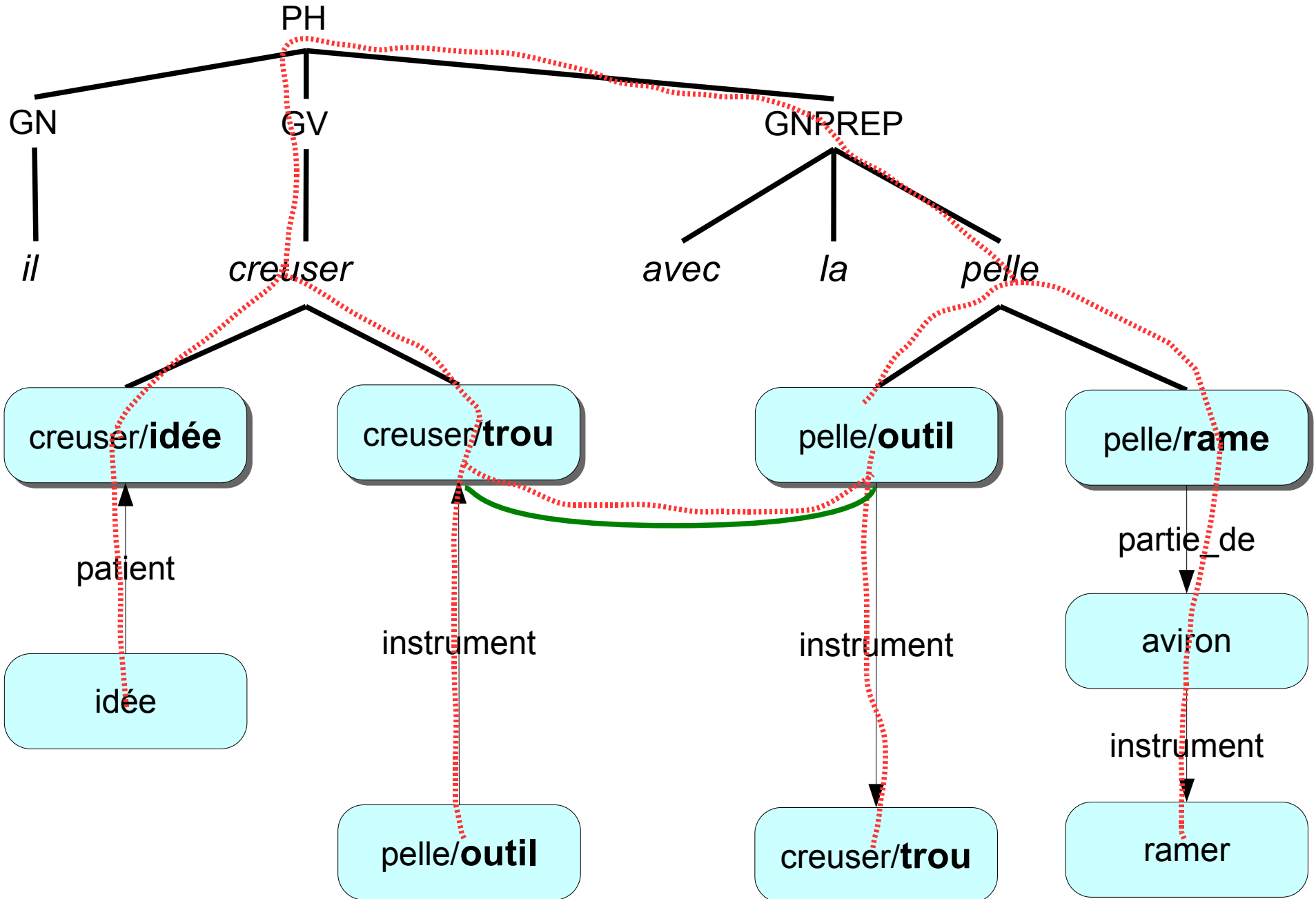


Exemple



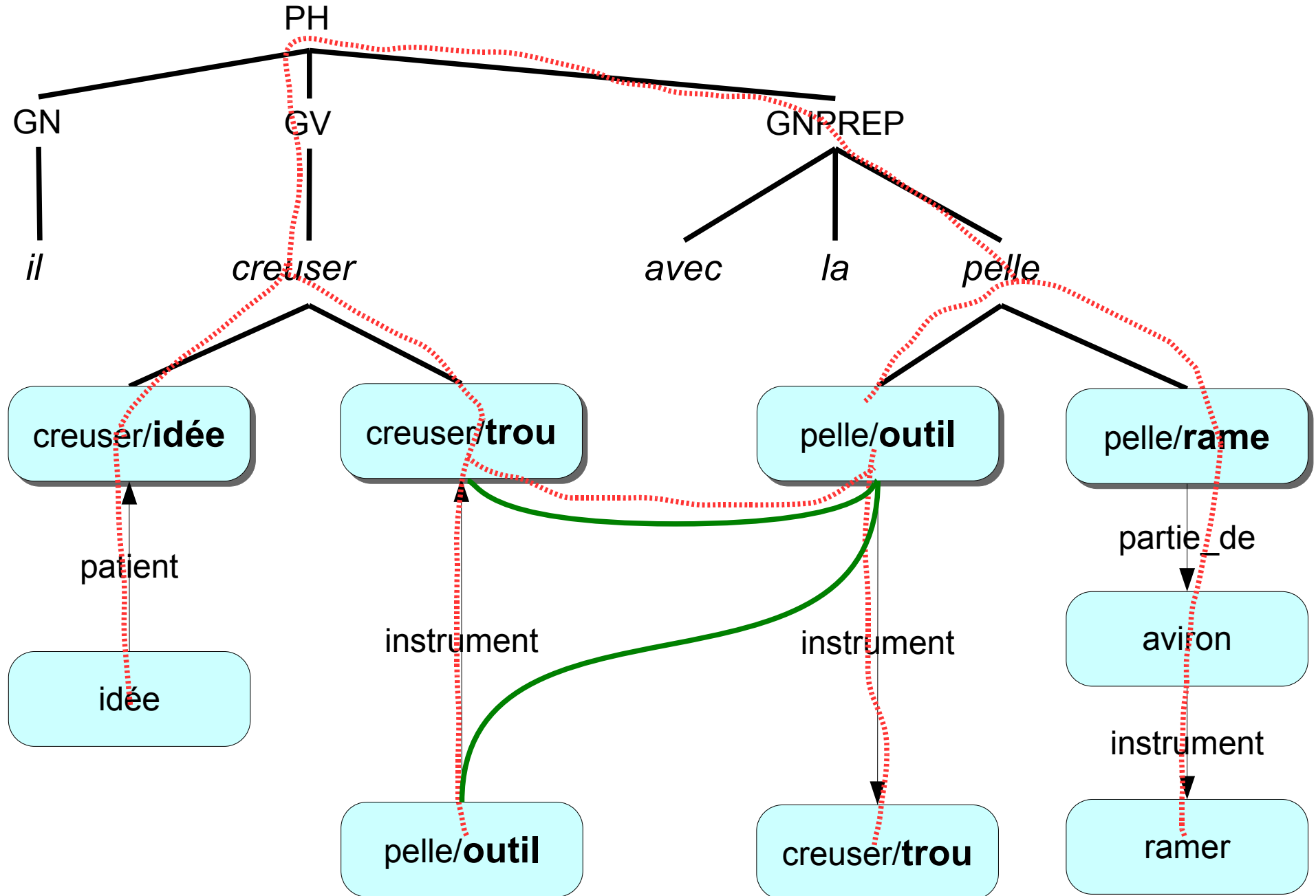


Exemple



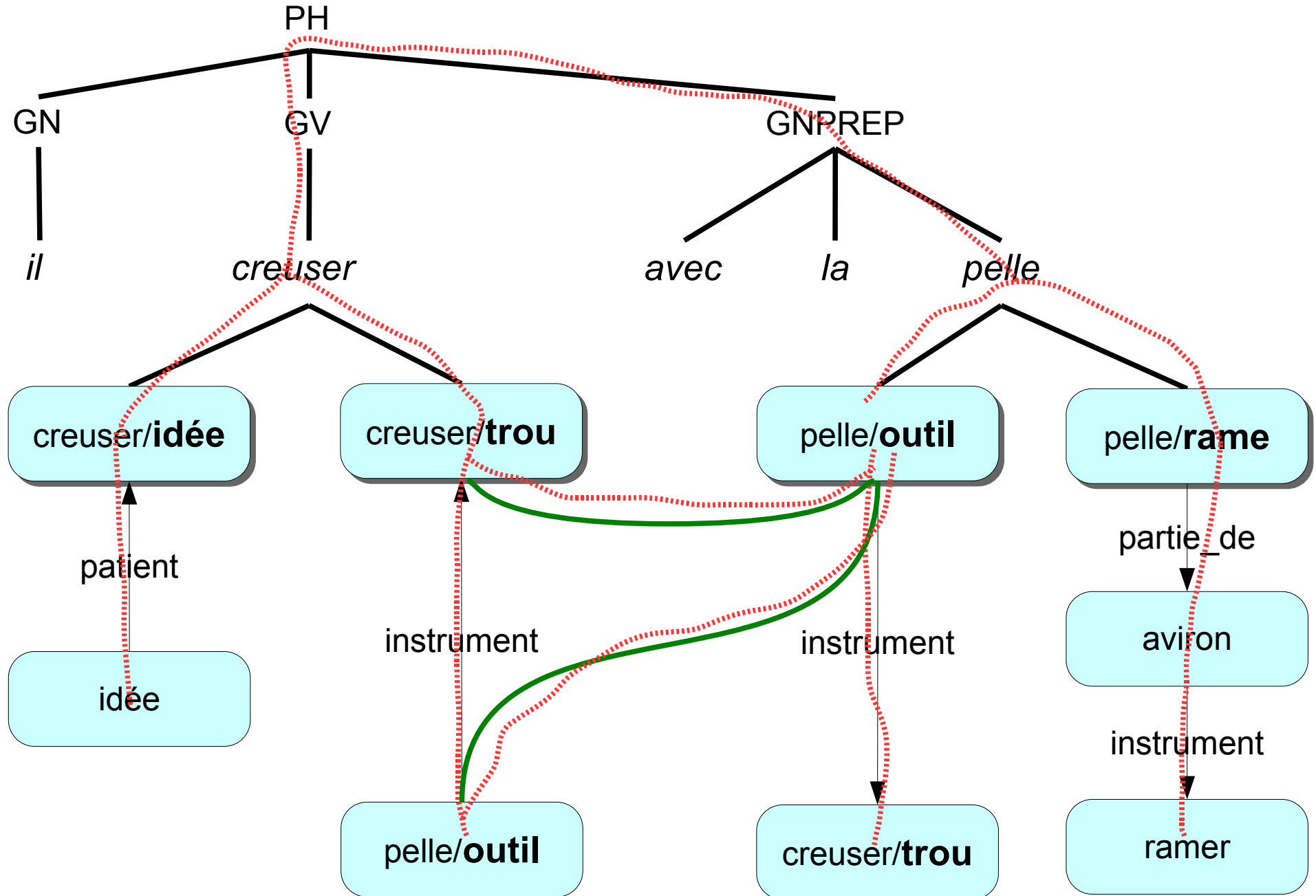


Exemple



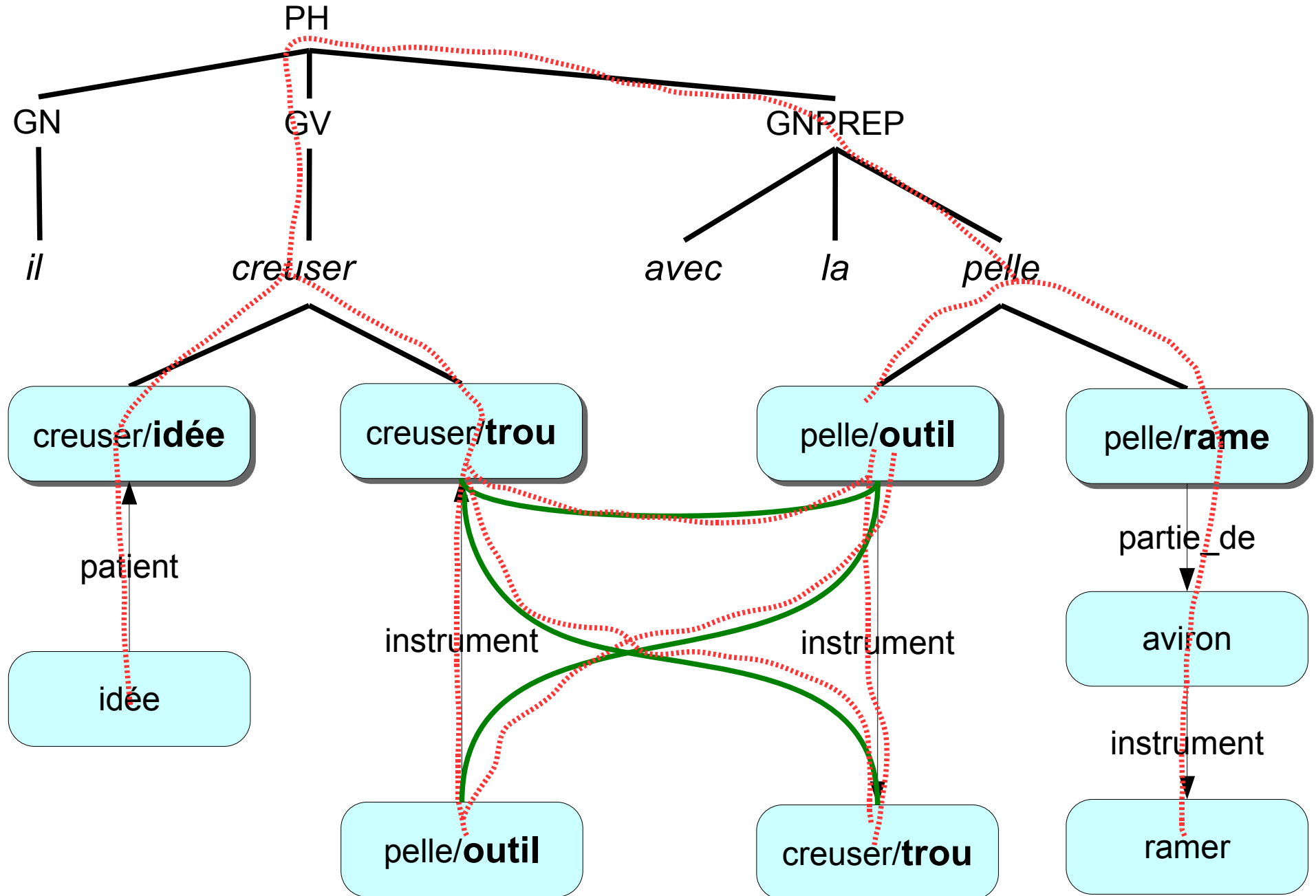


Exemple



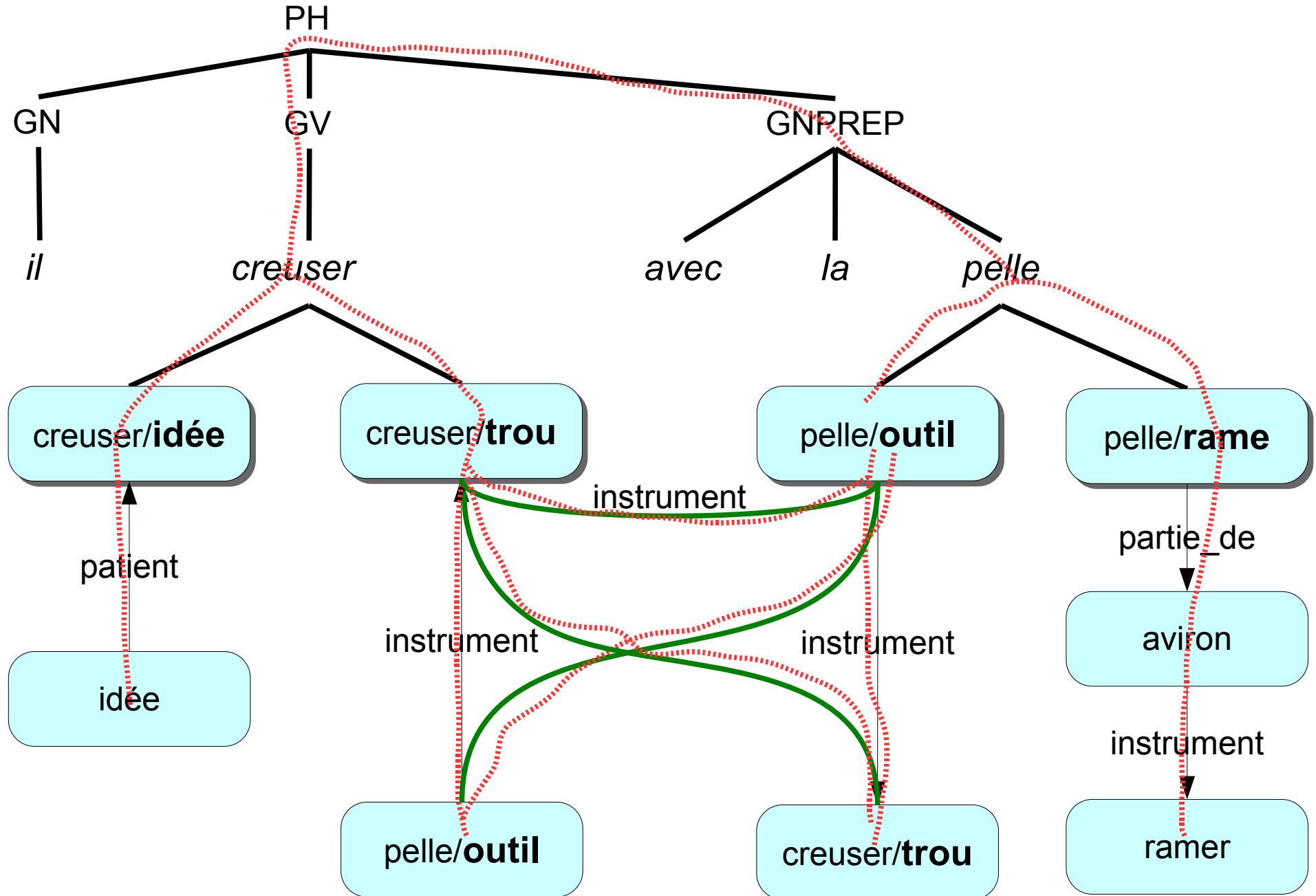


Exemple





Exemple





Sémantique = création de chemins entre objets lexicaux

Chemins = conjointement thématique et réseau lexical

Plus il y a de chemins, plus les ponts sont forts, plus il y a de circuits

Plus il y a de circuits, plus les fourmilières impliquées monopolisent de ressources (fourmis + énergie)

Les castes favorisent certaines circulations de fourmis par rapport à certains types d'arcs.



Castes

cherche_Syn, cherche_hyper, cherche_hypo

cherche_Magn, cherche_Ver, cherche_Bon

cherche_Mero, cherche_Holo

cherche_Instr

cherche_agt, cherche_pt, cherche_dest, cherche_pred

Corpus

41 textes courts (une à deux phrases)

Exemples caractéristiques des phénomènes étudiés

Web, œuvres classiques (Stendhal, Verne), jouets



acceptions	global	noms	adjectifs	verbes	adverbes
rappel	0,79 (+4%)	0,79 (+5%)	0,79 (+3%)	0,74 (+6%)	0,82 (+0%)
précision	0,78 (+3%)	0,78 (+2%)	0,82 (+4%)	0,76 (+4%)	0,8 (+2%)
arcs	global	interprétation	ratt.prep.	références	instanciations FLA
rappel	0,82 (+100%)	0,81 (+14%)	0,83 (+2%)	0,78 (+∞)	0,83 (+∞)
précision	0,85 (+98%)	0,82 (+12%)	0,89 (+6%)	0,81 (+∞)	0,87 (+∞)

instanciation FLA	Syn	Hyper	Magn	Ver	Bon	Mero	Holo	Instr	Agt	Pat	Dest
rappel	0,85	0,77	0,7	0,72	0,73	0,62	0,7	0,72	0,72	0,6	0,61
précision	0,86	0,78	0,72	0,74	0,75	0,66	0,69	0,7	0,74	0,62	0,62



Première étude

Faisabilité de l'approche

Permet de résoudre en partie les ambiguïtés sémantiques présentées

Pistes à explorer

Gestion de l'antonymie (transformation des vecteurs des fourmis)

Auto-arrêt du système

Problème des formes passives

« *L'avocat a été mangé.* »

Utiliser plus d'informations syntaxiques

Passage du discret au continu



Conclusion et perspectives



La détection des FLA améliore l'analyse sémantique

Pour la modélisation des FLA

- Modélisation et mise en œuvre d'une base lexicale sémantique
 - Approche hybride (RLV et vecteurs conceptuels)
 - Construction et exploitation conjointe
- FLA d'évaluation et de construction

Pour la détection et l'exploitation des FLA

- Extension des premiers modèles algorithmes à fournis pour l'AS



Tenir compte du domaine dans le réseau lexical

Approche dynamique

Introduction de nœuds de domaine liés aux acceptions pertinentes

Modification du modèle d'algorithmes à fourmis

Génération de sucre sur les nœuds domaines

Taux de génération en fonction du niveau d'excitation

Intérêt : domaine supposé au départ ou émergence du domaine

Multilinguisme

"Papilloniser" le réseau lexical

Introduire des nœuds axes (acceptions interlingues)

des attributs *langue* pour les acceptions et pour les FLACL

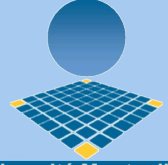
(les FLACM n'ont pas à être signées)



LIRMM



Centre National de la
Recherche Scientifique



Université Montpellier
II

MERCI