

Une analyse à la Gaume de la flexion verbale du français.

Gilles Boyé

Université Bordeaux-Montaigne & CLLE-ERSSàB

Analyse de la flexion verbale

- Analyser les désinences :
 - faire une description syntagmatique de la conjugaison capte une partie infime de la complexité du système
 - les désinences sont identiques pour tous les verbes sauf une poignée de formes exceptionnelles
 - la compositionnalité des exposants dépend d'hypothèses sur l'organisation des temps et des modes
- Analyser la flexion verbale :
 - faire une description paradigmaticque de la conjugaison et décrire les correspondances entre formes
 - contraintes générales (tous les verbes)
 - contraintes paroissiales (par *groupes*)

Espaces thématiques à la BoBo

- Des contraintes générales
 - partition du paradigme en zones correspondant à un même radical
 - relations régulières entre radicaux de zones différentes avec un graphe de régularité
- Des contraintes paroissiales
 - les verbes irréguliers échappent aux correspondances régulières et lexicalisent certains thèmes qui ne se correspondent pas

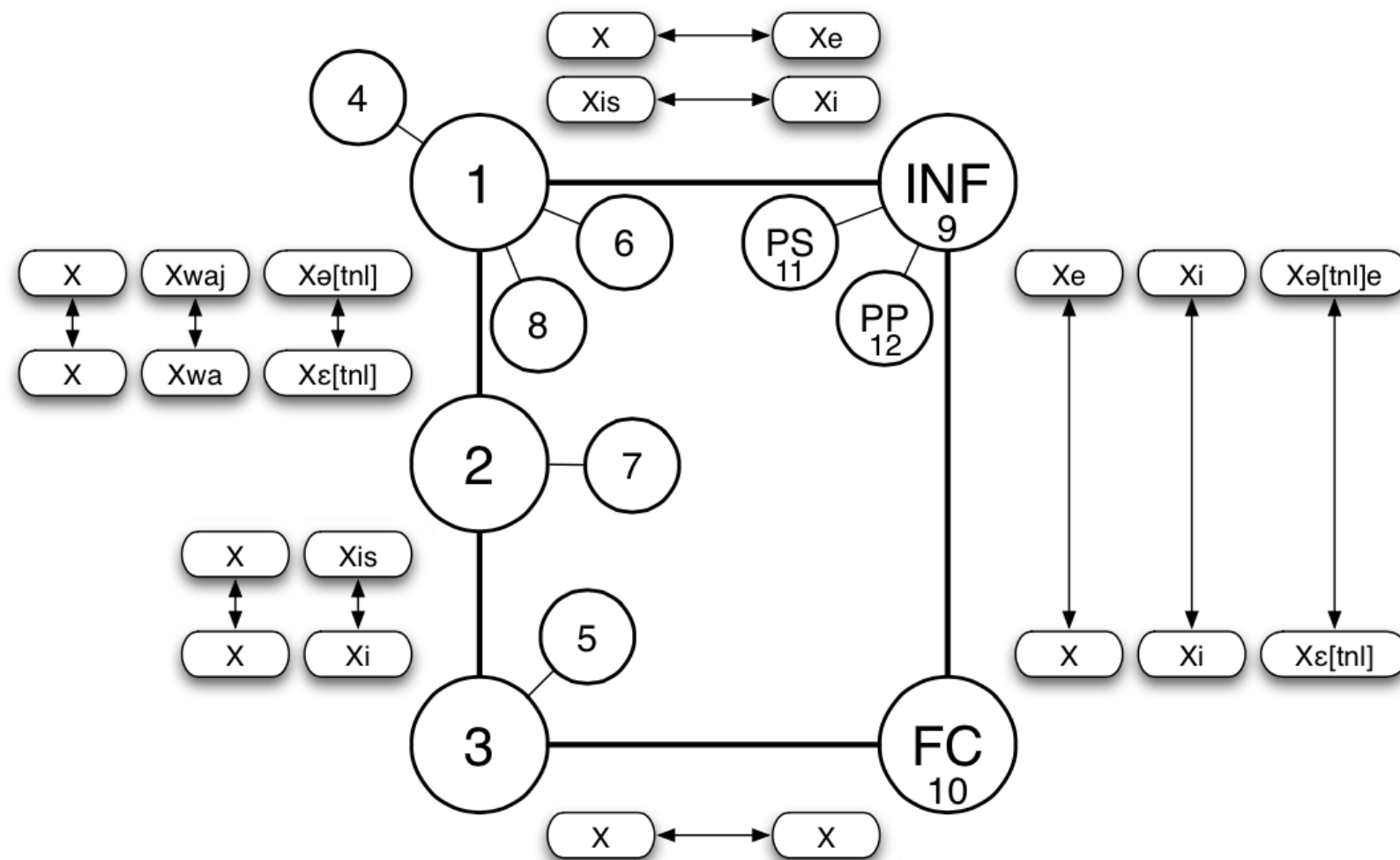
Espaces thématiques à la BoBo

Indice Formes

- 1 Radical par défaut
- 2 PRÉSENT 1SG, 3PL ; SUBJONCTIF SG, 3PL
- 3 PRÉSENT 23SG, IMPÉRATIF 2SG
- 4 PRÉSENT 1PL ; SUBJONCTIF 12PL

Indice Formes

- 5 PASSÉ 13SG, 3PL
- 6 FUTUR, CONDITIONNEL
- 7 PARTICIPE PASSÉ
- 8 INFINITIF



Quelques questions avec BoBo

- Des choix arbitraires, des représentations abstraites, du travail manuel
- Généralisations sur les régularités
 - Distillation du paradigme (abstraite)
 - Relations régulières entre thèmes (arbitraires)
- Contraintes sur la distribution des irrégularités
 - Réseau de relations (manuel)
 - Relations symétriques (simples mais inappropriées)
- Mais une problématique simple pour l'analyse de la flexion :
- Comment remplir un paradigme flexionnel ?

Le problème du remplissage

- Le problème du remplissage se pose à différents niveaux :
 - Remplissage d'une case
 - Connaissant une autre case (Albright 2002)
 - Connaissant plusieurs cases (Bonami)
 - Remplissage du paradigme (Boyé 2000)
- Et sous plusieurs formes
 - Génération de formes
 - syntagmatique
 - paradigmatic (Albright, Bonami, Boyé, ...)
 - Attribution de classe flexionnelle
 - entropie (Ackerman, Malouf, Blevins, Bonami, ...)
 - clustering (Bonami)

Albright 2002

- Minimal Generalization Learner
 - vise à générer le contenu d'une case du paradigme à partir du contenu d'une autre case
 - basé sur des analogies morpho-phonologiques (règles SPE)
 - étudie des paires de formes en entrée pour évaluer la confiance associée à chaque analogie
 - propose en sortie la forme candidate correspondant au meilleur taux de confiance

Principe de classification des alternances

- Pour une paire de formes du paradigme, on classe les alternances entre la forme d'entrée et la forme de sortie
- par exemple, pour l'imparfait indicatif 1 et l'indicatif présent 6 :

LEXEME	IPF.1	PRS.6
PASSER	pasε	pas
BOIRE	byvε	bwav
FINIR	finisε	finis
MENER	mənε	mən
SORTIR	sɔrtε	sɔrt
PRENDRE	prənε	prən
...

The diagram illustrates the classification of alternances between the IPF.1 and PRS.6 forms. Arrows point from the ε in the IPF.1 column to the corresponding alternance boxes on the right:

- From **PASSER** (pasε) to **ε → ∅**
- From **BOIRE** (byvε) to **yvε → wav**
- From **FINIR** (finisε) to **ε → ∅**
- From **MENER** (mənε) to **ənε → ən**
- From **SORTIR** (sɔrtε) to **ε → ∅**
- From **PRENDRE** (prənε) to **ənε → ən**
- From the bottom row (dots) to **... → ...**

Minimal generalizations

- Calcul de toutes les **analogies** existantes entre deux cases d'un paradigme
- Pour chaque analogie, le MGL calcul **tous les contextes** intermédiaires rencontrés
 - La première paire donne un contexte unique
 - Chaque nouvelle paire sert à généraliser l'analogie en se combinant avec tous les contextes antérieurs
- Au final, on obtient un treillis pour chaque analogie qui relie le contexte global à tous les contextes uniques
 - par exemple, du passé simple 2SG vers l'infinitif, on a :
 - 12900 règles pour 32 analogies
 - 6440 transformations spécifiques

Islands of reliability

- La plus grande partie du treillis est redondante
- il existe un contexte plus grand de même indice de confiance

y -> iR / X [J, N, b, d, g, k, m, n, p, t] uR	9	91.0	0.898
y -> iR / X kuR	8	81.0	0.887
y -> iR / X [2, 6, 9, @, E, H, O, R, S, Z, a, e, f, i, j, l, o, s, u, v, w, y, z, â, ê, è, ò, ô, û] kuR	7	71.0	0.872
y -> iR / X [2, 6, 9, @, E, H, O, R, a, e, i, j, l, o, u, w, y, â, ê, è, ò, ô, û] kuR	6	61.0	0.852
y -> iR / X [2, 6, 9, @, E, O, a, e, o, â, ê, è, ò, ô, û] kuR	5	51.0	0.825
y -> iR / X [6, @, R, S, Z, f, l, s, v, z] kuR	4	41.0	0.786
y -> iR / X [6, @, a, â, è, ò] kuR	4	41.0	0.786
y -> iR / X [6, @, O, o, â, ò, ô] kuR	4	41.0	0.786
y -> iR / X [6, @, a, è] kuR	3	31.0	0.718
y -> iR / X [6, @, O, o, ò, ô] kuR	3	31.0	0.718
y -> iR / X [6, @, â, ò] kuR	3	31.0	0.718
y -> iR / X [6, @, R, l] kuR	3	31.0	0.718
y -> iR / X [6, @, è] l	3	31.0	0.718
y -> iR / [a, â] kuR	2	21.0	0.57
y -> iR / X [O, o, â, ò, ô] kuR	2	21.0	0.57
y -> iR / [J, N, b, d, g, k, m, n, p, t] uR	2	21.0	0.57
y -> iR / X [R, f, l, s, v, z] kuR	2	21.0	0.57
y -> iR / [R, f, l, s, v, z] @kuR	2	21.0	0.57
y -> iR / X èl	2	21.0	0.57
y -> iR / X [R, l]	44	130.295	0.163
y -> iR / X l	21	40.19	0.123

- Une partie du treillis contient des contextes plus fiables que la généralisation globale
- les **ilots de confiance** sont formés par des contextes restreints où l’analogie est plus fiable que dans le cas général

y -> iR / X [J, N, b, d, g, k, m, n, p, t] uR	9	91.0	0.898
y -> iR / X [6, @, è] l	3	31.0	0.718
y -> iR / X [R, l]	44	130.295	0.163

Islands of reliability

y -> iR / X [J, N, b, d, g, k, m, n, p, t] uR			9	9	1.0	0.898
y -> iR / X kuR			8	8	1.0	0.887
y -> iR / X [2, 6, 9, @, E, H, O, R, S, Z, a, e, f, i, j, l, o, s, u, v, w, y, z, â, è, ê, ò, ô, û] kuR			7	7	1.0	0.872
y -> iR / X [2, 6, 9, @, E, H, O, R, a, e, i, j, l, o, u, w, y, â, è, ê, ò, ô, û] kuR			6	6	1.0	0.852
y -> iR / X [2, 6, 9, @, E, O, a, e, o, â, è, ê, ò, ô, û] kuR			5	5	1.0	0.825
y -> iR / X [6, @, R, S, Z, f, l, s, v, z] kuR			4	4	1.0	0.786
y -> iR / X [6, @, a, â, è, ò] kuR			4	4	1.0	0.786
y -> iR / X [6, @, O, o, â, ò, ô] kuR			4	4	1.0	0.786
y -> iR / X [6, @, a, è] kuR			3	3	1.0	0.718
y -> iR / X [6, @, O, o, ò, ô] kuR			3	3	1.0	0.718
y -> iR / X [6, @, â, ò] kuR			3	3	1.0	0.718
y -> iR / X [6, @, R, l] kuR			3	3	1.0	0.718
y -> iR / X [6, @, è] l			3	3	1.0	0.718
y -> iR / [a, â] kuR			2	2	1.0	0.57
y -> iR / X [O, o, â, ò, ô] kuR			2	2	1.0	0.57
y -> iR / [J, N, b, d, g, k, m, n, p, t] uR			2	2	1.0	0.57
y -> iR / X [R, f, l, s, v, z] kuR			2	2	1.0	0.57
y -> iR / [R, f, l, s, v, z] @kuR			2	2	1.0	0.57
y -> iR / X èl			2	2	1.0	0.57
y -> iR / X [R, l]			44	13	0.295	0.163
y -> iR / X l			21	4	0.19	0.123

ù

y -> iR / X [J, N, b, d, g, k, m, n, p, t] uR			9	9	1.0	0.898
y -> iR / X [6, @, è] l			3	3	1.0	0.718
y -> iR / X [R, l]			44	13	0.295	0.163

Islands of reliability

- La plus grande partie du treillis est redondante
- il existe un contexte plus grand de même indice de confiance

y ->	iR /	X [J, N, b, d, g, k, m, n, p, t] uR	9	91.0	0.898
y ->	iR /	X kuR	8	81.0	0.887
y ->	iR /	X [2, 6, 9, @, E, H, O, R, S, Z, a, e, f, i, j, l, o, s, u, v, w, y, z, â, ê, è, ò, ô, û] kuR	7	71.0	0.872
y ->	iR /	X [2, 6, 9, @, E, H, O, R, a, e, i, j, l, o, u, w, y, â, ê, è, ò, ô, û] kuR	6	61.0	0.852
y ->	iR /	X [2, 6, 9, @, E, O, a, e, o, â, ê, è, ò, ô, û] kuR	5	51.0	0.825
y ->	iR /	X [6, @, R, S, Z, f, l, s, v, z] kuR	4	41.0	0.786
y ->	iR /	X [6, @, a, â, è, ò] kuR	4	41.0	0.786
y ->	iR /	X [6, @, O, o, â, ò, ô] kuR	4	41.0	0.786
y ->	iR /	X [6, @, a, è] kuR	3	31.0	0.718
y ->	iR /	X [6, @, O, o, ò, ô] kuR	3	31.0	0.718
y ->	iR /	X [6, @, â, ò] kuR	3	31.0	0.718
y ->	iR /	X [6, @, R, l] kuR	3	31.0	0.718
y ->	iR /	X [6, @, è] l	3	31.0	0.718
y ->	iR /	[a, â] kuR	2	21.0	0.57
y ->	iR /	X [O, o, â, ò, ô] kuR	2	21.0	0.57
y ->	iR /	[J, N, b, d, g, k, m, n, p, t] uR	2	21.0	0.57
y ->	iR /	X [R, f, l, s, v, z] kuR	2	21.0	0.57
y ->	iR /	[R, f, l, s, v, z] @kuR	2	21.0	0.57
y ->	iR /	X èl	2	21.0	0.57
y ->	iR /	X [R, l]	44	130.295	0.163
y ->	iR /	X l	21	40.19	0.123

- Une partie du treillis contient des contextes plus fiables que la généralisation globale
- les **ilots de confiance** sont formés par des contextes restreints où l'analogie est plus fiable que dans le cas général

y ->	iR /	X [J, N, b, d, g, k, m, n, p, t] uR	9	91.0	0.898
y ->	iR /	X [6, @, è] l	3	31.0	0.718
y ->	iR /	X [R, l]	44	130.295	0.163

Islands of reliability

- La plus grande partie du treillis est redondante
- il existe un contexte plus grand de même indice de confiance

y -> iR / X [J, N, b, d, g, k, m, n, p, t] uR	9	91.0	0.898
y -> iR / X kuR	8	81.0	0.887
y -> iR / X [2, 6, 9, @, E, H, O, R, S, Z, a, e, f, i, j, l, o, s, u, v, w, y, z, â, è, ê, ò, ô, û] kuR	7	71.0	0.872
y -> iR / X [2, 6, 9, @, E, H, O, R, a, e, i, j, l, o, u, w, y, â, è, ê, ò, ô, û] kuR	6	61.0	0.852
y -> iR / X [2, 6, 9, @, E, O, a, e, o, â, è, ê, ò, ô, û] kuR	5	51.0	0.825
y -> iR / X [6, @, R, S, Z, f, l, s, v, z] kuR	4	41.0	0.786
y -> iR / X [6, @, a, â, è, ò] kuR	4	41.0	0.786
y -> iR / X [6, @, O, o, â, ò, ô] kuR	4	41.0	0.786
y -> iR / X [6, @, a, è] kuR	3	31.0	0.718
y -> iR / X [6, @, O, o, ò, ô] kuR	3	31.0	0.718
y -> iR / X [6, @, â, ò] kuR	3	31.0	0.718
y -> iR / X [6, @, R, l] kuR	3	31.0	0.718
y -> iR / X [6, @, è] l	3	31.0	0.718
y -> iR / [a, â] kuR	2	21.0	0.57
y -> iR / X [O, o, â, ò, ô] kuR	2	21.0	0.57
y -> iR / [J, N, b, d, g, k, m, n, p, t] uR	2	21.0	0.57
y -> iR / X [R, f, l, s, v, z] kuR	2	21.0	0.57
y -> iR / [R, f, l, s, v, z] @kuR	2	21.0	0.57
y -> iR / X èl	2	21.0	0.57
y -> iR / X [R, l]	44	130.295	0.163
y -> iR / X l	21	40.19	0.123

- Une partie du treillis contient des contextes plus fiables que la généralisation globale
- les **ilots de confiance** sont formés par des contextes restreints où l'analogie est plus fiable que dans le cas général

y -> iR / X [J, N, b, d, g, k, m, n, p, t] uR	9	91.0	0.898
y -> iR / X [6, @, è] l	3	31.0	0.718
y -> iR / X [R, l]	44	130.295	0.163

- Le MGL fonctionne bien localement pour une paire
 - système asymétrique ($1 \Rightarrow 2 \neq 2 \Rightarrow 1$)
 - modèle les ilots de confiance (spring/splang/splung en anglais)
 - rend compte de la défectivité par manque analogique (distraire)
- Difficultés
 - pas de traitement au delà d'une paire de cases
 - pas de notion de paradigme flexionnel
 - rien pour les deux autres types de défectivité
 - héritée : clore \Rightarrow ?closons
 - hésitation : nasal \Rightarrow ?nasals/?nasaux
 - lexique complet en ligne

Bonami & Boyé 2012

- Calculer l'ensemble des correspondances entre cases
- Système similaire au MGL pour calculer les **analogies**
 - pour chaque paire, on calcule la généralisation minimale couvrant l'ancienne généralisation et la paire ajoutée
 - **un seul contexte** par analogie au final
- Les indices de confiances sont remplacés par l'entropie des différentes **classes de distribution**
 - une classe de distribution est un ensemble d'analogie qui peuvent s'appliquer à une forme
 - pour une classe donnée, on compte quelle est la répartition des formes réelles entre les analogies et on calcule l'entropie correspondante
- Calcul pour toutes les paires de cases du paradigme complet
 - distillation et relations concrètes et motivées

Principe de classification par ensembles

- Pour une paire de formes du paradigme, chaque forme d'entrée est associée à l'ensemble d'alternances susceptibles de lui être appliquées
 - par exemple, pour l'imparfait indicatif 1 et l'indicatif présent 6 :

LEXEME	IPF.1 (→ PRS.6)	1	2	3	...	Ensemble
		$\varepsilon \rightarrow \emptyset$	$\gamma\nu\varepsilon \rightarrow \text{w}\alpha\nu$	$\varepsilon\nu\varepsilon \rightarrow \varepsilon\eta$...	
PASSER	$\text{pas}\varepsilon$	+				{1, ...}
BOIRE	$\text{by}\nu\varepsilon$	+	+			{1, 2, ...}
FINIR	$\text{finis}\varepsilon$	+				{1, ...}
MENER	$\text{m}\varepsilon\nu\varepsilon$	+		+		{1, 3, ...}
SORTIR	$\text{s}\omega\text{rt}\varepsilon$	+				{1, ...}
PRENDRE	$\text{pr}\varepsilon\nu\varepsilon$	+		+		{1, 3, ...}
...	...					{..., ...}

Principe de classification par ensembles

- Pour une paire de formes du paradigme, chaque forme d'entrée est associée à l'ensemble d'alternances susceptibles de lui être appliquées
 - par exemple, pour l'imparfait indicatif 1 et l'indicatif présent 6 :

50 verbes	{1, 3, 4}	1	3	4	Nombre
EXEMPLE	IPF.1 → PRS.6	$\varepsilon \rightarrow \emptyset$	$\text{\textcircled{a}}\text{n}\varepsilon \rightarrow \text{\textcircled{e}}\text{n}$	$\text{\textcircled{a}}\text{n}\varepsilon \rightarrow \text{\textcircled{j}}\text{n}$	
PRENDRE	$\text{pr}\text{\textcircled{a}}\text{n}\varepsilon$	$\text{\textcircled{p}}\text{r}\text{\textcircled{a}}\text{n}$	$\text{\textcircled{p}}\text{r}\text{\textcircled{e}}\text{n}$	$\text{\textcircled{p}}\text{r}\text{\textcircled{j}}\text{n}$	24
VENIR	$\text{v}\text{\textcircled{a}}\text{n}\varepsilon$	$\text{\textcircled{v}}\text{\textcircled{a}}\text{n}$	$\text{\textcircled{v}}\text{\textcircled{e}}\text{n}$	$\text{\textcircled{v}}\text{\textcircled{j}}\text{n}$	26
Entropie pour la classe {1,3,4}					0,999

$$H[X] = - \sum_{x \in X} p(x) \log_2 p(x)$$

Principe de classification par ensembles

- Pour une paire de formes du paradigme, chaque forme d'entrée est associée à l'ensemble d'alternances susceptibles de lui être appliquées
 - par exemple, pour l'imparfait indicatif 1 et l'indicatif présent 6 :

Ensemble	Nombre	Proportion	Entropie	Contribution
{1, 3, 4}	50	0,78 %	0,999	0,008
{1, 5, 6}	438	6,80 %	1,527	0,104
...
Entropie globale pour IPF.1 → PRS.6				0,155

Une analyse transparente ?

- Système de classification à deux niveaux
 - l'entropie générale classe les cases les unes par rapport aux autres et donne une distillation motivée
 - l'entropie locale donne des îlots de confiance par recoupement des contextes
- La description des correspondances entre formes est remplacée par un calcul d'entropie qui caractérise les interprédictibilités orientées entre cases
 - un calcul motivé sur la prédictibilité
 - les prédictions disparaissent en pratique derrière l'entropie

Une analyse transparente ?

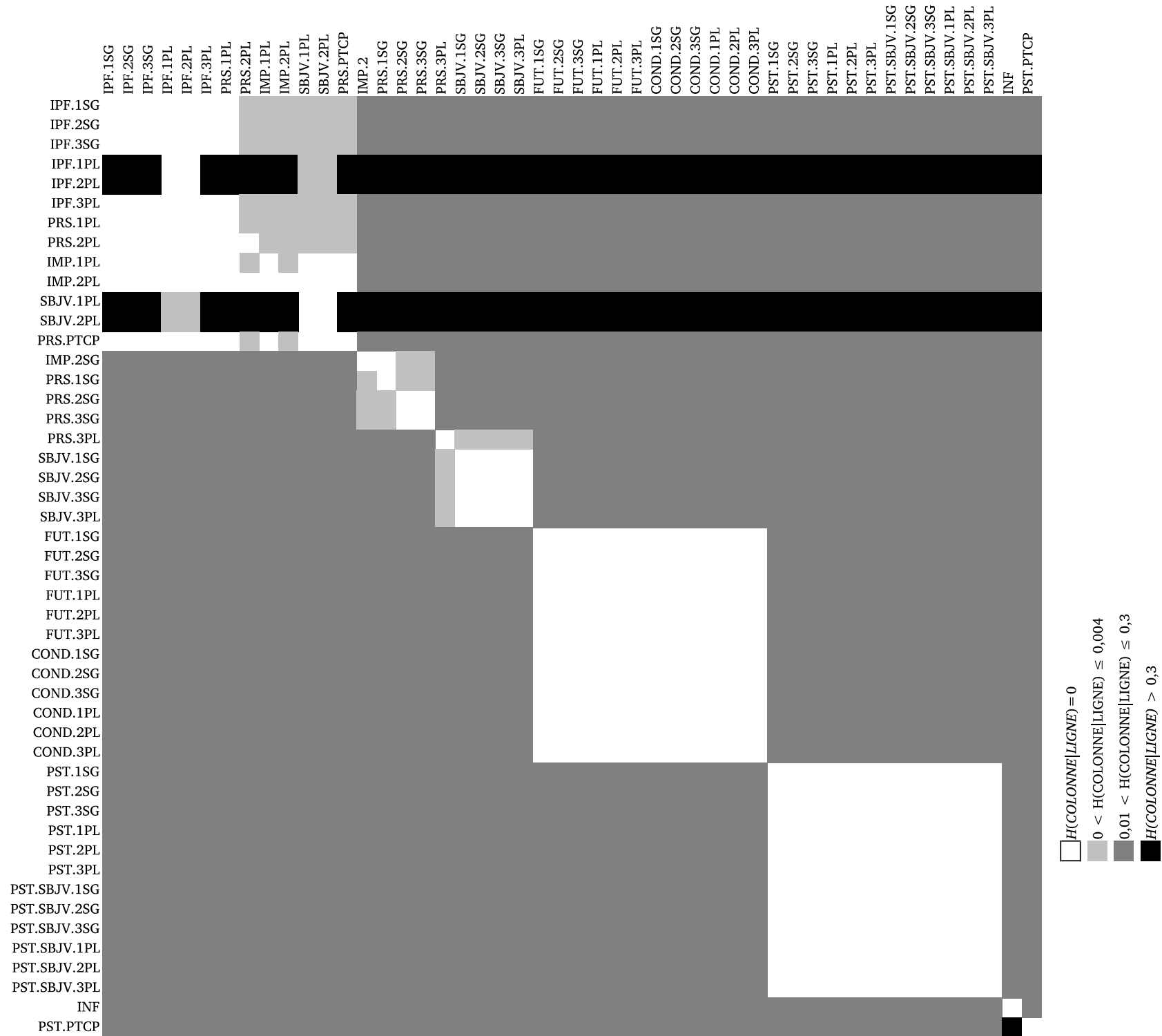
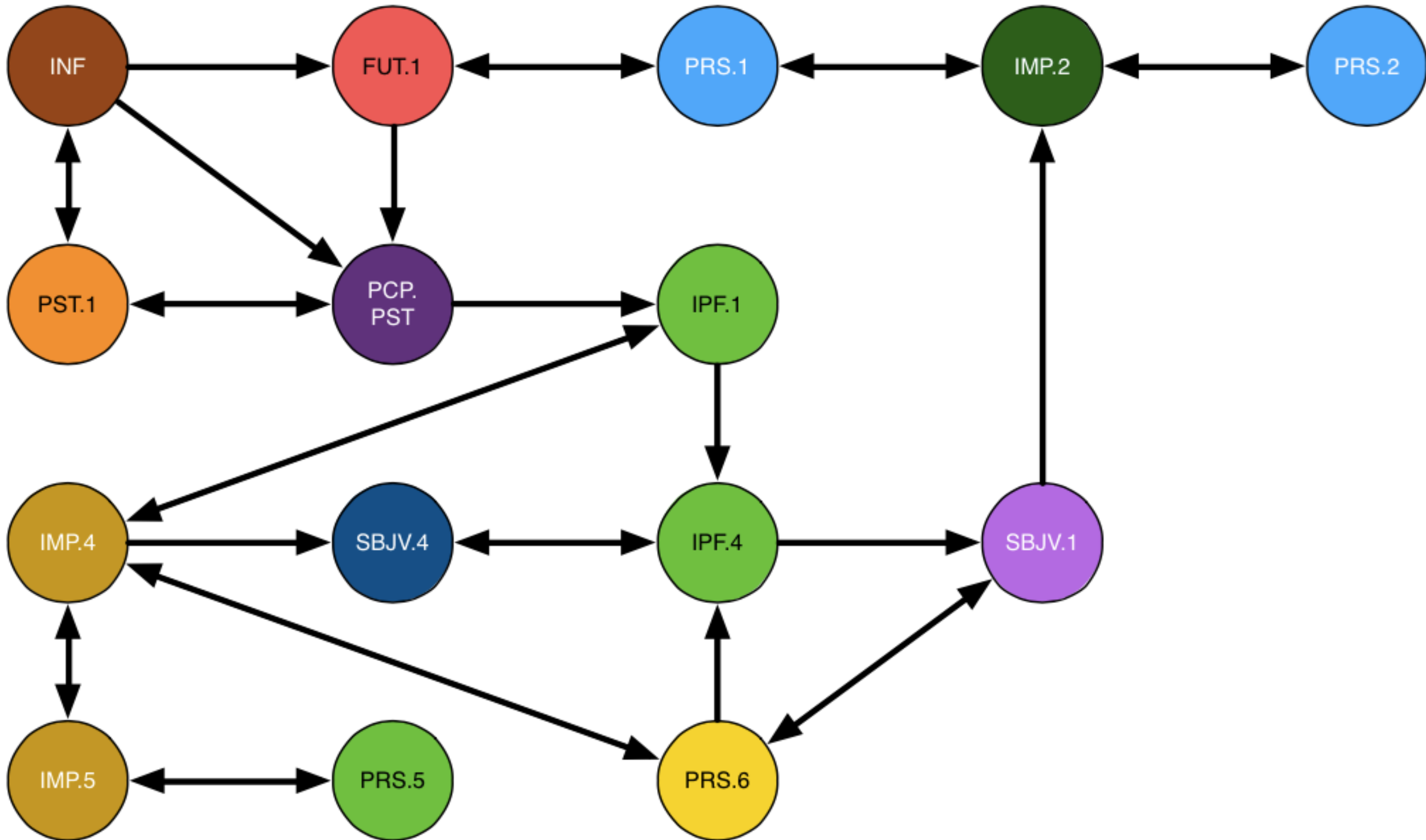


Figure 1 — Entropie conditionnelle entre les 48 formes des 6440 verbes non-défectifs de BDLX¹³

Boyé & Kupść 2014

- Une analyse en espaces thématiques basée sur BoBo12
 - reprise intégrale de la distillation
 - génération automatique d'un graphe de régularité
 - ajout incrémental de liens à partir des entropies minimales
 - formation d'un graphe connecté complet
- Pas de relations explicites entre les cases ou les formes
 - entièrement basé sur les entropies globales
 - des prédictions sur la distribution de la régularité mais pas sur les formes

Boyé & Kupść 2014



Paradigmes à la Gaume

- Objectif
 - trouver des relations régulières motivées
 - construire un réseau orienté entre formes
 - prédire l'ensemble du paradigme à partir d'une forme de plusieurs formes
- Une recette avec trois ingrédients
 - Analyser le lexique connu et produire les distributions pour toutes les classes d'analogies
 - Générer un bi-paradigme à partir des formes connues d'un lexème
 - Récolter les cliques de formes obtenues

1.a) Analogies à la Albright

- Pour toutes les paires de cases, on génère toutes les analogies repérées avec le contexte minimal couvrant l'ensemble des cas observés.
- le moteur utilisé est le même MGL simplifié que BoBo12, il génère des règles SPE :
 - $A \rightarrow B / C _ D$
- tout autre moteur permettant d'établir des analogies contextuelles minimales conviendrait

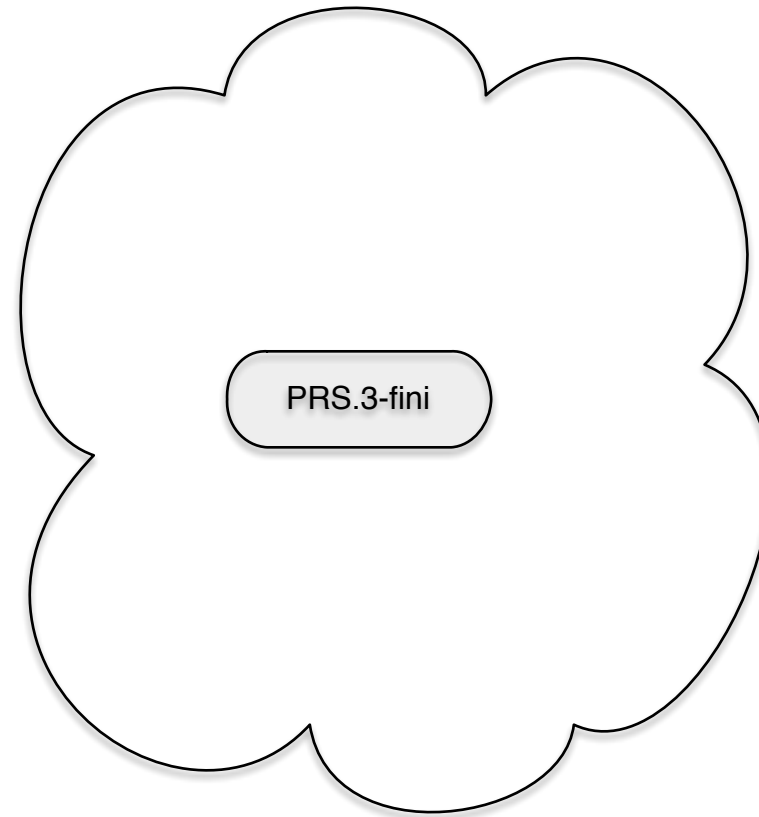
1.b) Distribution de BoBo

- L'algorithme de classement est également repris à BoBo12
- chaque forme de départ appartient à une classe d'analogies potentielles issues de l'analyse du MGL
- la distribution des formes de départ entre les analogies potentielles donne la répartition pour la propagation à l'étape suivante

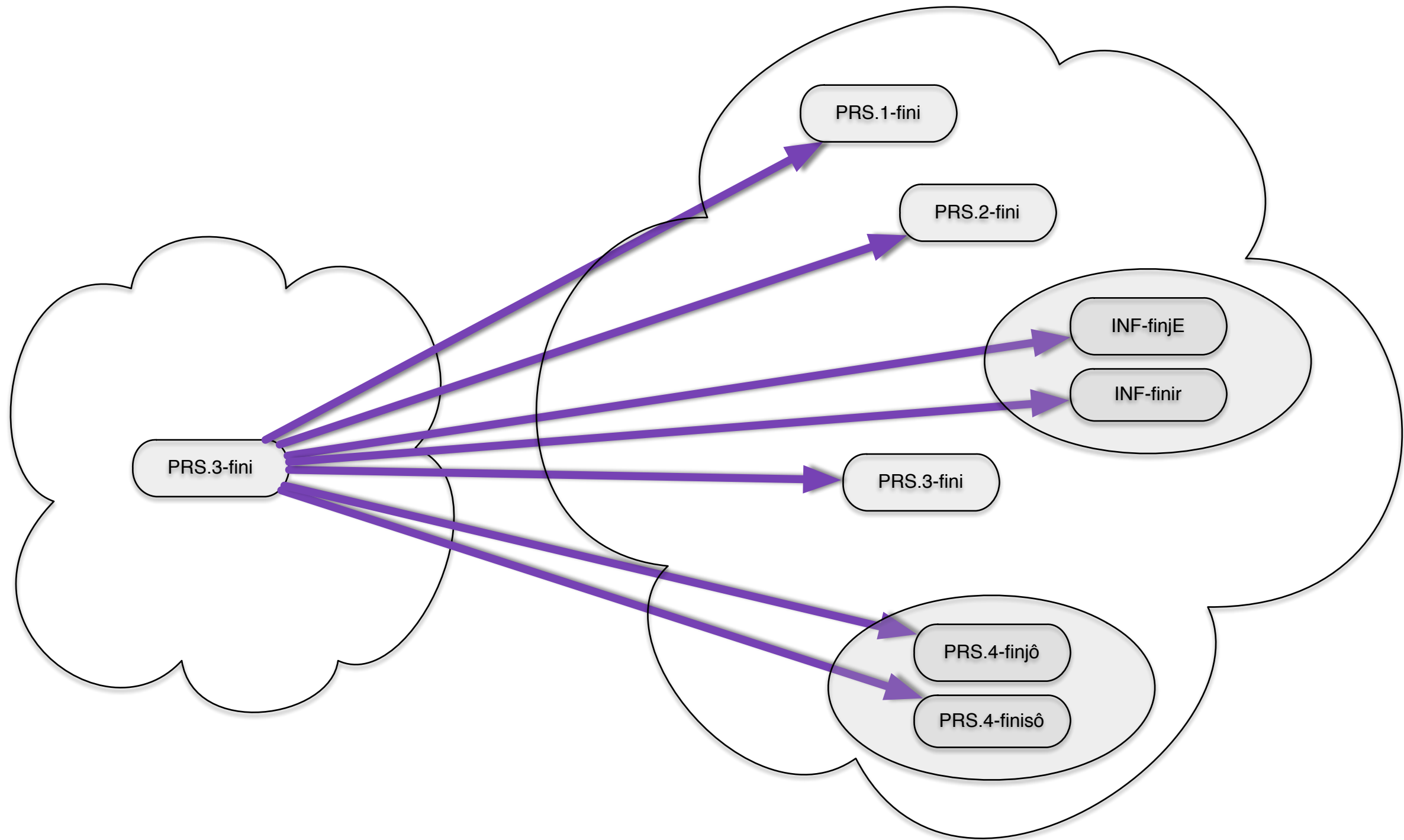
2) Propagation sur un *bigraphe*

- Dans cette phase, on utilise une représentation de départ et un bigraphe avec deux propagations
 - une représentation lexicale contenant les connaissances initiales sur le contenu du paradigme
 - deux propagations
 - une première permet d'obtenir des candidats dans toutes les cases du paradigme
 - une seconde permet de récupérer les candidats qui sont le plus soutenus par l'ensemble
- à chaque propagation, les formes de chaque cellule se propagent via leurs analogies potentielles vers toutes les cases en proportion de la distribution BoBo observée précédemment

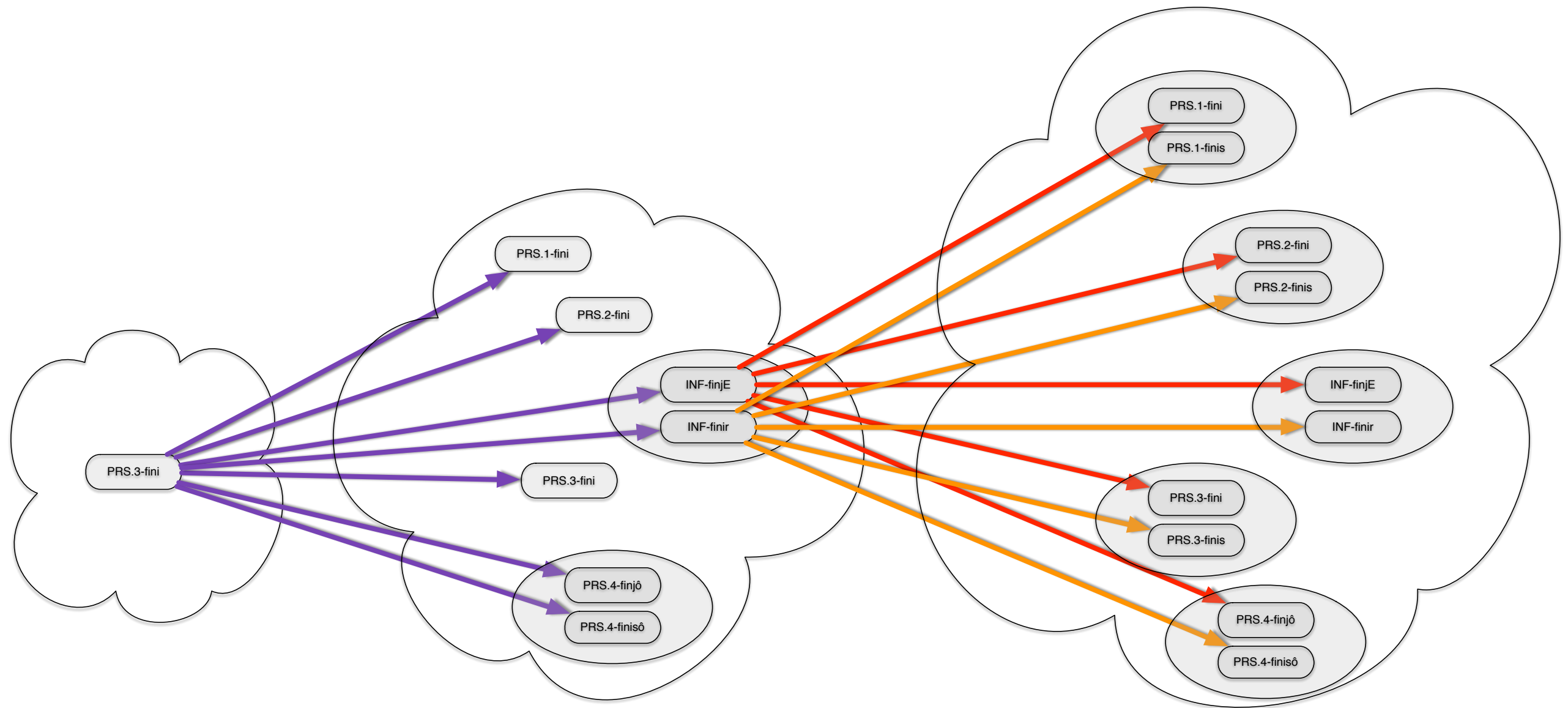
Semer des formes



Propager deux fois



Propager deux fois



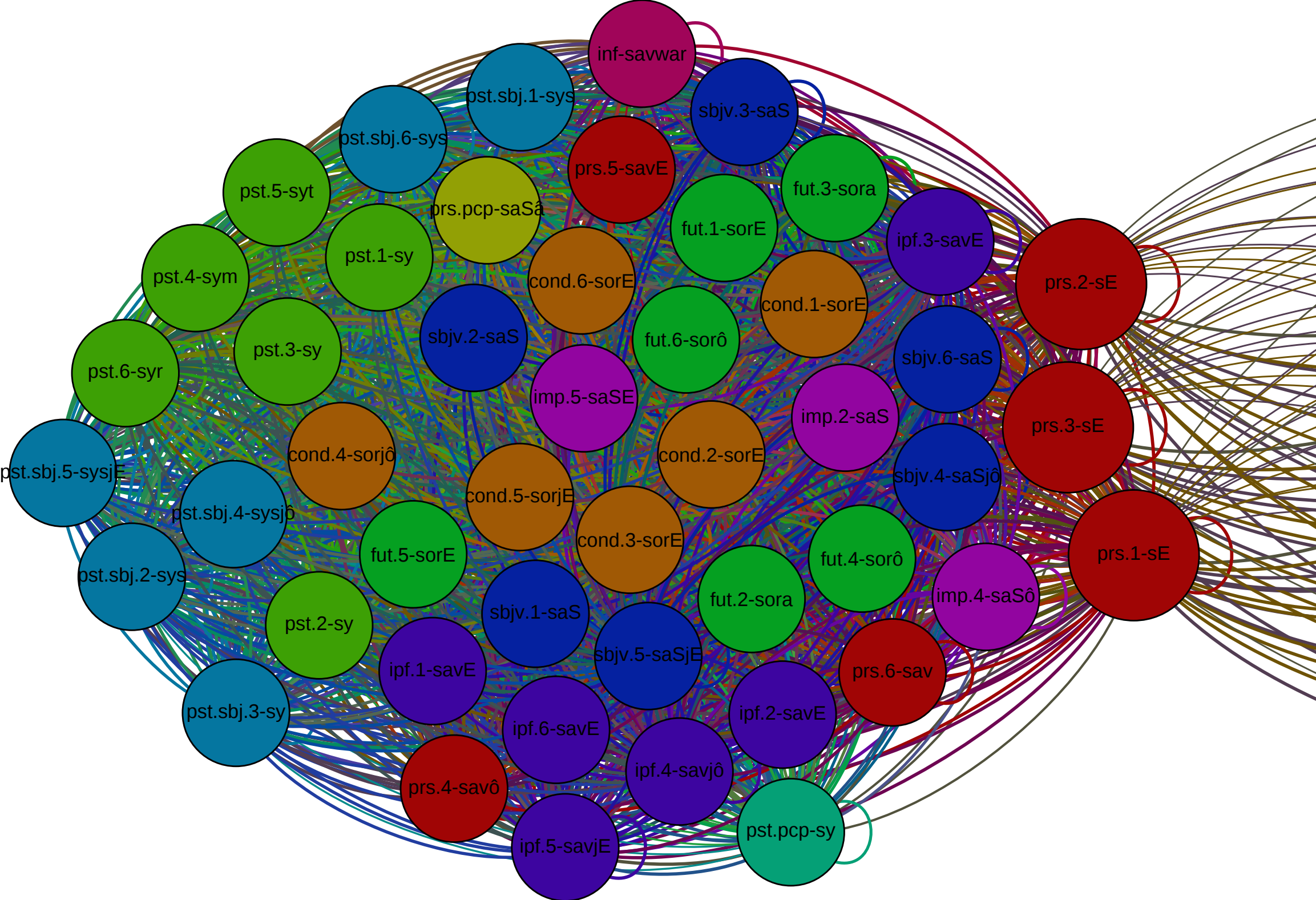
Distributions

- Deux niveaux de pondération :
 - Niveau Paradigme :
 - une distribution pour la contribution des cases
 - l'infinitif est plus important que le passé simple
 - Niveau Formes :
 - une distribution pour les différentes formes à l'intérieur d'une case
 - au départ, en fonction des proportions dans la surabondance
 - à l'arrivée, en fonction de la propagation

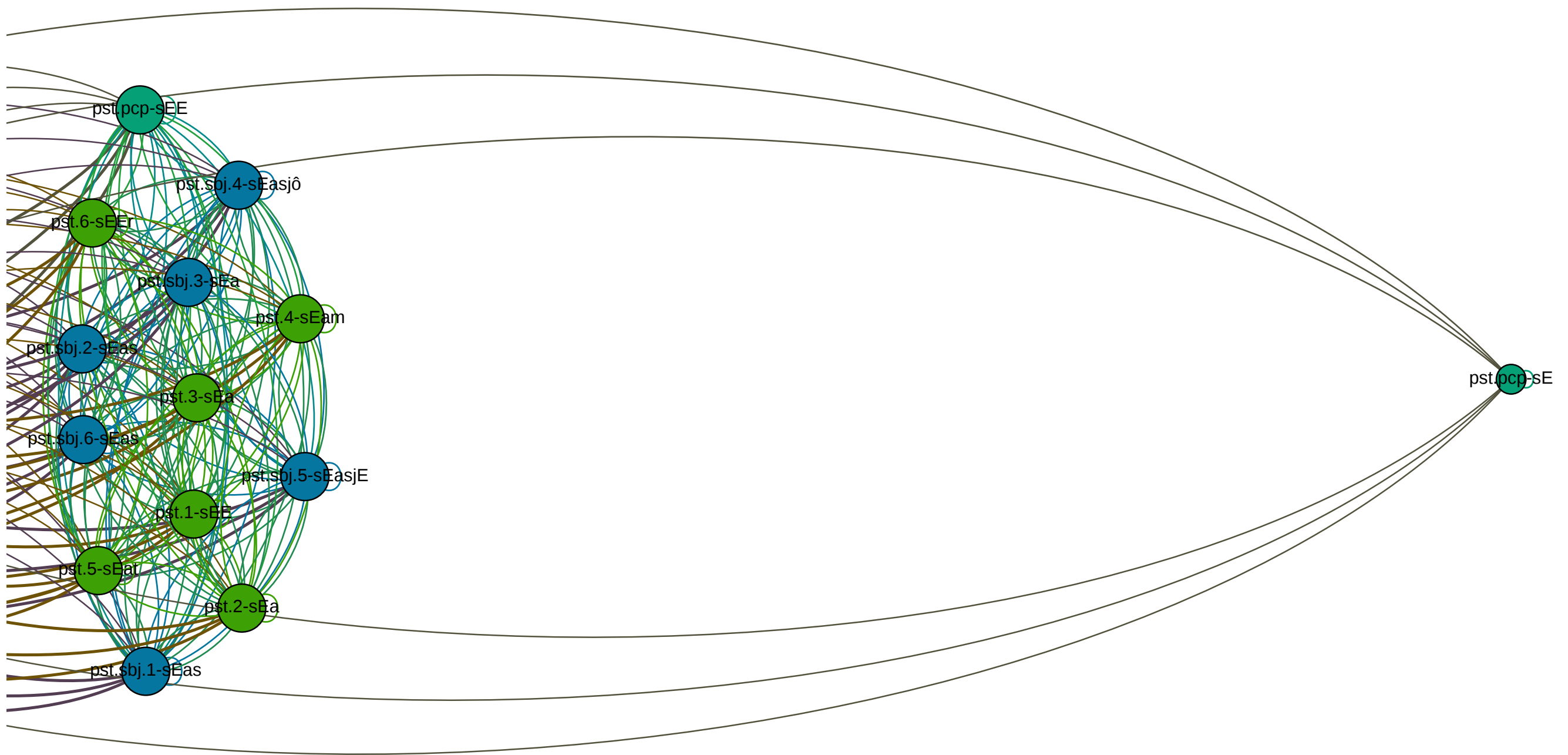
3) Cliques & paradigmes

- Dans l'idéal, un paradigme flexionnel correspond à un ensemble de formes qui se répondent en chœur
 - dans un graphe simple, les formes d'un paradigme constituent une clique
- dans le *bigraphe* obtenu par propagation, on peut réunir les deux derniers stades pour extraire des cliques qui correspondent aux paradigmes potentiels pour les formes introduites dans le premier
 - chaque nœud est associé à un coefficient d'accord avec les autres formes du graphe
 - on retrouve des organisations de petits mondes, avec des formes qui appartiennent à plusieurs cliques

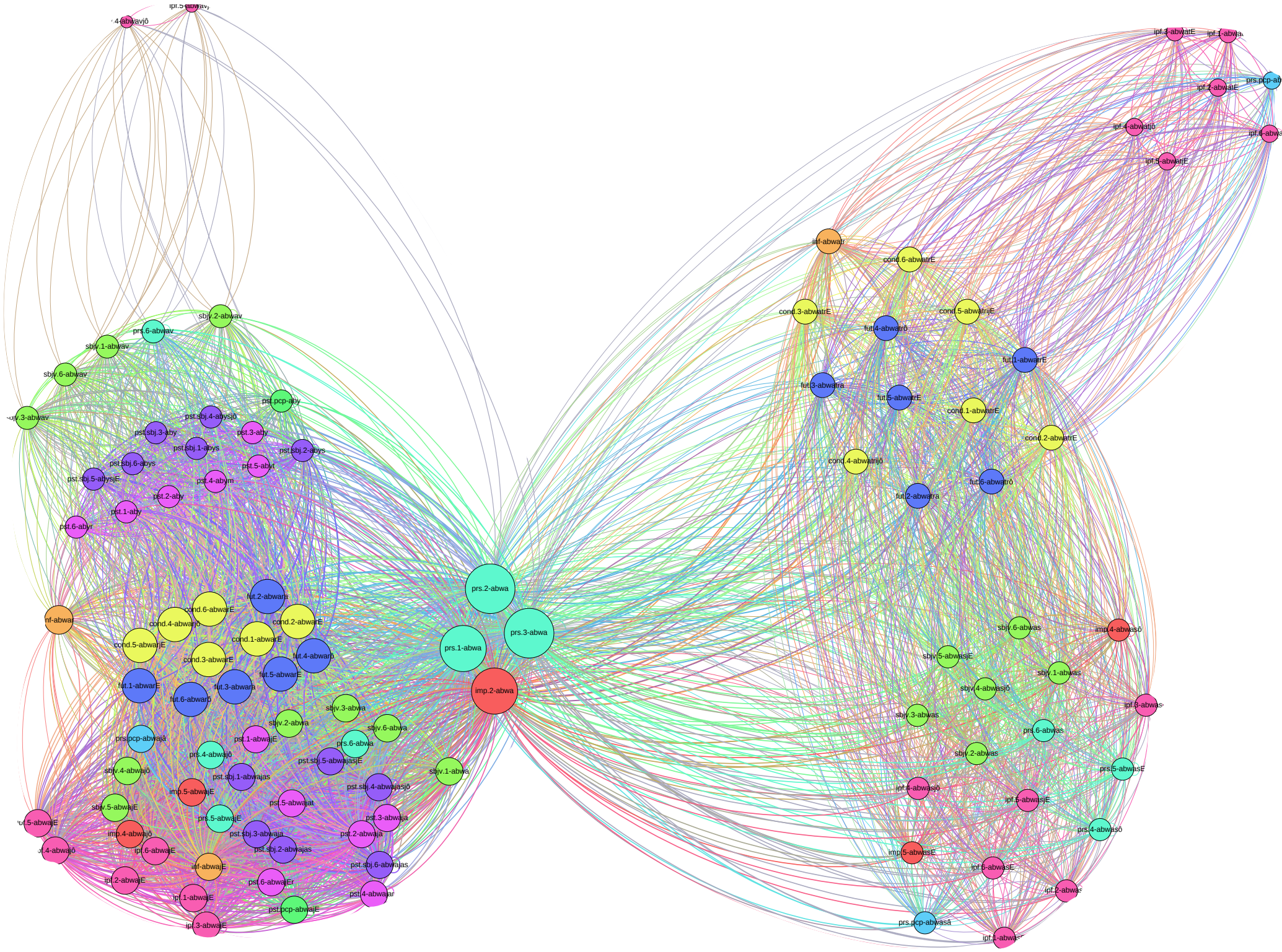
Savoir-IMP.2



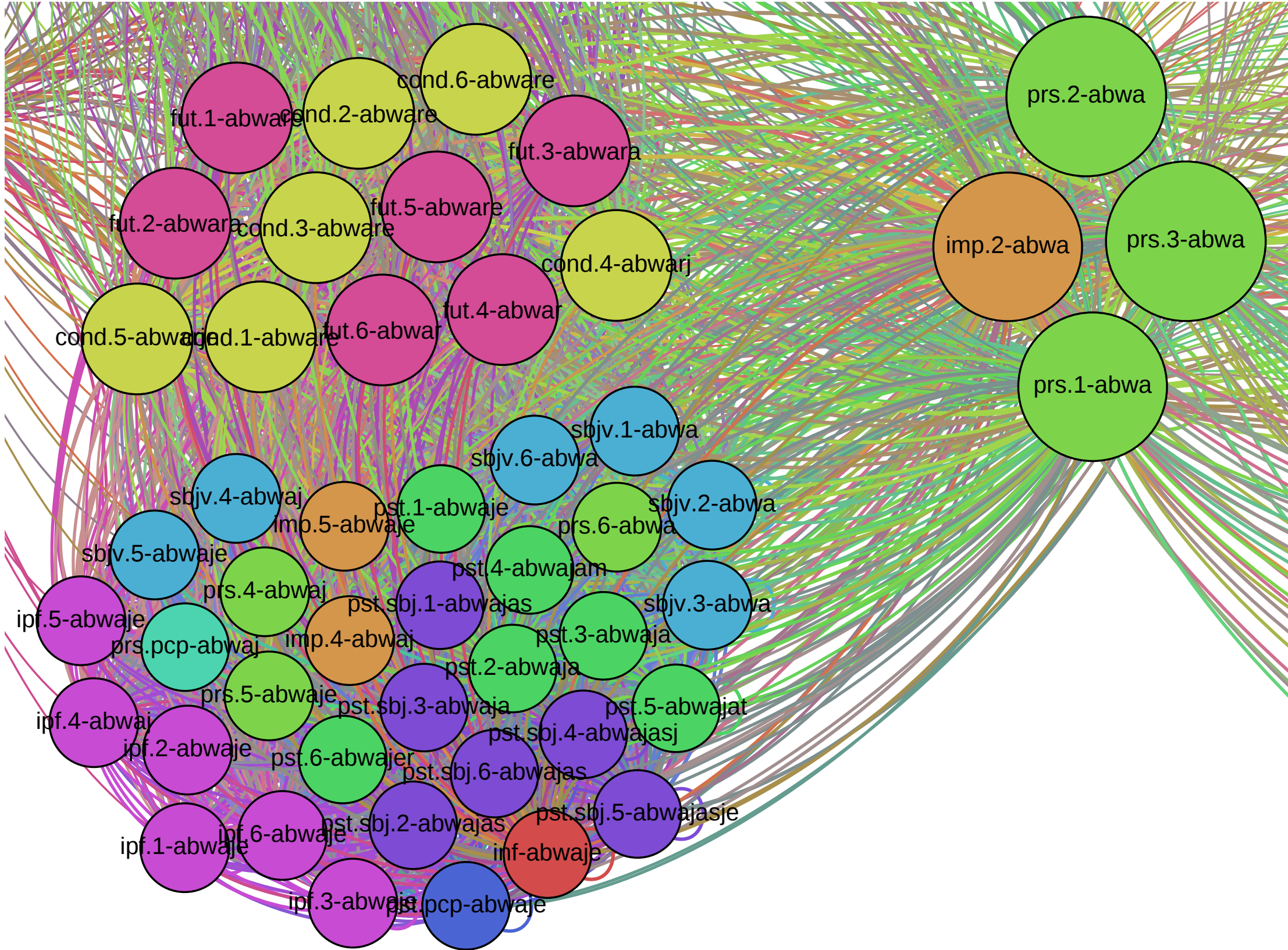
Savoir-IMP.2



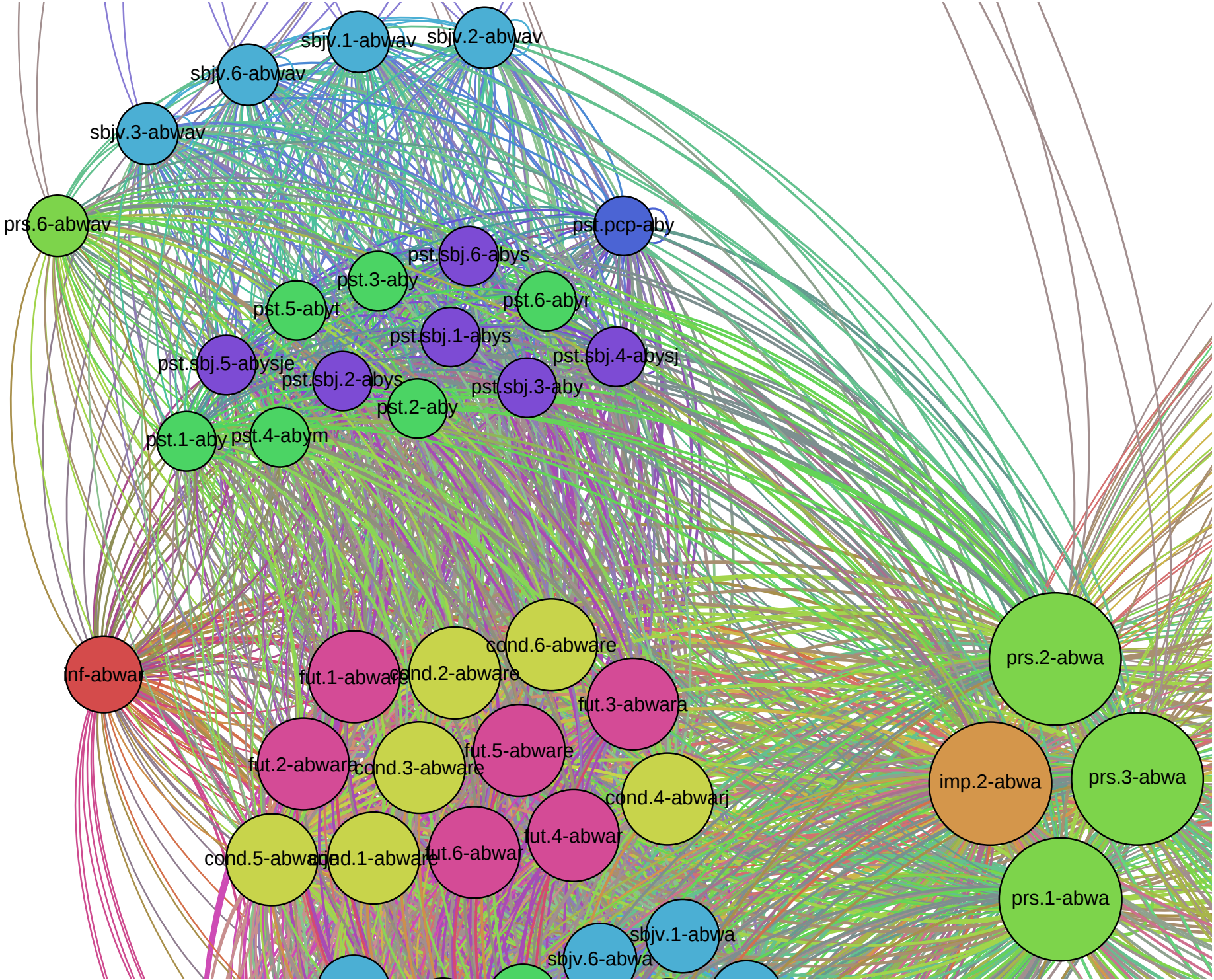
Aboyer-PRS.3



Aboyer-PRS.3



Aboyer-PRS.3



Aboyer-PRS.3

aboyer-PRS3 48 0.807708333333

temps	1SG	2SG	3SG	1PL	2PL	3PL
present	abwa	abwa	abwa	abwajǽ	abwaje	abwa
imperfective	abwaje	abwaje	abwaje	abwajǽ	abwaje	abwaje
past	abwaje	abwaja	abwaja	abwajam	abwajat	abwajer
future	abware	abwara	abwara	abwarǽ	abware	abwarǽ
present subj.	abwa	abwa	abwa	abwajǽ	abwaje	abwa
imperfective subj.	abwajas	abwajas	abwaja	abwajasjǽ	abwajasje	abwajas
conditional	abware	abware	abware	abwarjǽ	abwarje	abware
imperative		abwa		abwajǽ	abwaje	

inf	abwaje
prs.pcp	abwajǽ
pst.pcp	abwaje

Aboyer-PRS.3

aboyer-PRS3 35 0.541714285714

temps	1SG	2SG	3SG	1PL	2PL	3PL
present	abwa	abwa	abwa			abwav
imperfective						
past	aby	aby	aby	abym	abyt	abyr
future	abware	abwara	abwara	abwarǽ	abware	abwarǽ
present subj.	abwav	abwav	abwav			abwav
imperfective subj.	abys	abys	aby	abysjǽ	abysje	abys
conditional	abware	abware	abware	abwarjǽ	abwarje	abware
imperative		abwa				
		inf	abwar			
		pst.pcp	aby			

Évolutions

- Variation explicite
 - passer d'un lexique de paradigmes à un paradigme de formes
 - calculer les analogies entre formes concurrentes
- Défectivité explicite
 - ajouter des formes explicitement vides \emptyset
 - calculer les analogies contextuelles avec les formes \emptyset en sortie
- Gestion des informations initiales
 - filtrer les cliques d'après la représentation originale complète
- Utiliser un jeu d'essai standard
 - si possible écologique
 - morceaux de paradigmes correspondant aux fréquences