

Intégration de R dans le Trameur

Ce document décrit la mise à jour mise en œuvre à partir de la **version 11.00** du **Trameur** permettant de lancer des modules R (AFC, ACP, CAH) via la librairie *factoMineR* (http://factominer.free.fr/index_fr.html)

Pour illustrer ces fonctionnalités, on utilise ici la base textométrique intitulée « Base des états finaux du corpus BROUILLONS (ANR ECRITURES¹) » :

Cette base rassemble 19 « états finaux » issus de 19 dossiers² du corpus BROUILLONS (ANR ECRITURES). Chaque item de la *Trame* est associé à 3 niveaux d'annotation : forme, lemme, catégorie. On présente ci-dessous un extrait de la *Trame* de cette base via une concordance :

```
-----PARTIE(dossier=ravet)-----
un lycée professionnel standard mais nous craignons que adèle ne
compte tenu de ce que nous connaissons
relation avec les garçons que nous trouvons
compte tenu de ce que nous observons
un autre garçon accueilli) nous
. $ a ce jour nous n'avons aucu
vacances et elle a pu nous dire que sa grand-
madame richard, assistante familiale, nous a informé que adèle
de sa famille. $ nous avons reçu adèle pour
de la dap). $ nous lui avons rappelé qu'
famille non préparé. $ nous allons proposer à la
fragilités de adèle, il nous semble qu'un accueil
l'institut camille lamartine, nous avons organisé des rencontres
adèle a été demandeuse .nous allons en organiser une
foyer des oiseaux. $ nous sommes actuellement en contact
elle. $ toutefois, nous pensons qu'il y
```

```
Position:<482>
Forme:<nous>|Freq:317
Lemme:<Nous_Sujet_Sentiment>|Freq:13
Cat:<Nous_Sujet_Sentiment>|Freq:13
```

Les items associés à la forme (annotation n°1) « nous » ont été annotés via le **Trameur** : leur catégorie (annotation n°3) a été modifiée pour prendre les valeurs suivantes :

```
Nous_Faire
Nous_OBJ
Nous_Sujet
Nous_Sujet_Constat
Nous_Sujet_Dire
Nous_Sujet_Dire_Faire
Nous_Sujet_Dire_Preco
Nous_Sujet_Faire
Nous_Sujet_Pensee
Nous_Sujet_Pensee_Faire
Nous_Sujet_Preco
Nous_Sujet_Savoir
Nous_Sujet_Sentiment
```

¹ <http://www.univ-paris3.fr/anr-ecritures-96530.kjsp?RH=1295620557102>

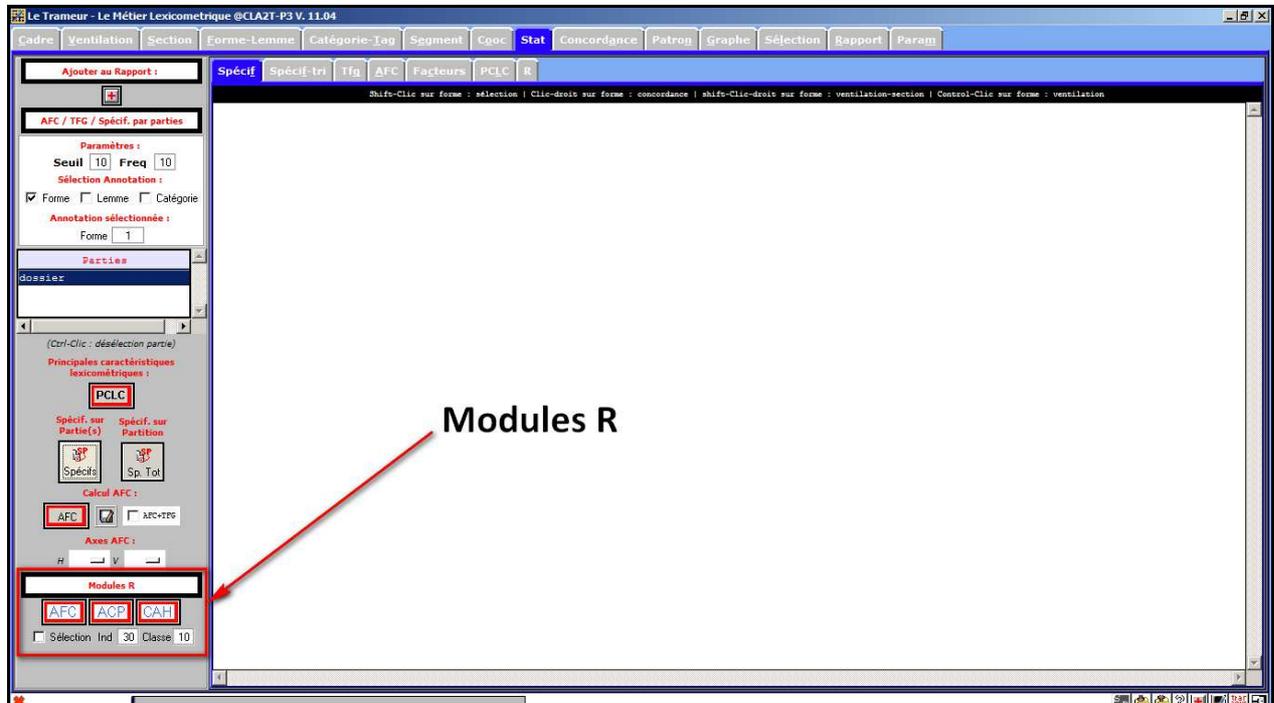
² Chaque dossier est initialement constitué d'un ensemble chronologique de textes allant du brouillon initial à la version finale, on ne retient ici que ce dernier état.

Sommaire

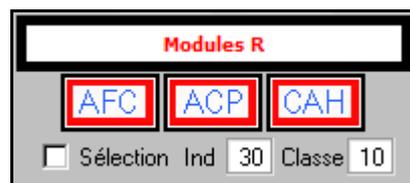
1. Mise en œuvre des modules R dans le Trameur.....	3
2. Illustration sur la base « Etats finaux BROUILLONS ».....	4
Calcul sur base complète.....	4
Calcul via le Gestionnaire de sélection.....	7
Calcul de l'AFC.....	9
Calcul de la CAH.....	10

1. Mise en œuvre des modules R dans le Trameur

Les modules R sont disponibles dans l'onglet **STAT** qui a désormais l'allure suivante :



Le fichier de données utilisés par R est généré dynamiquement dans le dossier contenant la base de travail. Les 3 modules produisent des sorties PDF (rassemblant les graphiques produits par R) qui sont stockées dans le sous-dossier *export* du dossier contenant la base de travail.



Par défaut, les 3 modules utilisent l'annotation sélectionnée dans le paramétrage de l'onglet **STAT** (en haut à gauche) et les items de fréquence supérieure à celle du paramètre *freq* (même zone de paramétrage, en haut à gauche).

Le paramètre *Ind* permet de spécifier le nombre d'item (forme, lemme, catégorie...) qui seront projetés sur les graphiques de l'AFC ou de l'ACP.

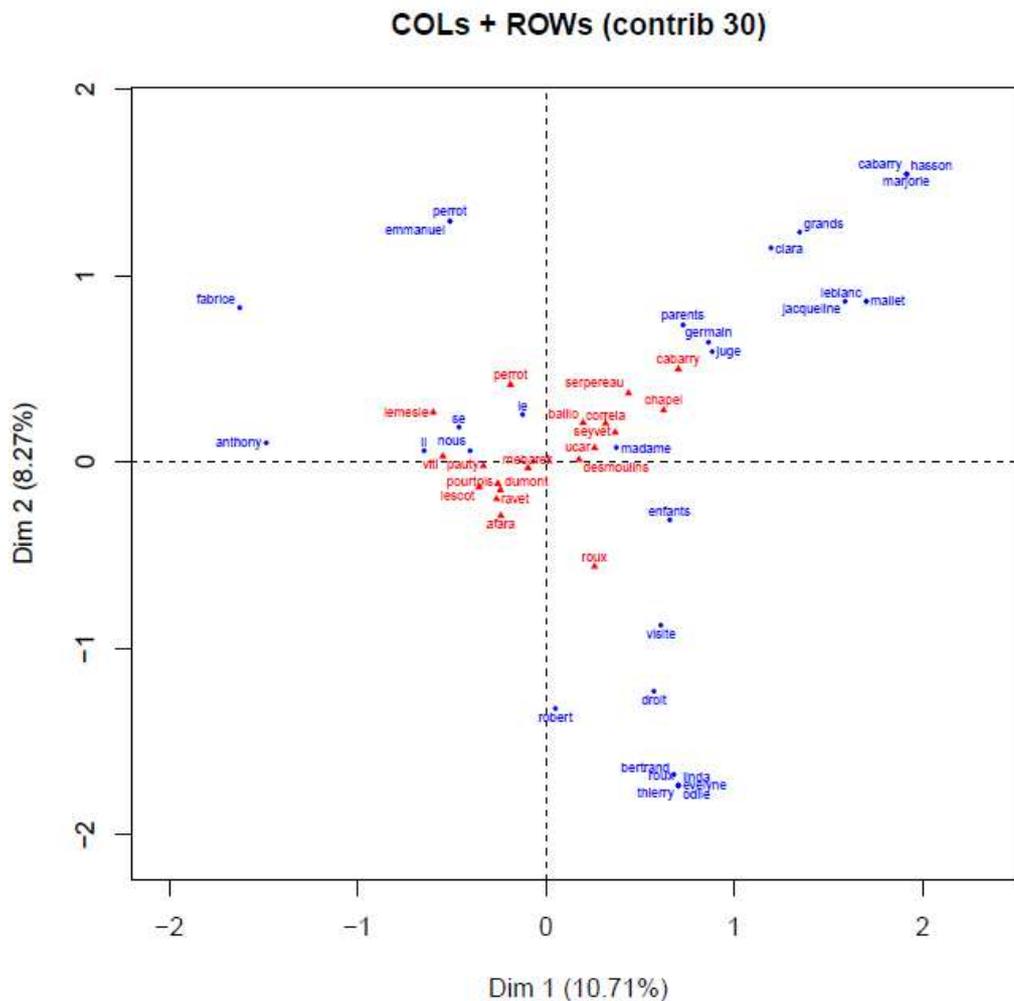
Le paramètre *Classe* permet de spécifier le nombre de classes à mettre au jour pour le module CAH.

Le paramètre *Sélection* permet de modifier le tableau de données en entrée de R en ne gardant que les items préalablement sélectionnés via le *Gestionnaire de Sélection*.

2. Illustration sur la base « Etats finaux BROUILLONS »

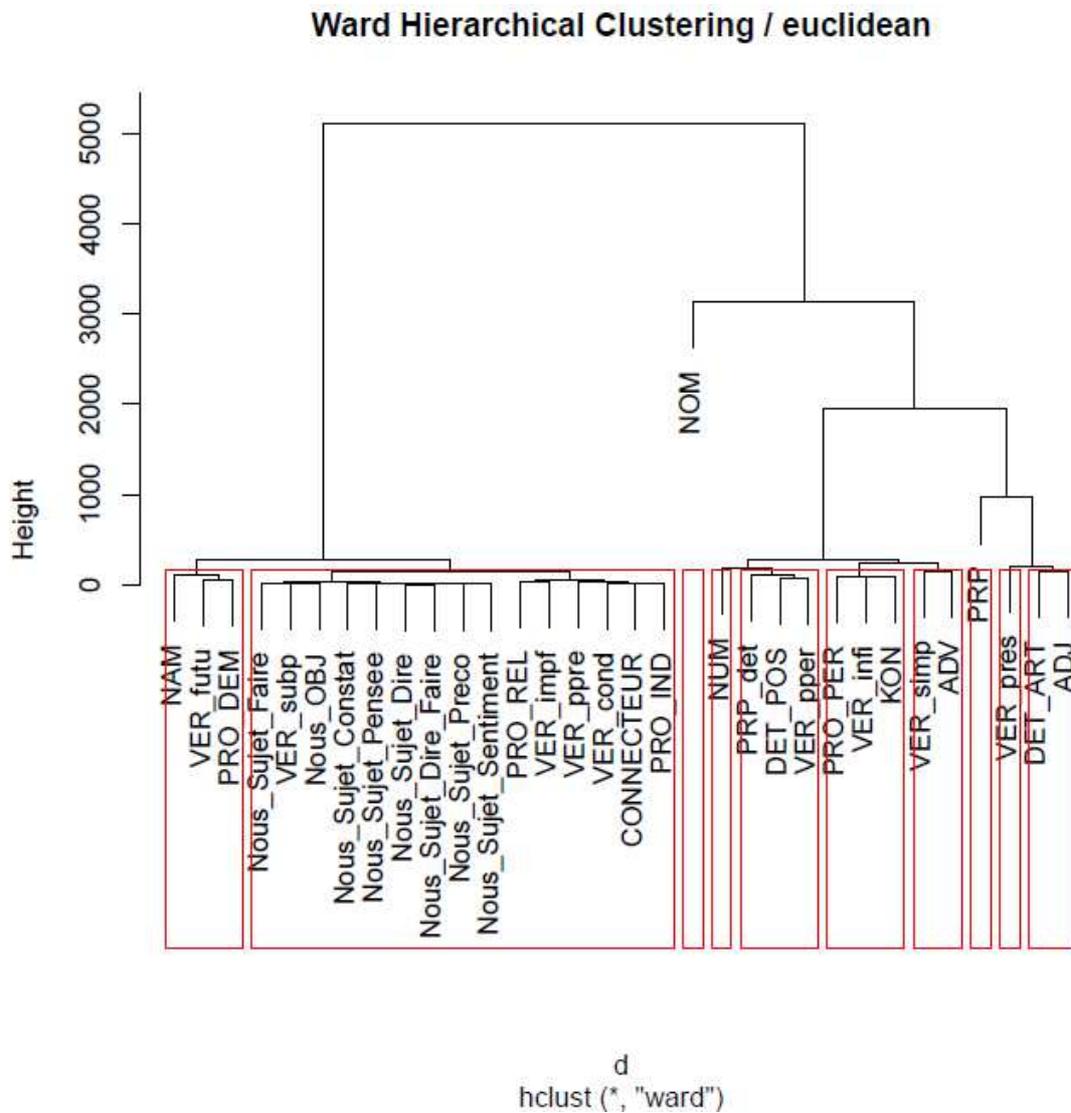
Calcul sur base complète

La base étant chargée, on active le calcul de l'AFC via R (paramétrage par défaut) via le bouton idoine, on donne à voir ci-dessous un extrait de la sortie PDF produite :



```
----- AFC OK -----  
Fichier de résultats (PDF) : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export/AFC1375361223.pdf  
Fichier de résultats (TXT) : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export/AFC-Result-1375361223.txt  
Fichier de données : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/dataForR1375361223.txt  
Répertoire de travail : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/export  
Répertoire de logs : C:/Users/UTILIS~1/Dropbox/ANR-EC~1/WIP/BASE-A~1/logR  
Répertoire de R : C:/SergeFleury/SF-WORK/07-DEVEL/ANR-textometrie/mkGraphTrameCadre/R/bin/x64/R.exe
```


De même pour le module CAH en modifiant ici le calcul pour travailler sur l'annotation n°3 (catégorie) :



Calcul via le Gestionnaire de sélection

On souhaite désormais que les calculs soient faits uniquement sur les items de type VERBE et ceux correspondant aux différents types de « nous ». Pour repérer ces items, on commence par les sélectionner via le *Gestionnaire de Sélection* : recherche en 2 temps, les « nous » puis les verbes (cf figures suivante) :

Recherche des différents types de "nous"
Les items étant mis au jour (1), il suffit de les ajouter au Gestionnaire de Sélection (2), (3)

Gestionnaire de Sélections
Le gestionnaire de sélections permet de rechercher, sélectionner des items sur la trame (ou des ensembles d'items). On peut aussi sélectionner ou modifier des éléments du cadre (parties ou régions).
- L'onglet "Liste Sélections" donne accès aux items sélectionnés dans le gestionnaire et aux opérations sur ces items.
- L'onglet "Recherche Forme" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Recherche Annotation" permet de rechercher puis de sélectionner des items de la Trame en filtrant sur tous les niveaux d'annotation.
- L'onglet "Recherche Segment" permet de rechercher puis de sélectionner des segments sur la Trame en filtrant sur les formes (annotation n°1).
- L'onglet "Sélections sur Cadre" permet de sélectionner des items sur ces sélections : facilités, cooccurrences,

Recherche
Annotations (RegExp):
Forme: VERB
Catégorie: VERB
Chercher

Position	Annotation 1	Annotation 2	Annotation 3	Type
10	adèle			VER_pres
20	adèle			VER_pres
24	bénéficié	bénéficié		VER_pper
26	d	d		VER_simp
30	suiivi	suivre		VER_pper
36	d			VER_simp
56	limoges	limoger		VER_pres
65	bénéficié	bénéficié		VER_pper
67	d	d		VER_simp
73	effectuée	effectuer		VER_pper
89	d	d		VER_simp
104	l	l		VER_simp
112	permis	permettre		VER_simp
116	progresser	progresser		VER_infi
131	n			VER_simp
151	adèle	adèle		VER_pres
159	adèle	adèle		VER_pres
161	est	être		VER_pres
165	3eme	3eme		VER_pres
168	segpa	segpa		VER_simp
180	adèle	adèle		VER_pres
184	situe	situer		VER_pres
190	élève	élever		VER_pres
192	ayant	avoir		VER_ppre
198	place	placer		VER_pres
202	segpa	segpa		VER_simp
237	liée	lier		VER_pper
256	adèle	adèle		VER_pres
260	effectué	effectuer		VER_pper
276	ont	avoir		VER_pres
279	permis	permettre		VER_simp
285	prendre	prendre		VER_infi
311	confronter	confronter		VER_infi
324	bus	boire		VER_pper
332	représenté	représenter		VER_pper

Le Trameur : documentation complémentaire

Au final, le *Gestionnaire de Sélection* contient tous les items recherchés, on prend soin de tous les sélectionner :

Position	Annotation 1	Annotation 2	Annotation 3	Type
100820	nous	Nous_Sujet_Prec	Nous_Sujet_Prec	multi-regex
101001	nous	Nous_Sujet_Prec	Nous_Sujet_Prec	multi-regex
520	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
1021	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
7097	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
13175	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
13492	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
22653	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
23060	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
40475	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
58124	nous	Nous_Sujet_Savo	Nous_Sujet_Savoir	multi-regex
482	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
7140	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
7368	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
8421	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
30069	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
33379	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
46320	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
51363	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
51820	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
53128	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
71915	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
81100	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
100915	nous	Nous_Sujet_Sent	Nous_Sujet_Sentiment	multi-regex
26027	aurait	avoir	VER_cond	multi-regex
26231	aurait	avoir	VER_cond	multi-regex
33477	aurait	avoir	VER_cond	multi-regex
44692	aurait	avoir	VER_cond	multi-regex
55646	aurait	avoir	VER_cond	multi-regex
74327	aurait	avoir	VER_cond	multi-regex
33438	correspondrait	correspondre	VER_cond	multi-regex
19279	devrait	devoir	VER_cond	multi-regex
26688	devrait	devoir	VER_cond	multi-regex
36886	devrait	devoir	VER_cond	multi-regex
39599	devrait	devoir	VER_cond	multi-regex
53000	devrait	devoir	VER_cond	multi-regex

Les items pertinents étant sélectionnés, on peut désormais lancer les modules R sur ces items uniquement :

Modules R

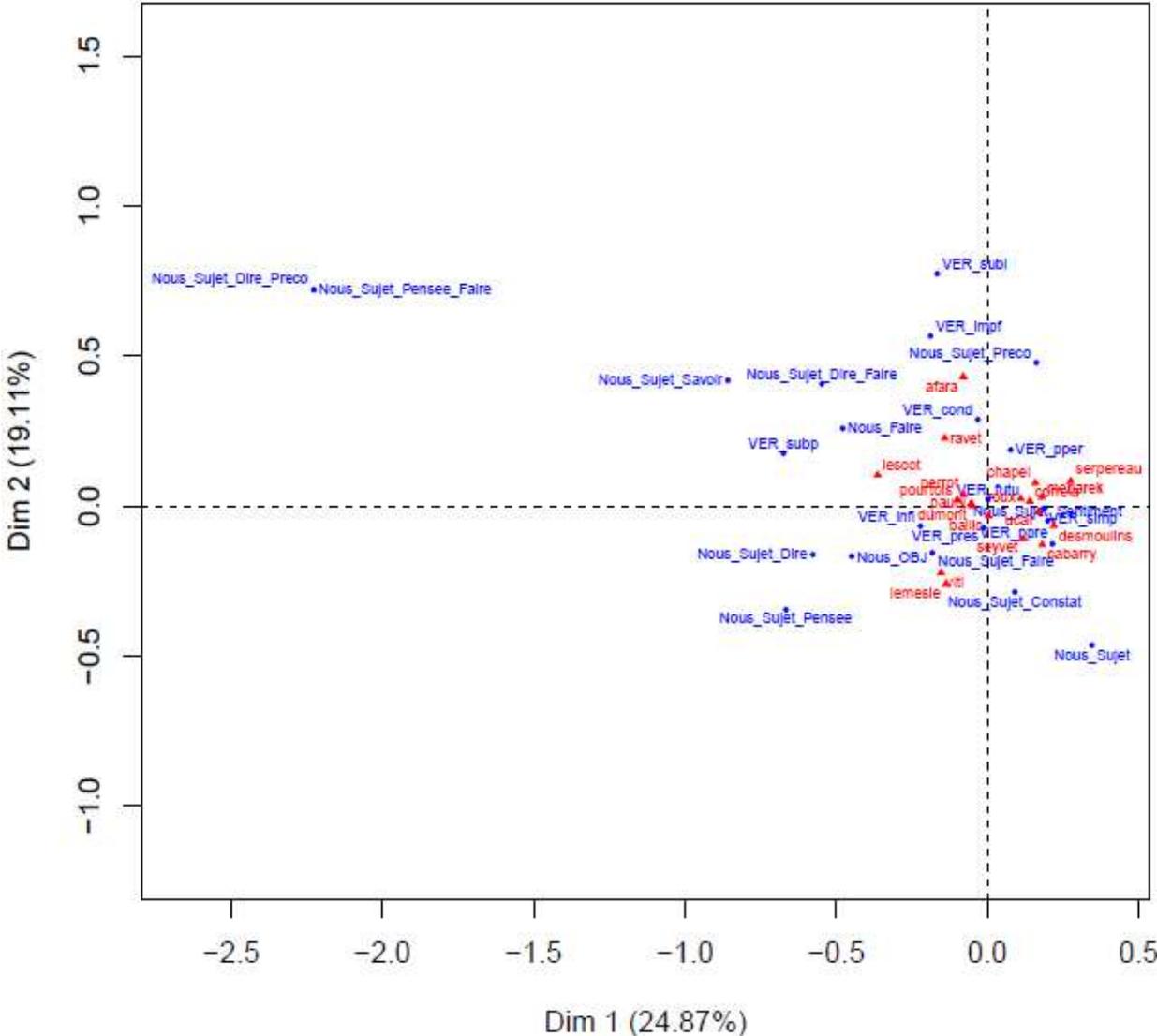
AFC ACP CAH

Sélection Ind 30 Classe 10

Il peut être nécessaire/utile de modifier aussi les paramètres *Ind* et/ou *Classe*.

Calcul de l'AFC

COLs + ROWs (contrib 30)



Calcul de la CAH

