

TalN 5-8 juin 2007

Fabienne Venant

***Utiliser des classes distributionnelles pour
désambiguïser les adjectifs***

Présentation - Enjeux

⇒ Analyse textuelle

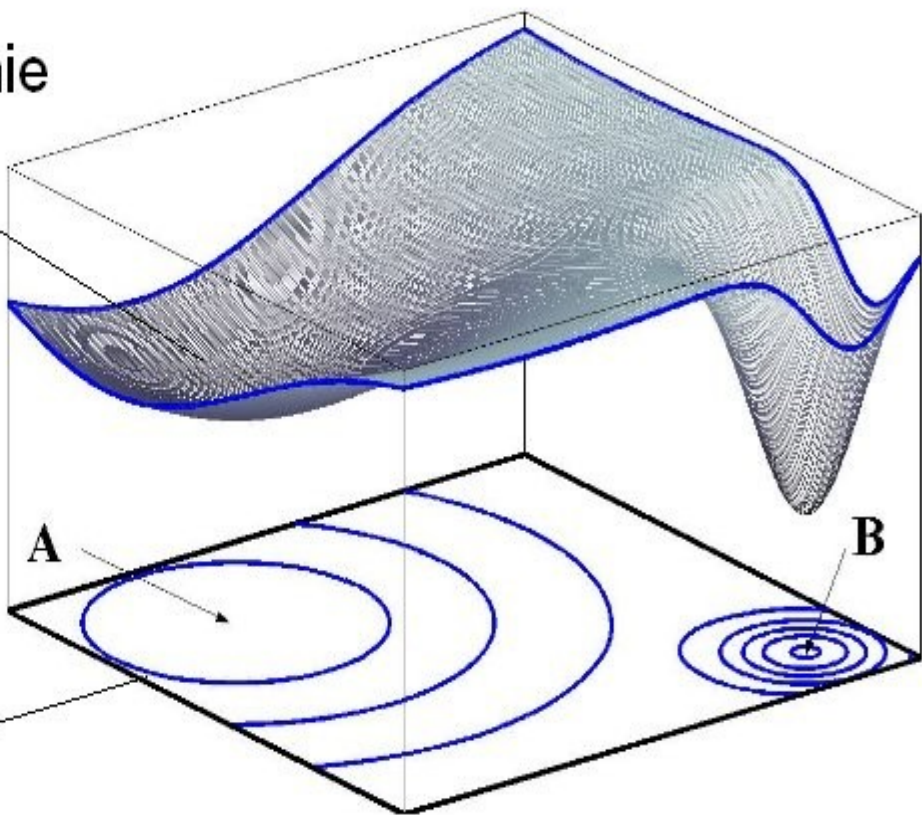
- Prendre en compte la polysémie
- Désambiguïsation de haut niveau

⇒ Modèle dynamique géométrique

- Contraintes définies par le co-texte
- Sémantique adjectivale
 - Influence du nom recteur
une méchante voiture, un méchant garçon
 - Influence de la place
un curieux homme, un homme curieux
« Curieux homme qui ne s'intéresse à rien » (Wilmet)

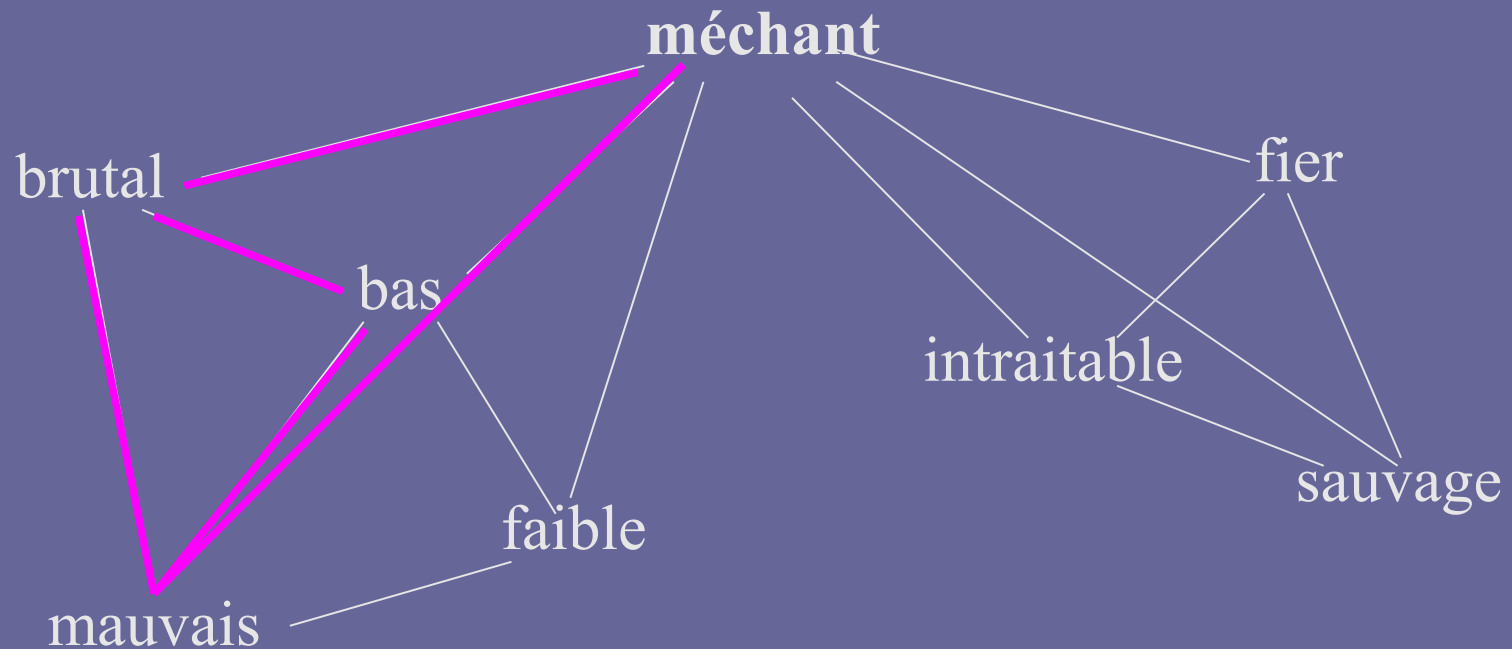
Un modèle dynamique (Victorri et Fuch 96)

Fonction potentielle définie
par le cotexte



Espace sémantique

Construction des espaces sémantiques (Ploux et Victorri 98)



- Cliques :
- bas ; brutal ; mauvais ; méchant
 - bas ; mauvais ; méchant
 - fier ; intraitable ; méchant ; sauvage

Construction des espaces sémantiques

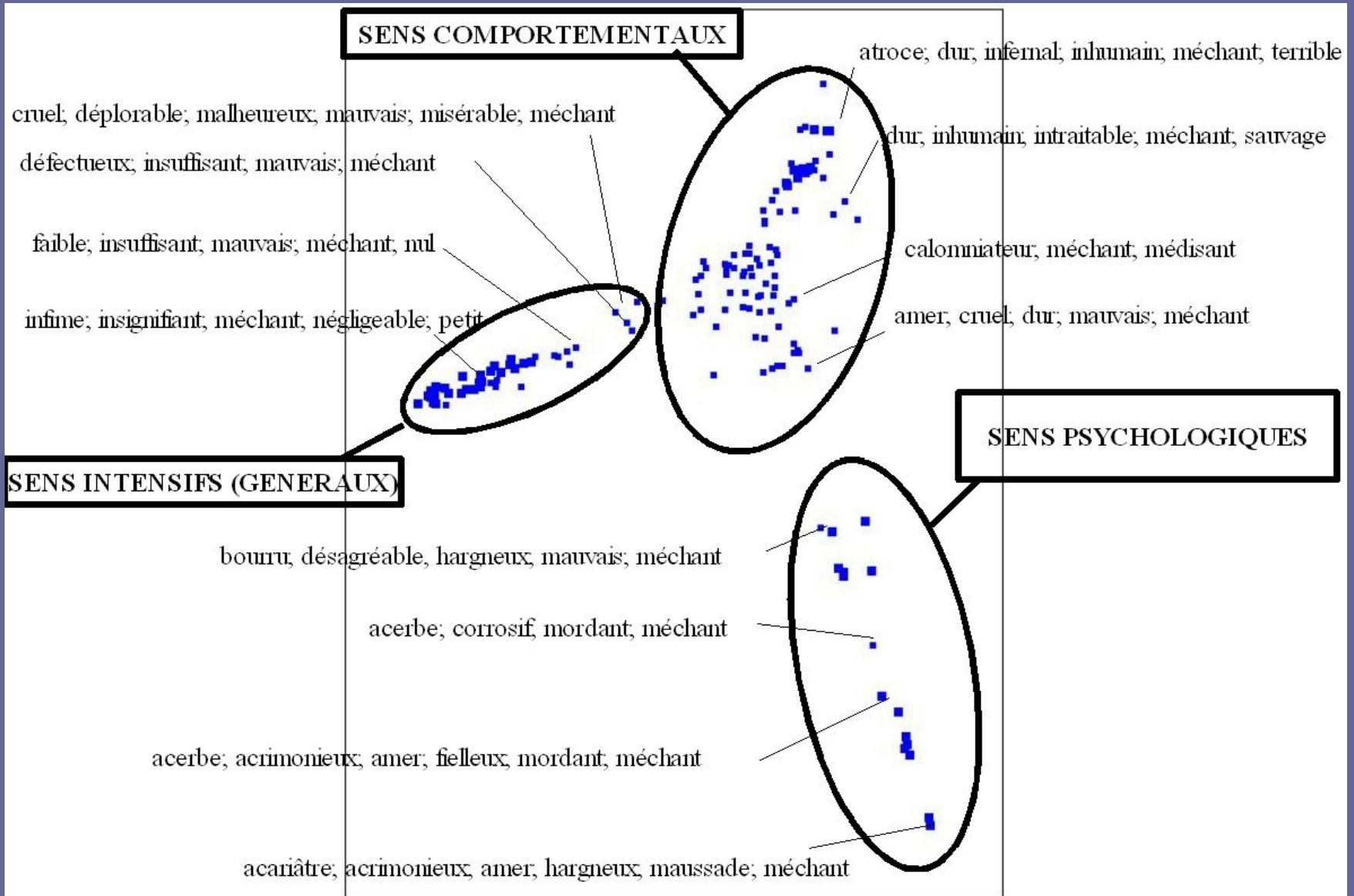
- ⇒ En première approximation, une clique est une nuance de sens
- ⇒ L'espace sémantique est l'espace vectoriel engendré par les synonymes (les sommets du graphe)
- ⇒ Chaque clique est un point de cet espace
- ⇒ Les coordonnées d'une clique sont fonction des synonymes qu'elle contient :

Soient u_1, u_2, \dots, u_n les synonymes engendrant les axes de l'espace,

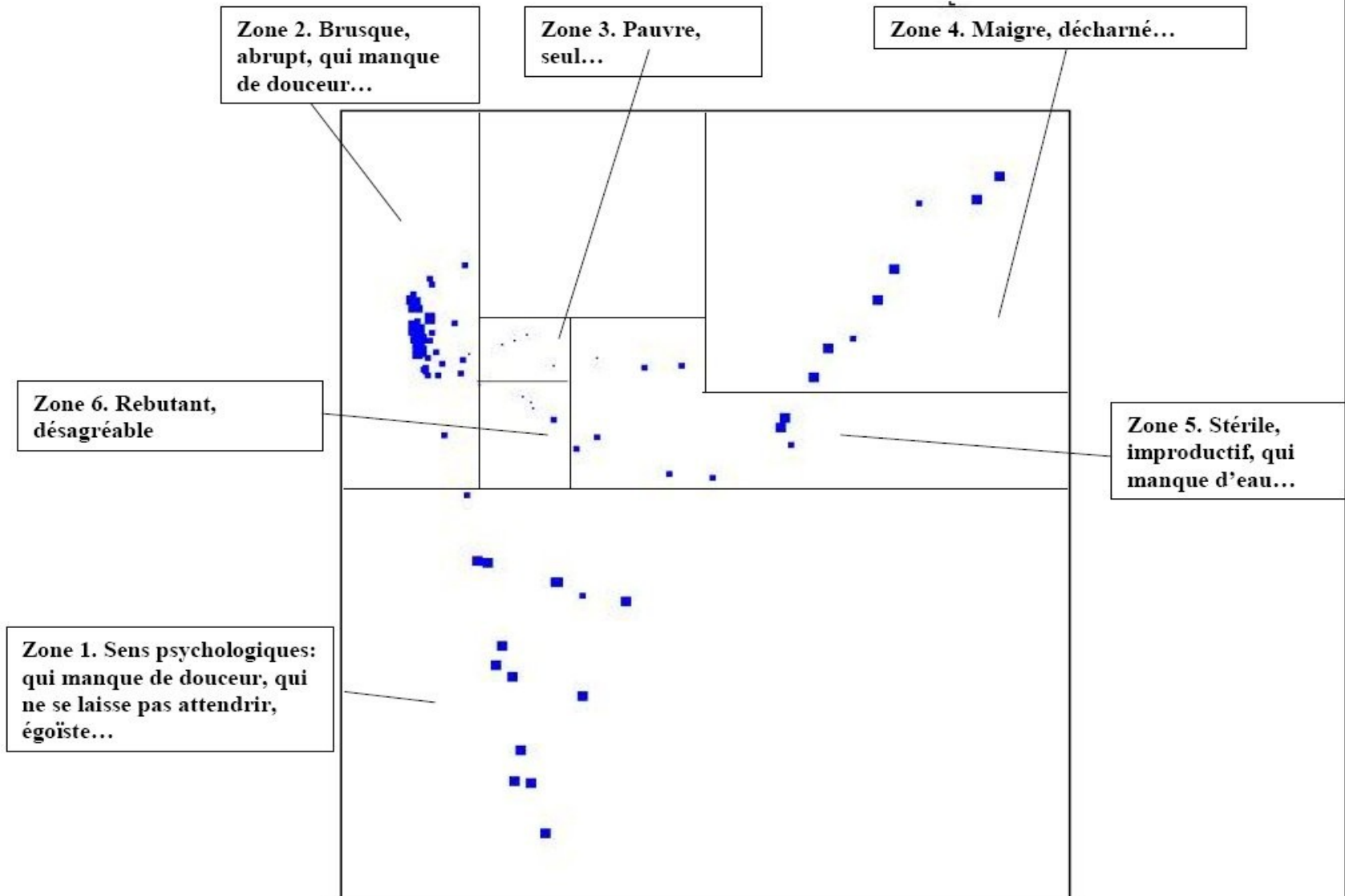
appelons (x_1, x_2, \dots, x_n) les coordonnées de c ,

$$x_i = 1 \text{ si } u_i \in c \text{ et } x_i = 0 \text{ si } u_i \notin c$$

Espace sémantique de méchant



Espace sémantique de sec



Classes de sélection distributionnelle

coup sec

(*coup, bruit, geste*)

Brusque

coup audacieux

(*coup, tentative, expérience*)

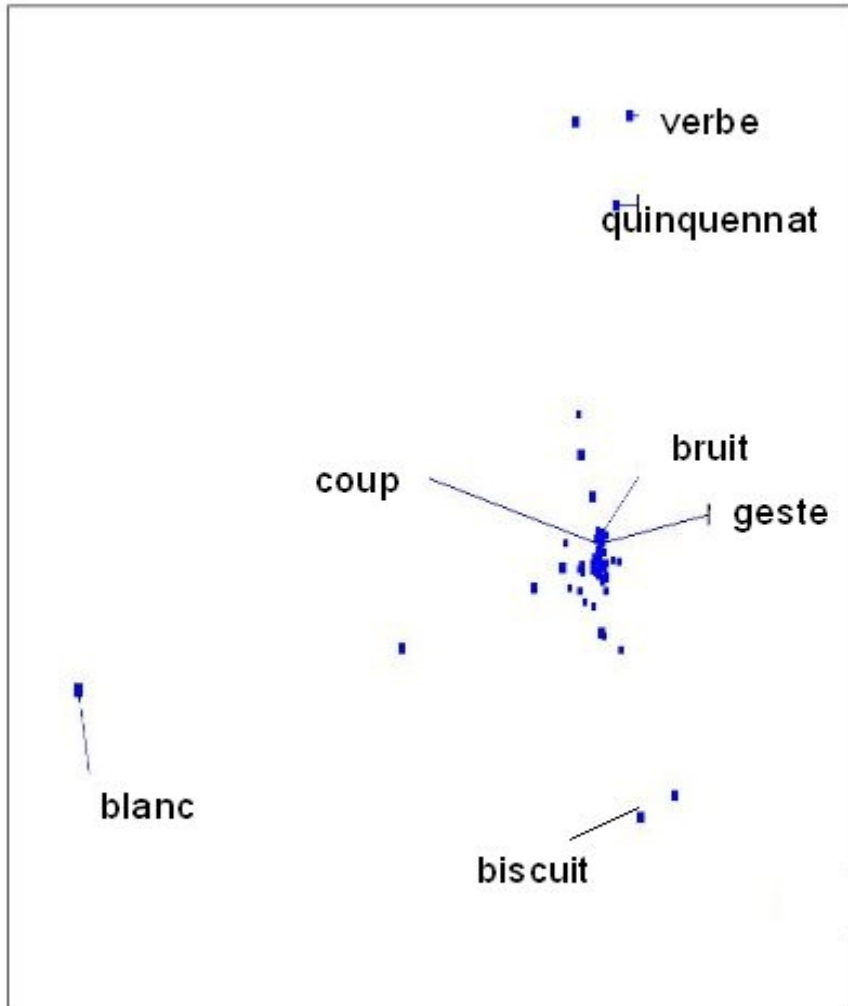
hardi, nouveau

- ⇒ Calculées sur corpus
- ⇒ A partir des sorties de syntax
 - des contextes lexico-syntaxiques
 - être recteur de *sec* en tant qu'épithète (*sec.EPI*)
 - être COD du verbe *donner* (*donner.OBJ*)
 - des mots *sec, coup, geste*
 - des fréquences

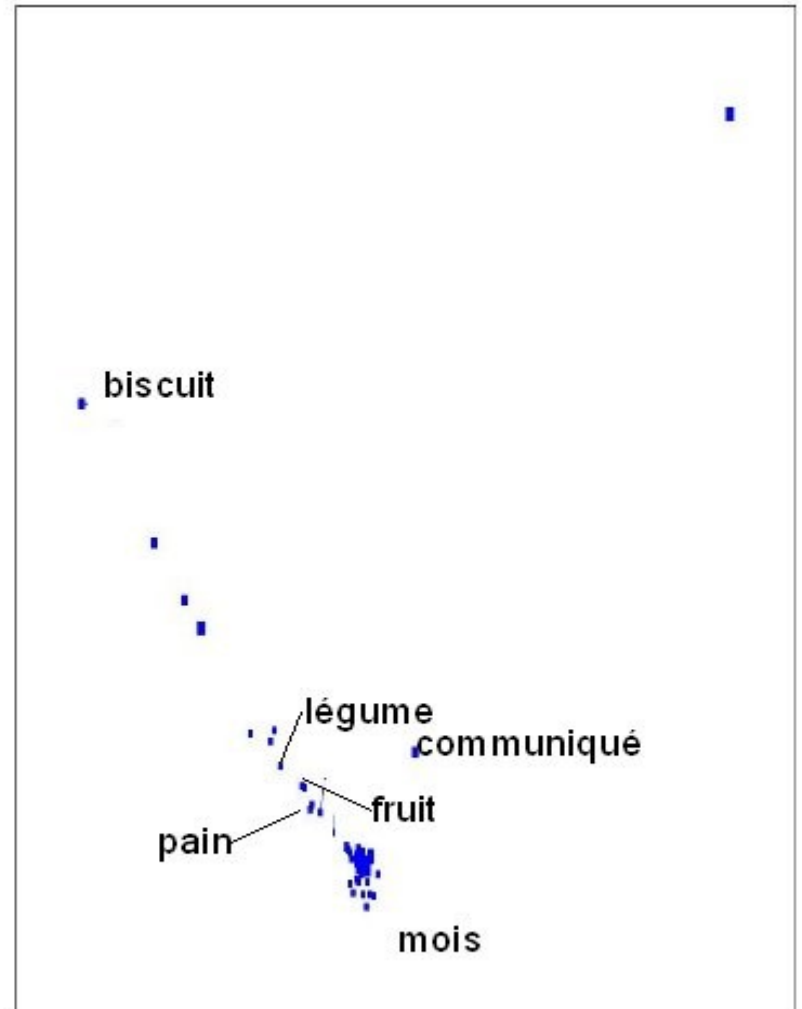
Un espace distributionnel

- ⇒ Engendré par les contextes lexico-syntaxiques
- ⇒ Les points de l'espace sont les mots du corpus
- ⇒ Les coordonnées d'un point sont fonction de la fréquence d'occurrence du mot dans le contexte considéré
- ⇒ Permet le calcul de distances et la catégorisation
- ⇒ Classe d'un nom dans un contexte donné:
 - restriction au sous-espace pertinent
 - *coup sec*: étude des noms attestés comme recteur de sec dans le corpus, dans l'espace engendré par les CLS concernés (au moins un des noms est utilisé dans ce contexte)

Espace distributionnel noms recteurs de sec



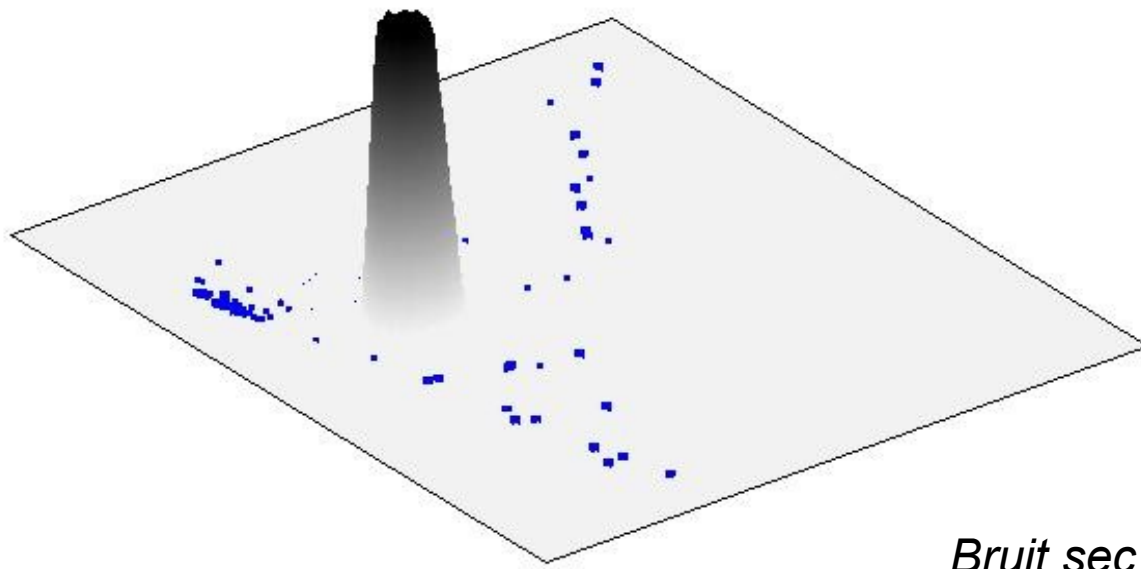
Etude de *bruit sec* – axes 3 et 4



Etude de *fruit sec* – axes 1 et 2

Calcul du sens

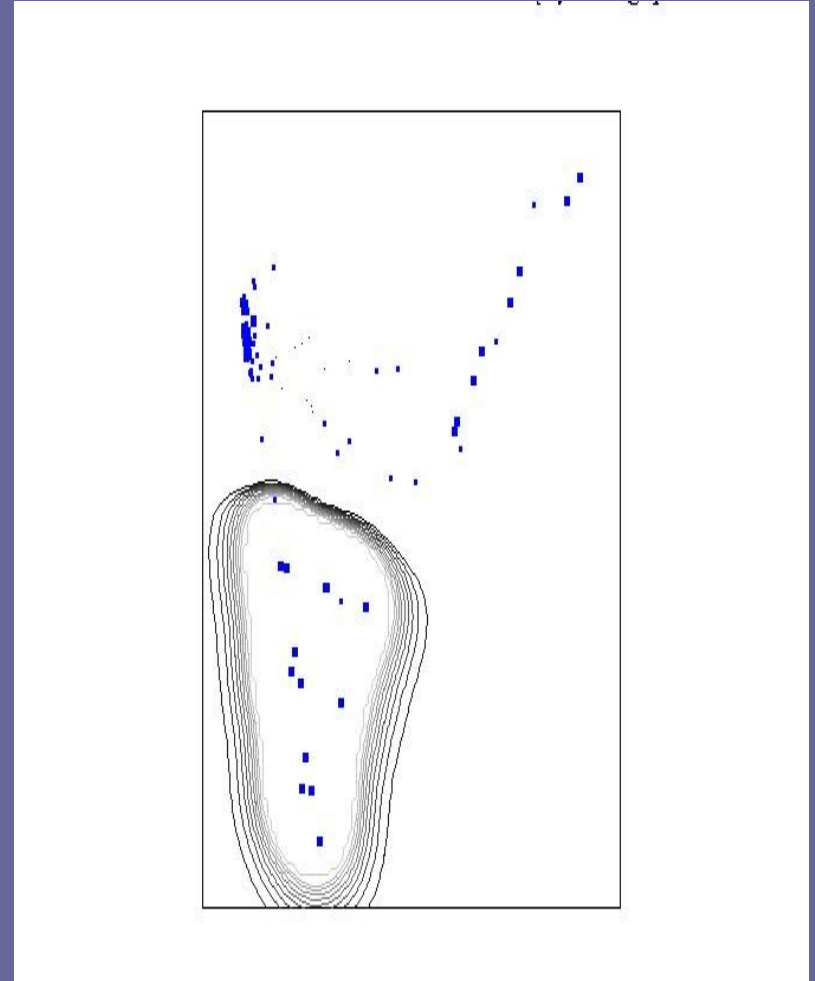
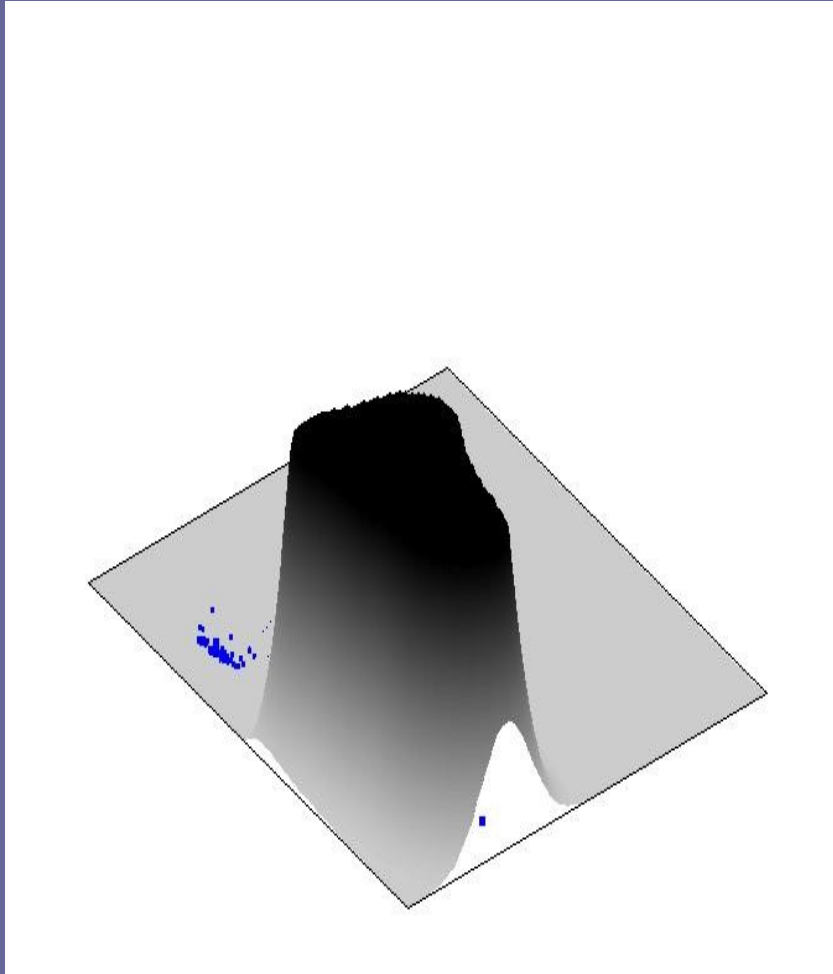
Potentiel désambiguïsateur du nom



Bruit sec

Calcul du sens

Fonction potentiel associée à une zone de sens



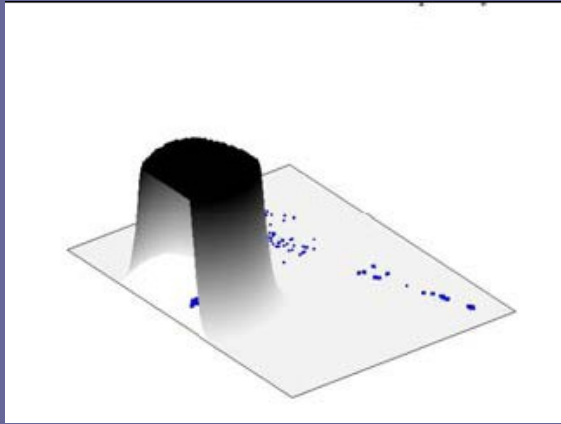
Sens psychologiques de sec

nom	Zone(s)	% affinité
communiqué	manque de douceur	100
corps	maigre, décharné,...	93
coup	manque de douceur	78
humour	psychologique	25
	manque de douceur	22
pain	manque d'eau, improductif,	90
récit	manque de douceur	100
son	manque de douceur	42
	psychologique	33
terrain	manque d'eau, improductif	92
ton	psychologique	90
mois	psychologique	56
		31

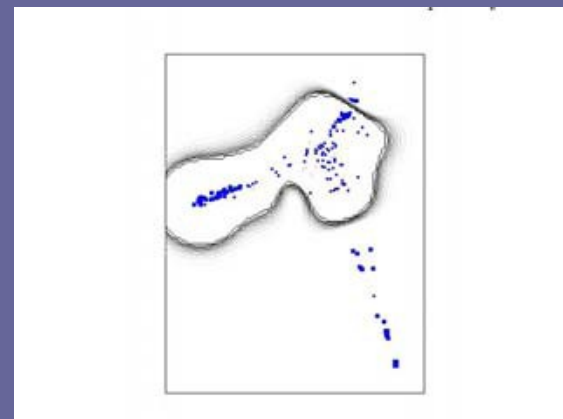
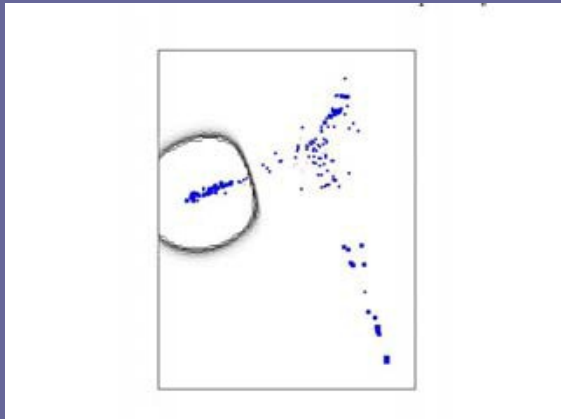
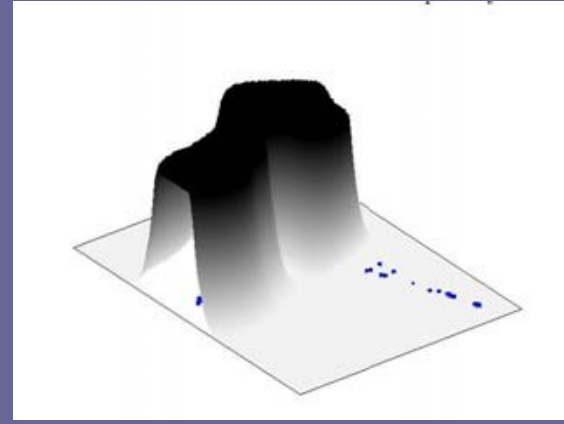
Calcul du sens

Influence de la place de l'adjectif

Classe ANTE de *bête*



Classe POST de *bête*



Résultats

Antéposition

⇒ Sens intensifs omniprésents (extension)

⇒ Repérage des ambiguïtés :

cheval, couleur, eau, espèce, farce, matin, mot, nature, parole, taureau, terre

--> valeur comportementale ou valeur intensive

A. *Sur une route quelconque, à travers la plaine immense et quelconque, au trot de deux **méchants chevaux**, nous allons.* » [Barrès M.]

B. « *Non, non, je veux voir maman; ces **méchants chevaux** ont emporté maman.* » [Comtesse de Ségur]

Résultats

Changement de sens ANTE --> POST

- ➔ *Cheval, couleur, dent, eau, espèce, farce, maison, mot, nature, parole, part, société, taureau, terre :*
 - En antéposition, valeur intensive ou comportementale
 - En postposition, valeur comportementale ou psychologique

- ➔ *Bois, bruit, chemin, chose, corps, rire, voix :*
 - En antéposition, valeur intensive uniquement
 - En postposition, valeur comportementale ou psychologique

- ➔ *Coup, regard, vérité*

Pas de changement de sens: valeur comportementale-

Conclusion

- Désambiguïsation fine
 - Utilisation d'informations lexicales globales
 - Limite de la synonymie comme description du sens
 - Données générales et spécifiques au corpus
 - Pertinence des outils géométriques
 - Mathématiques du continu
- Perspectives
 - Confrontation à des données massives
 - Passage à l'énoncé
 - Interface sémantique / syntaxe
 - Collaboration avec d'autres modèles
 - RI
 - Construction de ressources termino-ontologiques (extraction de métadonnées)