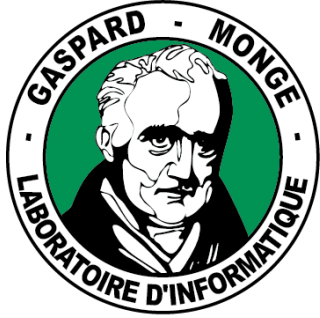




**LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE**

15 janvier 2018
Réunion de consortium Parseme-FR
Marseille

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Collocations ou expressions verbales lexicalisées ?

Éric Laporte, Fabrice Beltran, Takuya Nakamura



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sommaire

Objectif

Méthode

Résultats



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Objectifs

Distinction entre expressions verbales lexicalisées
et collocations

Poursuivre la construction du lexique-grammaire
des expressions verbales figées du français

Conserver les entrées lexicalisées mais non les
collocations



Définitions

Définitions distributionnelles

Collocation

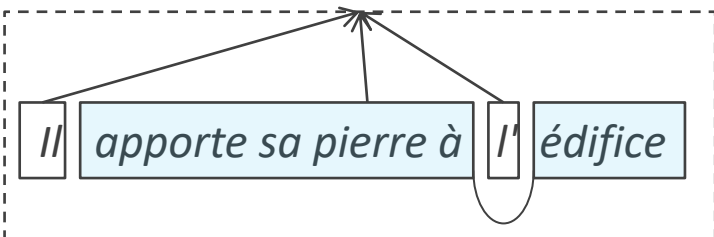
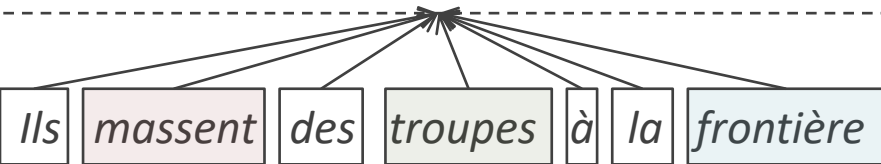
Combinaison compositionnelle, c'est-à-dire dont il est possible de prédire l'emploi uniquement à partir de connaissances sur l'emploi de ses éléments quand ils sont utilisés séparément et dont la fréquence plus élevée (que d'autres combinaisons compositionnelles) n'a pas d'explication extra-linguistique

Ils regroupent des militaires dans la zone frontalière

Expression polylexicale lexicalisée

Expression non compositionnelle

**Il apporte sa pierre au bâtiment*





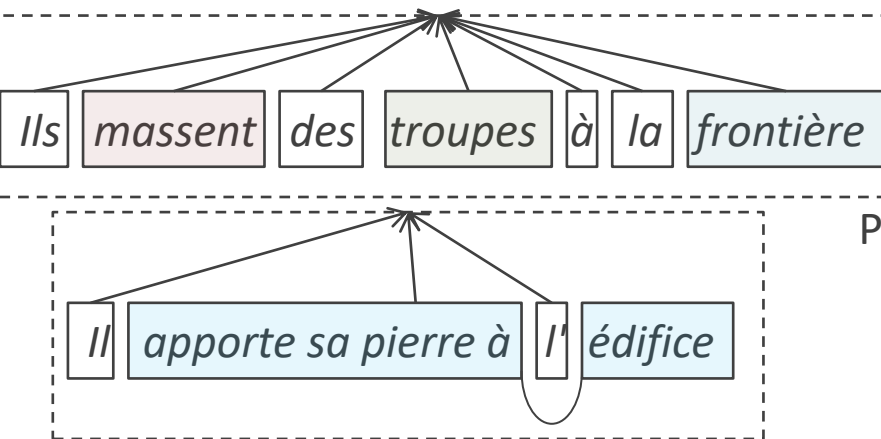
Pourquoi supprimer les collocations du lexique-grammaire

Inutiles pour l'analyse de textes

Les éléments d'une collocation peuvent être décrits par ailleurs dans des entrées lexicales séparées

C'est suffisant pour l'analyse syntaxique

Pour les expressions lexicalisées, des entrées lexicales spécifiques sont nécessaires



Classe c1i *casser un arrêt*
casser un jugement
mais pas *casser une décision*
casser un mariage

prendre son temps pour

Données moins fiables

Le critère statistique (fréquence plus élevée) a été évalué à l'intuition et non sur un corpus

Pour les expressions lexicalisées, les critères sont distributionnels et non statistiques

Point de départ

	figé		N0 =: Nhum	N0 =: N-hum	Ppv =: se fig	<ENT>Ppv	<ENT>V	N0 V	<ENT>DetI	<ENT>C1	C1 =: Npc	[passif]
+	+	+	+	+	<E>	avancer	-	un	argument	-	+	
	+	+	+	+	<E>	avancer	-	un	pion	-	+	
	+	+	+	+	<E>	aviver	-	une	plaie	-	+	
	+	+	+	+	<E>	balancer	-	un	bras	+	-	
	+	+	+	+	<E>	balancer	-	une	jambe	+	-	

Lexique-grammaire

Type de dictionnaire attribuant des propriétés discrétisées à des entrées lexicales discrétisées

Maurice Gross, 1982, Une classification des phrases « figées » du français, *Revue québécoise de linguistique* 11(2), 1982, pp. 151–185

Version de 2001 de ces tables, librement disponibles mais inédites : travail en cours à la mort de Maurice Gross

8 510 entrées

38 658 entrées



Expérience en cours

Séparer les expressions lexicalisées des collocations

Pour l'instant dans un échantillon de 7 classes

C1D	N_0 <i>V le N_1</i>	<i>Il a loupé le coche</i>
C1DPN	N_0 <i>V le N_1 Prép N_2</i>	<i>Il met le cap sur l'Asie</i>
C1G	N_0 <i>V Poss N_1</i>	<i>Il a perdu sa langue</i>
C1GPN	N_0 <i>V Poss N_1 Prép N_2</i>	<i>Il met son nez dans tout</i>
C1i	N_0 <i>V Dind N_1</i>	<i>Il a avalé une couleuvre</i>
C1P2	N_0 <i>V N_1 Prép N_2</i>	<i>Il met de l'eau dans son vin</i>
CFF	<i>P Conj P</i>	<i>Compte dessus et bois de l'eau</i>



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sommaire

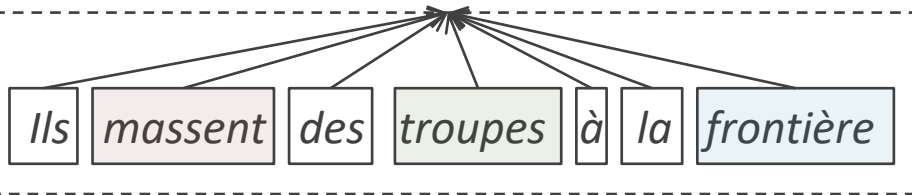
Objectif

Méthode

Résultats



Complémentarité



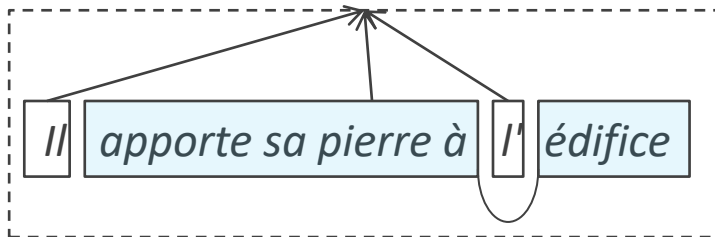
masser

troupes

frontière

Dans un dictionnaire syntaxique

- pour une construction compositionnelle, des entrées lexicales séparées pour les éléments
- pour une expression figée, une entrée lexicale spécifique



apporter sa pierre à Dét édifice



Complémentarité

Complémentarité avec les verbes libres

Une construction ne peut pas être décrite à la fois et pour le même sens

- comme expression figée
- comme construction compositionnelle construite autour d'un verbe simple

Une expression mieux décrite comme figée

Et toc ! reprit Simon qui comptait les points

Lors de ces échanges, la salle comptait les coups

≠ Le groupe compte les absents

Une expression mieux décrite avec un verbe libre

classer : la distribution du complément est un ensemble de substantifs réduit, mais clos par synonymie

Classe 32R3 de verbes libres

compter les points

compter les coups

N_0 compter N_1

classer une affaire

classer un dossier



Complémentarité

Complémentarité avec les constructions à verbe support

Une construction ne peut pas être décrite à la fois et pour le même sens

- comme expression figée
- comme construction à verbe support

Une expression mieux décrite comme figée

perdre sa respiration

Le spectateur en perd sa respiration

N_0 avoir N_1
 $= N_0$ perdre N_1

$\neq$ *Le sucre brésilien a des débouchés internationaux*

$=$ *Le sucre brésilien a perdu ses débouchés internationaux*

Une expression mieux décrite avec un verbe support

défaite : le verbe support *infliger* a un verbe support
converse *subir* avec changement de position des actants

Classe IS2 de noms prédicatifs

subir une défaite



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Expressions terminologiques et techniques

donner la parole à l'accusation

Dans la langue technique, lexicalisées

Dans la langue courante, souvent
compositionnelles

commettre un avocat d'office

mais pas toujours



Cas intermédiaires

vanter sa marchandise à

?*vanter ses marchandises à*

?*vanter ses téléphones à*

*vanter les qualités de ses
marchandises à*

Distribution du complément direct :

- *sa marchandise* OK
- autres concrets beaucoup plus rares
- abstraits OK

Si on considère acceptable N_0 *vanter Nconc à N_1*
dans une entrée de *vanter*, alors *vanter sa
marchandise* est une collocation

Si on considère inacceptable N_0 *vanter Nconc à N_1* ,
vanter sa marchandise est lexicalisé



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sommaire

Objectif

Méthode

Résultats



Résultats

Classe	Effectif	Marquées	Lexicalisées		Collocations		Non marquées
c1p2	1 201	1 201	933	78%	268	22%	0
c1d	1 678	1 050	764	73%	286	27%	628
c1g	671	48	29	60%	19	40%	623
c1gpn	404	370	246	66%	124	34%	34
c1dpn	554	0	0	-	0	-	554
c1i	1 731	358	157	44%	201	56%	1 373
cff	206	162	150	93%	12	7%	44
total	6 445	3 189	2 279	71%	910	29%	3 256



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Conclusion

La distinction entre entrées lexicalisées et collocations
compositionnelles est un découpage fin

L'automatisation est un vrai défi



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Merci

CONTACT

ÉRIC LAPORTE

00 +33 (0)1 60 95 75 52

ERIC.LAPORTE@UNIV-PARIS-EST.FR