

Repérer la relation d'*Élaboration* : Combiner un marqueur ambigu, le gérondif, avec des indices lexicaux, les voisins distributionnels

Clémentine Adam & Marianne Vergez-Couret

CLLE - Université de Toulouse

10 novembre 2009

Plan

- 1 Introduction
- 2 État de l'art
- 3 Genèse de notre travail
- 4 Expérience
- 5 Conclusions et perspectives

Introduction

- Étude très circonscrite sur corpus (Wikipédia)
- avec un objectif précis : le repérage automatique d'une relation de discours, la relation d'*Élaboration*
- en utilisant des éléments repérables automatiquement (gérondif et voisins distributionnels)
- et un objectif plus vaste : montrer la pertinence d'indices lexicaux pour le repérage de la structure du discours
- qui fait l'objet d'une soumission pour le colloque MAD10 : "Signalling Text Organisation"

Plan

- 1 Introduction
- 2 État de l'art
- 3 Genèse de notre travail
- 4 Expérience
- 5 Conclusions et perspectives

Plan

- 1 Introduction
- 2 État de l'art
- 3 Genèse de notre travail
- 4 Expérience
- 5 Conclusions et perspectives

Cohérence et cohésion

- L'analyse du discours bénéficie actuellement d'un regain d'intérêt dû à ses enjeux pour le TAL
- Elle repose sur l'observation selon laquelle un texte n'est pas une simple succession de phrases, mais un tout cohérent.
- Cette cohérence est reflétée par les observables que sont les marques de cohésion.
- Procédés cohésifs classiquement considérés (Halliday & Hasan, 1976) : référence, substitution, ellipse, conjonction et cohésion lexicale
- La cohésion lexicale est vue comme "the dominant mode of creating texture" (Hoey, 1991)

Relations mises en jeu dans la cohésion lexicale

Suivant Halliday & Hasan (1976) :

- Des relations de réitération : répétition lexicale, reprise par un synonyme ou par un hyperonyme, etc.

(1) I turned to the ascent of the peak.
The [ascent|climb|task|thing] is perfectly easy.

- Des relations dites de collocation : associent des mots présentant une tendance à apparaître ensemble, mais ne relevant pas de la réitération.

*malade/docteur, essayer/réussir, abeille/miel,
porte/fenêtre*

Cohésion lexicale et relations *non classiques*

- Les relations les plus pertinentes pour le repérage des structures discursives sont dans la plupart des cas des relations échappant aux typologies traditionnelles (Morris & Hirst, 2004)
 - Lorsqu'il s'agit d'interpréter un texte, les relations comme la synonymie, l'antonymie, etc. cèdent le pas à des relations *non classiques*
- Ces relations sont plus difficiles à appréhender dans les textes → la cohésion lexicale est peu exploitée sur le plan applicatif

Quelles ressources pour capter la cohésion lexicale ?

- La finesse de la détection des liens de cohésion lexicale est tributaire de la ressource utilisée pour estimer les liens présents entre les mots
 - Sans ressource, seuls les phénomènes de répétition lexicale sont accessibles
 - Quelles ressources ?
 - Roget's Thesaurus (Morris & Hirst, 1991)
 - WordNet (Hirst & St-Onge, 1998)
- Permettent d'appréhender des relations classiques

Quelles ressources pour capter la cohésion lexicale ?

Quelles ressources pour dépasser les relations classiques ?

- JeuxDeMots (LIRMM) : ressource alimentée grâce aux contributions d'internautes, à travers un jeu (Lafourcade *et al.*, 2009) → travail de Morlane-Hondère (2009).
- Collocations ou cooccurrences (Choi *et al.*, 2001; Ferret, 2002) : la proximité sémantique entre les mots est évaluée en fonction de leur co-présence au sein de mêmes phrases, paragraphes ou textes dans un corpus
- Voisins distributionnels : mots rapprochés par l'analyse distributionnelle d'un corpus

Le projet VOILADIS

- VOILADIS (VOIsinage Lexical pour l'Analyse du DIScours) : Projet du PRES Toulouse coordonné par C. Fabre impliquant des chercheurs des laboratoires CLLE-ERSS et IRIT
- Ressource mobilisée : une base de *voisins distributionnels*
- Objectif : utiliser des indices lexicaux pour la mise au jour de phénomènes discursifs
- À un niveau macroscopique : certaines zones de texte peuvent émerger parce qu'elles présentent une forte cohésion lexicale.
- À un niveau micro/ascendant : les indices lexicaux peuvent permettre d'inférer des relations de discours entre segments

Qu'est ce qu'une structure du discours ?

- textes cohérents = discours structurés
- Approche ascendante = structure via des relations de discours entre segments (annodis)
- Analyse exhaustive (textes cohérents = textes dont tous les segments sont reliés par des relations de discours)

Reconnaître la structure du discours : pourquoi ?

- Interpréter les expressions référentielles
- Résoudre les anaphores pronominales interphrastiques
- Identifier l'ordre temporel des événements
- etc.

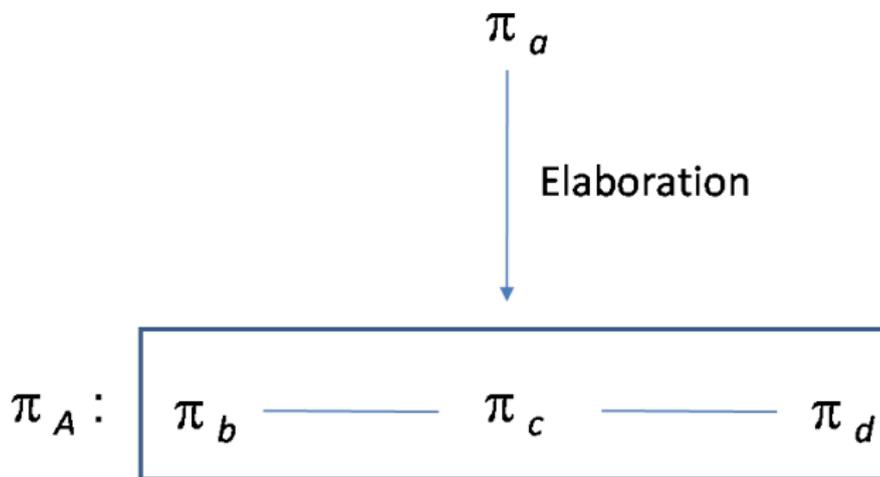
nécessitent des connaissances sur la structure du discours
(Grosz & Sidner 1986, Asher & Lascarides 2003)

Reconnaître la structure du discours : comment ?

- Segments de discours et Relations de discours
 - (i) nature exacte des segments discursifs
 - (ii) nature exacte des relations de discours
 - (iii) structure hiérarchique que ces relations engendrent sur la structure du discours
 - (iv) reconnaissance des relations de discours

Nature exacte des segments discursifs

- Segments minimaux (contenu sémantique clair)
- Segments complexes = emballage organisateur (en formant des groupes de segments minimaux)

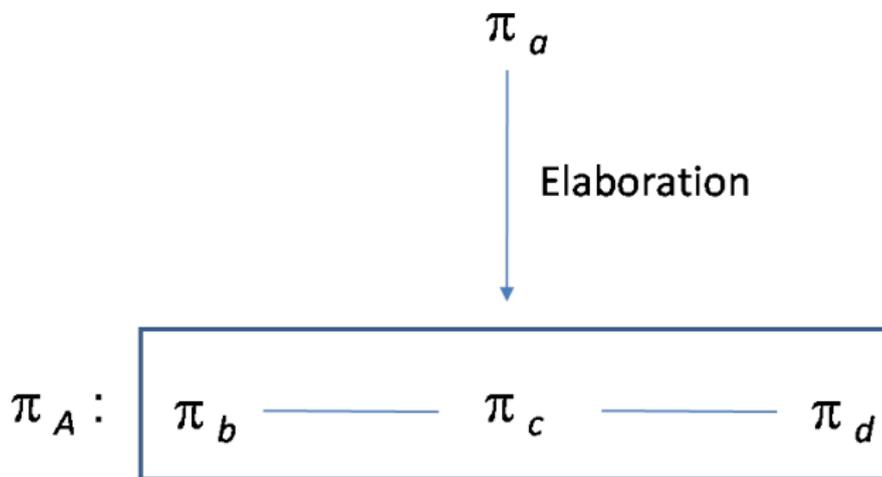


Nature exacte des relations de discours

- Nature différente des relations de discours
 - relations intentionnelles (Mann & Thompson 1987)
 - relations structurelles (Grosz & Sidner 1986)
 - relations sémantiques (Asher 1993, Asher & Lascarides 2003)
- Nombre de relations de discours
 - liste fermée (Hobbs 1990, Asher 2003)
 - liste ouverte (Mann & Thompson 1987)
 - hiérarchie (Hovy & Maier 1991)

Structure hiérarchique

- Relations coordonnantes (narration, continuation, contraste)
- Relations subordonnantes (élaboration, explication)



Identification des relations de discours (1)

- Basé sur la reconnaissance des intentions du locuteur (Mann & Thompson 1987)
- Basé sur des marqueurs de discours (Knott 1996)
- Basé sur la sémantique des énoncés individuels pour atteindre compositionnellement le sens du discours (+ marqueurs lexicaux, structures syntaxiques, marqueurs aspectuo-temporels, sémantique lexicale et connaissances du monde) (Asher & Lascarides 2003)

Identification des relations de discours (2)

- Sources d'information très riches
- mais souvent peu explicitées

Identification des relations de discours en SDRT

- La SDRT cherche à rendre compte de la façon dont différents types d'information interagissent dans l'identification des relations de discours
 - sémantique lexicale et grammaticale
 - structure syntaxique
 - marqueurs lexicaux
 - connaissances du monde
- La SDRT s'appuie sur une logique non monotone chargée d'inférer les liens de cohérence entre le nouveau constituant et le contexte discursif

Reconnaître la structure du discours : Applications

- Résumé automatique (Marcu 2000)
- Système de question/réponse

Le repérage systématique des relations de discours requiert un inventaire aussi riche que possible des marqueurs disponibles pour le repérage automatique.

Nécessité d'identifier des marques de surface pour repérer les relations de discours

- Cela permet d'une part de valider les structures de discours à l'aide de forme de surface leur correspondant.
- Cela permet d'autre part de découvrir de nouvelles structures et d'affiner les connaissances des structures existantes sur la base d'études systématiques des marqueurs discursifs.

Plan

- 1 Introduction
- 2 État de l'art
- 3 Genèse de notre travail**
- 4 Expérience
- 5 Conclusions et perspectives





Définition en SDRT

- Relation subordonnante
- *Élaboration* peut être inférée de façon non monotone grâce à la présence d'une relation de subsumption (*Subtype_D*) entre les types des éventualités des segments à relier indiquant que le type de la seconde éventualité est un sous-type de celui de la première
- dans la sémantique lexicale des éventualités ou grâce à des connaissances du monde restreintes au contexte discursif.

Illustration

(2) Max ate a lovely meal. He devoured lots of salmon.

- eat/devour
- meal/salmon

Repérage automatique de la relation d'Élaboration

- Relation difficile à repérer automatiquement
- Pas de signaux de surface (Scott & Souza, 1990 ; Knott 1996 ; Knott *et al.*, 2001)
- Marqueur du discours *specifically*
- Marqueur non linguistique basé sur le nombre de phrases dans un paragraphe ou le nombre de paragraphes dans une section : si ce nombre est petit et qu'aucun marqueur du discours n'est utilisé, alors la relation entre les phrases ou entre les paragraphes est généralement *Élaboration* (Marcu 2000)
- Des marqueurs du discours ambigus : *d'abord*, le gérondif...

Du Sous-type à la notion de similarité lexicale

- Partant de l'hypothèse de la SDRT selon laquelle l'*Élaboration* est marquée lexicalement, le but est de mettre en place une détection automatique
- Nécessité d'une ressource
- Sous-type lexical → hyperonymie ?
- L'observation de données réelles (annotations issues de la phase exploratoire du projet ANNODIS) montre que la relation d'*Élaboration* met en jeu des phénomènes lexicaux plus complexes que de strictes relations lexicales de sous-type

Du Sous-type à la notion de similarité lexicale

- (3) [Un véhicule a effectué une spectaculaire sortie de route, hier vers 18 h 15, sur l'A36.]¹ [La voiture circulait dans le sens Mulhouse-Montbéliard]² [lorsqu'après être passée à hauteur du 35e RI,]³ [elle a quitté la chaussée sur sa droite.]⁴
- Ici, trois liens permettent d'inférer $Subtype_D(\pi_1, \pi_4)$:
 - "véhicule"/"voiture" → relation d'hyponymie
 - "sortie"/"quitter" → proximité sémantique évidente, mais relation intercatégorielle
 - "route"/"chaussée" → relation relevant plutôt de la méronymie

Du Sous-type à la notion de similarité lexicale

(4) (...) [qui rappelle la vocation des bénévoles de l'association :]32 [être un soutien pour la paroisse,]33 [apporter une petite contribution financière aux travaux grâce aux manifestations et aux dons,]34 [accomplir de multiples tâches et démarches touchant aux bâtiments paroissiaux,]35 [contribuer à la convivialité entre les paroissiens.]36

- Au niveau discursif : $Subtype_D(\pi_{32}, [\pi_{33} - \pi_{36}])$
- Au niveau lexical : liens entre "vocation" et "soutien", "contribution", "accomplir", "tâches", etc.
- Relations de proximité lexicale établies dans le discours, non recensées par des ressources classiques.

Notre choix de ressource : les Voisins de Wikipédia

- Nous avons exclu l'emploi d'une ressource générique
- Notre choix s'est porté sur une ressource construite par l'analyse distributionnelle d'un corpus : les *Voisins de Wikipédia*.
 - Ressource à notre disposition, et mobilisée dans le cadre de voiladis
 - Ressource susceptible de mettre au jour des liens d'ordre paradigmatique
- Ressource construite à partir de Wikipédia (version d'avril 2007) : environ 194 millions de mots, 470 000 articles
- Chaîne de traitement : chaîne Syntex (analyse syntaxique)
 - Upéry (analyse distributionnelle), développée par D. Bourigault et adaptée à Wikipédia par F. Sajous.

Construction de la ressource



Le Kilimandjaro ou Kilimanjaro est une montagne située au nord-est de la Tanzanie et composée de trois volcans éteints. Le plus à l'est, culminant à 5895 mètres d'altitude, se situe à l'est. Il fait à l'inverse à 5140 mètres d'altitude et le Kibo, le plus récent géographiquement, situe entre les deux autres et culme le 91 000 à 5895 mètres d'altitude contre le point culminant de l'échelle. Cette cote caractéristique, le Kilimanjaro est connu pour sa cime glaciaire sommitale en phase de retraite accélérée depuis le début de l'ère glaciaire et qui devrait disparaître totalement d'ici 2025 à 2030. La base des pentes latérales magiques qui est elle-même est souvent attribuée au réchauffement climatique mais la déforestation est également un facteur majeur. Ainsi, malgré la grande quantité de précipitations de 1570 et alors même qu'elle joue un rôle essentiel dans la régulation bioclimatique du cycle de l'eau, le glacier fondra continu à la monter.

En effet, le marigage est notamment le facteur des pertes de la nuit et à l'aube qui ont besoin de prises d'altitude pour faire passer leurs troupeaux et des individus sauvages en fait à l'est qui culminent des pentes latérales plus élevées sur les plateaux, malgré les prises de conscience depuis le début du 20^{ème} siècle.

Après la surprise engendrée dans le milieu scientifique avec sa découverte contemporaine par Johannes Rebmann en 1849, le Kilimanjaro a été le fruit des explorations comme celle de Hans Meyer ou Ludwig Reichert qui parvint au sommet en 1889 accompagné de leur guide Tlibao Kinyasi Luvu. Par le suite, il a constitué une base d'exploration qui se sont étendues catholiques et protestants. Enfin, après plusieurs années de colonisation allemande par le territoire, il a vu l'émergence d'une élite chagga qui a été un rôle dans la naissance d'une identité nationale avec comme point d'orgue l'indépendance de la Tanzanie en 1961.

Depuis, le Kilimanjaro est devenu une montagne emblématique, souvent se représente dans les arts et symbolise une de nombreux produits à valeur commerciale. Elle est très présente par son relief, le randonneur qui réalise son ascension tout en profitant de la grande diversité de sa faune et de sa flore.

Le nom utilisé pour désigner le marigage dans son ensemble est orthographe Kilimanjaro en français et Kilimanjaro en anglais. Elle est aussi appelée C. Dainton. C'est en fait, une "Montagne blanche" ou "Montagne éternelle". Son nom a été adopté en 1880 et venait du swahili Kilima Njaro. Son nom a été adopté en 1880 et venait du swahili Kilima Njaro.

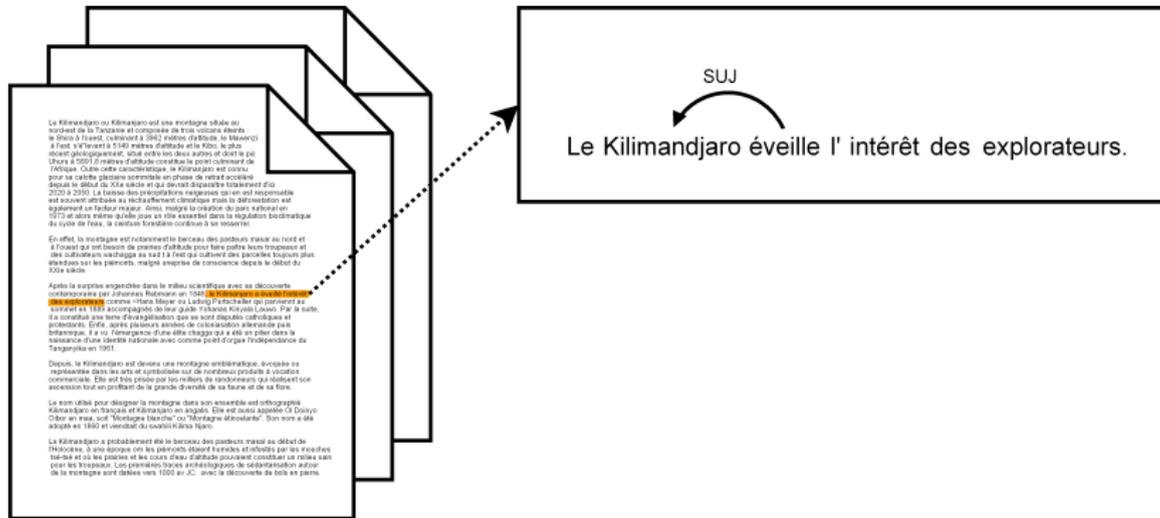
Le Kilimanjaro a probablement été le berceau des premiers peuples du début de l'histoire. Il est associé aux plateaux élevés et isolés par les vallées profondes et où les plantes et le cours d'eau d'altitude pouvaient constituer un refuge sûr pour les troupeaux. Les premières traces archéologiques de sédentarisation autour de la montagne sont datées vers 1000 av. J.C. avec la découverte de bols en pierre.

Construction de la ressource

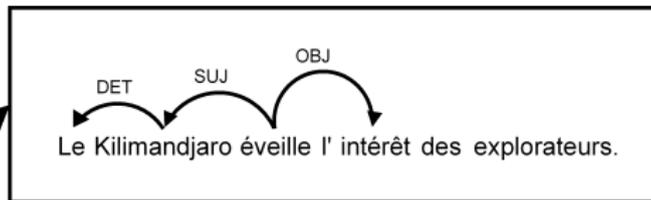


Le Kilimandjaro éveille l'intérêt des explorateurs.

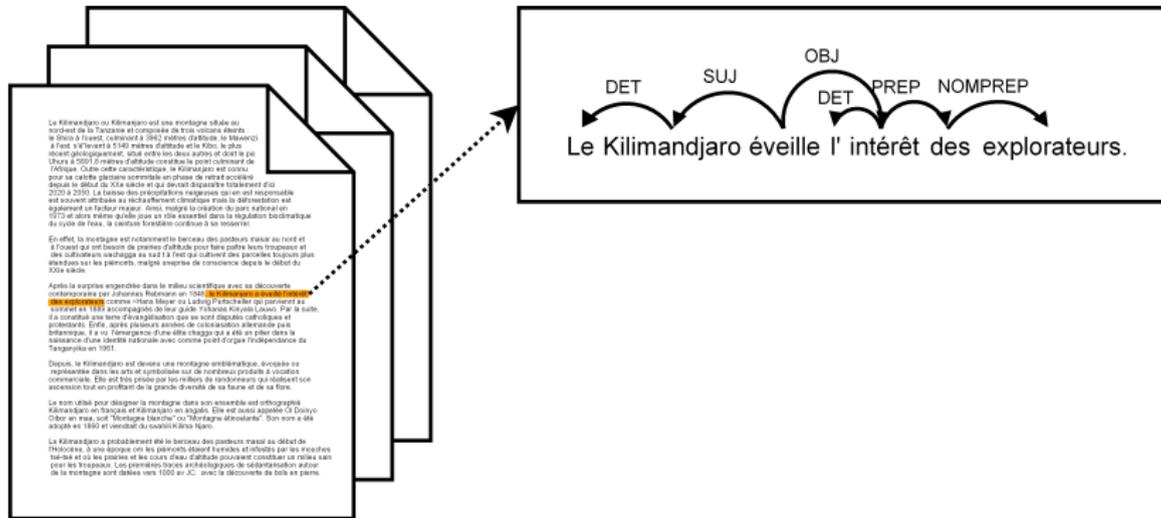
Construction de la ressource



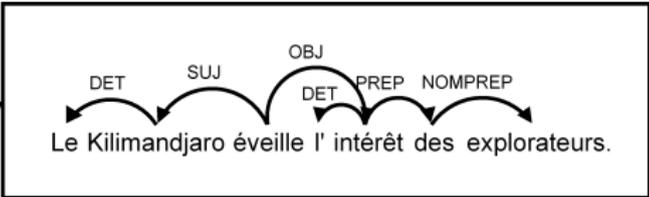
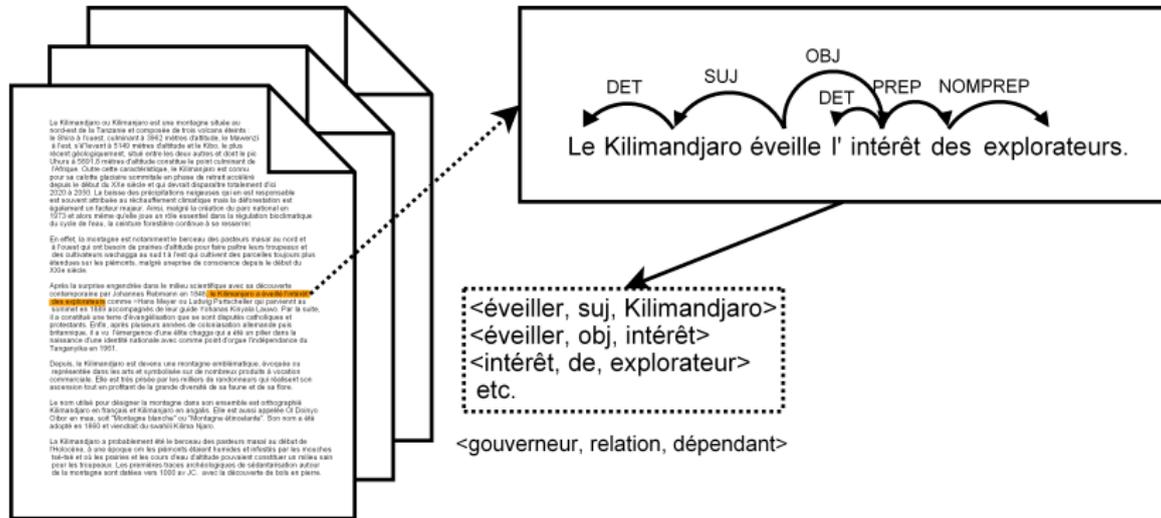
Construction de la ressource



Construction de la ressource



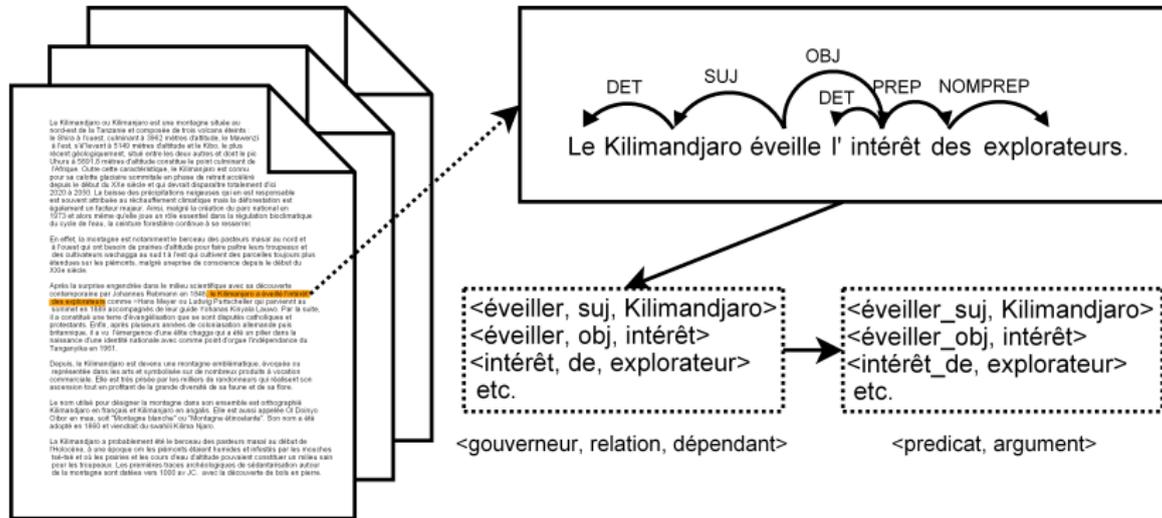
Construction de la ressource



- <éveiller, suj, Kilimandjaro>
- <éveiller, obj, intérêt>
- <intérêt, de, explorateur>
- etc.

<gouverneur, relation, dépendant>

Construction de la ressource



Construction de la ressource

Autres couples <prédicat, argument> extraits dans le corpus :

Construction de la ressource

Autres couples <prédicat, argument> extraits dans le corpus :

- <éveiller_obj, soupçon>
- <éveiller_obj, curiosité>
- <éveiller_obj, conscience>
- <éveiller_obj, attention>

etc.

Construction de la ressource

Autres couples <prédicat, argument> extraits dans le corpus :

- <éveiller_obj, soupçon>
- <éveiller_obj, curiosité>
- <éveiller_obj, conscience>
- <éveiller_obj, attention>
- etc.
- <taux_de, intérêt>
- <défendre_obj, intérêt>
- <groupement_de, intérêt>
- <servir_obj, intérêt>
- <porter_obj, intérêt>
- etc.

Construction de la ressource

Autres couples <prédicat, argument> extraits dans le corpus :

- <éveiller_obj, soupçon>
- <éveiller_obj, curiosité>
- <éveiller_obj, conscience>
- <éveiller_obj, attention>
- etc.
- <taux_de, intérêt>
- <défendre_obj, intérêt>
- <groupement_de, intérêt>
- <servir_obj, intérêt>
- <porter_obj, intérêt>
- etc.

La similarité des distributions est évaluée grâce au score de *Lin*. Les mots ainsi rapprochés sont nommés *voisins distributionnels*

Construction de la ressource

Relations de voisinage distributionnel établies par Upéry :

éveiller_obj / exciter_obj

Contextes partagés

curiosité
convoitise
appétit
imagination
désir
passion
esprit
intérêt

Construction de la ressource

Relations de voisinage distributionnel établies par Upéry :

éveiller_obj / exciter_obj
éveiller_obj / raviver_obj

Contextes partagés

soupçon
nostalgie
crainte
souvenir
désir
inquiétude
sentiment
espoir
passion
intérêt

Construction de la ressource

Relations de voisinage distributionnel établies par Upéry :

éveiller_obj / exciter_obj
éveiller_obj / raviver_obj
éveiller_obj / endormir_obj

Contextes partagés

méfiance
vigilance
conscience
enfant
homme

Construction de la ressource

Relations de voisinage distributionnel établies par Upéry :

éveiller_obj / exciter_obj
éveiller_obj / raviver_obj
éveiller_obj / endormir_obj

sortie_de / quitter_obj

Contextes partagés

territoire
système éducatif
conservatoire
hôpital
autoroute
etc.

Construction de la ressource

Relations de voisinage distributionnel établies par Upéry :

éveiller_obj / exciter_obj

éveiller_obj / raviver_obj

éveiller_obj / endormir_obj

sortie_de / quitter_obj

véhicule / voiture

Contextes partagés

circuler_à bord de
pare-brise_de

portière_de

percuter_suj

immatriculation_de

etc.

Notre choix de ressource : les Voisins de Wikipédia

- La base obtenue pour l'ensemble de l'encyclopédie Wikipédia compte environ 4 millions de couples, couvrant un large éventail de relations de proximité sémantique (Fabre & Bourigault, 2006)

Un véhicule a effectué une spectaculaire sortie de route hier vers 18 h 15 , sur l' A36 . La voiture circulait dans le sens Mulhouse-Montbéliard lorsqu' être passée à hauteur du 35e RI, elle a quitté la chaussée sur sa droite. Frôlant le début d' une glissière de sécurité, le véhicule a gravi le talus, basculé de l' autre côté, traversé un champ, est entré dans un secteur boisé, pour finalement plonger vers le centre Leclerc dans une zone à pic .

- Liens pertinents pour notre tâche : "véhicule"/"voiture", "sortie"/"quitter", "route"/"chaussée"
- Liens participant à la cohésion lexicale de l'extrait : "route"/"circuler"
- Liens non-pertinents dans ce contexte : "route"/"traverser"

Gérondif = candidat parfait

- Certaines propositions gérondives sont considérées comme des élaborations de la proposition principale mais pas toutes.
- Extraction des gérondifs avec l'analyseur syntaxique syntex

Sur le gérondif

- Le gérondif établit un rapport de subordination syntaxique entre deux verbes.
- Deux procès sont liés.
- La valeur sémantique exprimée par la proposition gérondive n'est pas véhiculée par le gérondif lui-même mais pas dépend de la combinaison des deux verbes liés
- L'interprétation est faite *a posteriori* et déterminée par la relation sémantique entre les verbes et des éléments donnés par le contexte (Halmøy, 1982)

Illustration

- (5) Les Britanniques *réagissent* **en emprisonnant** ou **en tuant** les derniers chefs.
- (6) Puis on *irrigua* les alentours **en creusant** un canal dérivé du Zab Supérieur.
- (7) Dans la ville de Koriko, Kiki, accompagné de son chat noir Jiji, va *distribuer* des colis **en volant** sur son balais.

Plan

- 1 Introduction
- 2 État de l'art
- 3 Genèse de notre travail
- 4 Expérience**
- 5 Conclusions et perspectives

Motivations

But : repérer des élaborations de manière fiable.

- Afin de contribuer au repérage automatique de cette relation
- Afin de contribuer à sa description / formalisation :
 - En collectant des exemples
 - En validant l'hypothèse de la SDRT selon laquelle elle est marquée lexicalement

Stratégie

Stratégie :

- On combine les *Voisins de Wikipédia* avec le gérondif, marqueur ambigu
- On compare la fiabilité de cette signalisation avec l'utilisation du gérondif seul

Stratégie

Deux modes de combinaison sont testés :

- GV L'*Élaboration* est détectée si le gérondif est voisin du verbe principal (8)
 - GVOV L'*Élaboration* est détectée si le gérondif est voisin du verbe principal *ET* qu'il y a au moins un lien de voisinage entre les objets des deux verbes (9)
- (8) ... et les villages *contribuaient* également à ce grand projet religieux *en envoyant* des vivres.
- (9) Les Skrulls (...) *élargissent* leur **empire** *en englobant* dans celui-ci les **mondes** moins avancés qu'ils rencontrent.

Extraction des candidats

- Corpus : Fraction de l'encyclopédie en ligne Wikipédia (version avril 2007) 45823899 mots pour 5106831 phrases
- Prétraitement par syntaxe

G	GV	GVOV
18571	375	193

- Peu de cas, mais en l'état des recherches sur la relation d'*Élaboration*, identifier un mode de signalisation fiable présente un grand intérêt, même si le nombre de cas couverts est faible.

Annotation des candidats

Première phase d'annotations :

- 314 exemples (un peu plus de 100 pour chaque cas)
- sélectionnés et présentés de manière aléatoire
- à deux annotateurs experts
- via une interface php

VOILADIS - Annotation des données

Extrait

Lors de la cérémonie de transe khlysty , les participants dansent en tournant sur eux -mêmes au rythme des cantiques .

Annotations

Le gérondif marque-t-il ici une élaboration ? Oui Non

Commentaires éventuels sur cet extrait :

Enregistrer les annotations

Annotation des candidats

- Matrice de confusion :

	elab	non elab	tot.
elab	223	21	244
non elab	13	57	70
tot.	236	78	314

- Taux d'accord : 89% (280 accords contre 34 désaccords)
- Kappa modéré à bon : 0.70
- Ce score montre la difficulté de la tâche, sans pour autant exclure la possibilité d'une automatisation

Annotation des candidats

Seconde phase d'annotation :

- Les 34 cas de désaccord sont à nouveau présentés aux annotateurs pour la discussion
- Dans 25 cas, la discussion aboutit à un accord
- Dans 9 cas, le désaccord subsiste (les annotateurs conviennent que deux interprétations concurrentes sont possibles)

(10) Le but est de ramasser le plus de pièces d'or en utilisant ses corsaires pour défendre ses vaisseaux et attaquer ceux des adversaires.

Résultats

	Extraits	Annotés	Elab.	Non Elab.	Précision	Int. conf.
G	18571	102	62	40	60.8%	9.45%
GV	375	100	81	19	81.0%	6.59%
GVOV	193	104	99	5	95.2%	2.8%

- Le gérondif est effectivement un marqueur faible
- Le nombre de candidats annotés est faible par rapport au nombre de gérondifs extraits → intervalle de confiance important
- Néanmoins, les scores obtenus avec les voisins sont suffisamment élevés pour qu'on puisse considérer qu'on a affaire à une amélioration significative
- Ces résultats sont extrêmement encourageants

Perspectives

- Résultats encore perfectibles
- Les cas dans lesquels notre signalisation a échoué ont été analysés
 - Lien de voisinage non-pertinent (11)
 - Présence d'un autre marqueur (12)

(11) Ils *marchent* la campagne en *incendiant* toutes les habitations.

(12) Le roi d'Espagne lui accorda une décoration qu'il accepta, *mais* en refusant la pension qui y était attachée.

Plan

- 1 Introduction
- 2 État de l'art
- 3 Genèse de notre travail
- 4 Expérience
- 5 Conclusions et perspectives

Conclusions générales

- Expérience pratique dédiée à la détection automatique de l'*Élaboration*
- Mise au jour d'un marqueur : <principale, gérondive> dont les verbes et objets sont voisins
- Confirmant l'hypothèse que la relation d'*Élaboration* est marquée lexicalement
- Mais rappel faible

Conclusion sur les voisins

- Importance des indices lexicaux pour la structuration du discours est communément admise
- Mais ils sont souvent négligés.
- Cette expérience valide les voisins distributionnels comme une ressource pertinente pour le repérage de la relation d'*Élaboration*

Perspectives

- Détecter les *Élaborations* entre deux phrases
- En combinant avec des marqueurs ambigus : *dans un premier temps, d'abord*
- En calculant des scores de densité de voisinage selon la position syntaxique des voisins

Bibliographie I

- Choi, Freddy Y. Y., Wiemer-hastings, Peter, & Moore, Johanna. 2001. Latent semantic analysis for text segmentation. *Pages 109–117 of : In Proceedings of EMNLP.*
- Fabre, C., & Bourigault, D. 2006. Extraction de relations sémantiques entre noms et verbes au-delà des liens morphologiques. *In : Actes de la 13^e conférence sur le Traitement Automatique de la Langue Naturelle.*
- Ferret, Olivier. 2002 (24–27 juin). Segmenter et structurer thématiquement des textes par l'utilisation conjointe de collocations et de la récurrence lexicale. *In : TALN.*

Bibliographie II

- Halliday, M. A. K., & Hasan, Ruqaiya. 1976. *Cohesion in English*.
- Hirst, G., & St-Onge, D. 1998. Lexical chains as representation of context for the detection and correction of malapropisms. *In* : Fellbaum, C. (ed), *WordNet : An Electronic Lexical Database and Some of its Applications*. Cambridge, MA : The MIT Press.
- Hoey, M. 1991. *Patterns of lexis in text*. Oxford University Press (Oxford).
- Lafourcade, Mathieu, Joubert, Alain, & Riou, Stéphane. 2009. Sens et usages d'un terme dans un réseau lexical évolutif. *In* : *TALN'09*.

Bibliographie III

- Morlane-Hondère, François. 2009. *Deux approches de la cohésion lexicale pour le repérage de phénomènes discursifs*. M.Phil. thesis, Université de Toulouse.
- Morris, J., & Hirst, G. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, **17**(1), 21–48.
- Morris, J., & Hirst, G. 2004. Non-classical lexical semantic relations. *Pages 46–51 of : Proceedings of the HLT Workshop on Computational Lexical Semantics*.