

Entités nommées et lexique-grammaire

Tita Kyriacopoulou

Plan de présentation

- 1/ Extraction d'Entités Nommées
- 2/ Ressources linguistiques
- 3/ Lexique-grammaire
- 4/ Entités nommées et lexique-grammaire
- 5/ Conclusion

Extraction d'entités nommées

Exemple : *Un raid aérien a fait au moins 11 morts et 12 blessés sur le village de Menakro le mardi 12 février*

ENTITES	SEGMENTS	Information extraite
DATE	<i>le mardi 12 février</i>	12/02/03
LIEU	<i>sur le village de Menakro</i>	Menakro
FAIT	<i>Un raid aérien</i>	Attaque militaire
IMPACT	<i>au moins 11 morts</i>	Pertes humaines (Q<50)
	<i>(au moins) 12 blessés</i>	Dommmages humains (Q<50)

Extraction d'entités nommées

LEGENDE: NUMERO de DEPECHE LIEU ETAT FAITS ORGANISATION PERSONNE COORDONNEES MOYEN OEUVRE MESURE MONNAIE NUMERO POURCENTAGE DATE DUREE FREQUENCE HEURE NON RECONNU

###1000:

Le président ivoirien **Laurent Gbagbo***, assistera au sommet des chefs d'Etat et de gouvernement d'Afrique et de France qui se tiendra du 19 au 21 février à Paris, a déclaré mardi soir à Abidjan le sénateur français **Guy Penne***, au sortir d'un entretien avec **M. Gbagbo***, jlh-ban/sd

###10066:

Selon le dernier journal de BBC-Afrique, le président **Laurent Gbagbo*** a sollicité et obtenu de l'Angola un renfort en hommes convoyé par deux avions de

###1009:

Le président ivoirien **Laurent Gbagbo*** s'est engagé vendredi devant la presse à instaurer un "cessez-le-feu total" en Côte d'Ivoire, à l'issue d'un entretien :

###1013:

Ni le général **Robert Gué***, ni l'ancien Premier ministre ivoirien **Alassane Ouattara*** "ne sont dernière" la mutinerie du 19 septembre en Côte d'Ivoire, a affirmé jeudi devant la presse à Bouaké (centre) l'adjudant **Tuo Fozie***, se présentant comme le chef des militaires rebelles qui contrôlent cette ville. eg/ban/sd

###10157:

Le président ivoirien **Laurent Gbagbo***, devrait s'adresser à la nation vendredi soir conseiller en communication de **Laurent Gbagbo***, celui-ci prononcera une allocution de ces accords et notamment l'attribution des portefeuilles de la Défense et de l'Intérieur. **France**, n'étaient que des "propositions", alors que la rébellion qui contrôle le nord

###10189:

Après une première journée où chacun a pu exprimer ses principales revendications par l'ancien ministre **Pierre Mazeaud***, a été consacrée aux questions de l'ivoirité, de la nationalité, du code foncier et du statut des immigrés. Autant de sujets jugés prioritaires par le chef de la diplomatie française, **Dominique de Villepin***. Dans la matinée, le «facilitateur» **Pierre Mazeaud*** avait rencontré un à un les chefs des dix délégations qui participent à la réunion. Hier soir le gouvernement ivoirien a accusé les rebelles de violer le cessez-le-feu dans l'ouest du pays. Les négociateurs disposent d'une semaine pour parvenir à un compromis.

###10192:

Jacques Chirac* et **Tony Blair*** ont appelé mardi les Ivoiriens à "appliquer de bonne foi" l'accord de Marcoussis, dans une déclaration commune publiée à l'issue du 25e sommet franco-britannique du Touquet (Pas-de-Calais). "Nos deux pays se félicitent de la signature par l'ensemble des forces politiques ivoiriennes" de cet accord, peut-on lire dans la déclaration. "Ils appellent tous les Ivoiriens à l'appliquer de bonne foi et réaffirment leur volonté de faire tout leur possible pour une paix durable en Côte d'Ivoire".

###10195:

Le chef de l'Etat ghanéen, également président en exercice de la Communauté économique des Etats d'Afrique de l'ouest (CEDEAO), **John Kufuor***, a quitté Accra lundi matin pour Yamoussoukro, la capitale administrative ivoirienne où il doit participer à une réunion des chefs d'Etats du "groupe de contact" de la CEDEAO sur la crise ivoirienne. Le groupe de contact de la CEDEAO comprend les chefs d'Etat et de gouvernement du Mali, du Nigéria, du Ghana, du Togo, de la Guinée-Bissau, du Niger et a été récemment élargi au Sénégal. Le Sommet, auquel participent le président **Laurent Gbagbo*** ainsi que **M. Seydou Diarra**, le Premier ministre* qu'il a désigné à la tête du futur gouvernement ivoirien de réconciliation nationale, est boycotté par les rebelles du Mouvement patriotique de Côte d'Ivoire (MPCI). **M. Diarra*** est arrivé lundi matin à Yamoussoukro, en provenance d'Accra, où il était arrivé dimanche soir pour des entretiens avec des responsables de la CEDEAO à la veille de ce mini-sommet.

Date: du 19 au 21 février
Normalisation: -02-19/-02-21

Nom: Ouattara
Prénom : Alassane
Fonction: Premier ministre
Nationalité: ivoirien

cinq mois, a annoncé jeudi l'un de ses aides. Selon **Toussaint Alain***, r les accords de Marcoussis depuis leur signature fin janvier. La teneur bo* avait toutefois déclaré que ces accords signés, sous l'égide de la

du sujet à **Marcoussis (Essonne)**. La première séance plénière, dirigée

par le chef de la diplomatie française, **Dominique de Villepin***. Dans la matinée, le «facilitateur» **Pierre Mazeaud*** avait rencontré un à un les chefs des dix délégations qui participent à la réunion. Hier soir le gouvernement ivoirien a accusé les rebelles de violer le cessez-le-feu dans l'ouest du pays. Les négociateurs disposent d'une semaine pour parvenir à un compromis.

Ressources linguistiques

- Les entrées des dictionnaires peuvent jouer plusieurs rôles dans la tâche d'extraction d'entités nommées :
 - entité nommée en tant que telle (*Napoléon, Côte d'ivoire, SNCF*)
 - mots déclencheurs présents dans le contexte immédiat d'une EN

Exemple: *Le **groupe** **Vivendi** prend le contrôle de **Neuf** **Télécom***

preuve externe preuve interne

Recherche du plus grand contexte

Principe: Elargir le contexte pour lever l'ambiguïté

Le *Quai d'Orsay se trouve dans l'impossibilité* d'affirmer que...

[Dictionnaire]

Orsay,.N+PR+Toponyme+Ville:fs

[Preuve interne]

Quai d'Orsay :lieu_micro-toponyme

[Preuve interne/externe]

Quai d'Orsay_se trouve :lieu_micro-toponyme

[Contexte éloigné]

Quai d'Orsay_se trouve dans l'impossibilité:organisation

Ressources linguistiques pour le FR

1° le système DELA qui comporte :

un lexique d'environ 90 000 mots simples ([DELAS](#)),

un lexique de plus de 100 000 noms composés ([DELAC](#) incluant des termes techniques de divers domaines) ;

2° le [lexique-grammaire](#) des phrases élémentaires du français :

15 000 types de [phrases libres](#) construites sur des verbes,

25 000 types de [phrases figées](#),

50 000 types de [phrases à verbe support de noms](#),

7 000 phrases à [adverbes figés](#) ;

3° des **grammaires locales**, représentant des phrases de domaines spécifiques :

[expressions de dates](#),

[expressions de durée](#),

[expressions de température](#),

[expression des titres : l'exemple de "ministre"](#),

[variations des cours de la Bourse](#).

Le lexique-grammaire

Dans le lexique grammaire, l'unité minimale de sens est la phrase élémentaire :

Exemple

N_0 tourner N_1 = : Pierre **a tourné** la bouteille

N_0 tourner à N_1 = : Cette affaire **a tourné** à la catastrophe

Le lexique-grammaire est organisé en un ensemble de tables, chaque table regroupant les usages des mots prédicatifs qui partagent les propriétés dites définitoires de la table.

Exemple de Table du Lexique-Grammaire

Microsoft Excel interface showing a Lexique-Grammaire table. The table columns are labeled with grammatical categories and the rows contain example entries.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
1	<ID>	N0 = Nhum	N0 = N-hum	Ppv = en fige	Ppv = y fige	Ppv = Neg	<ENT>Ppv	<ENT>V	N0 V	<ENT>Possl	autre Dat1	<ENT>C1	C1 = Npc	[passif]	N0 de V Ddef C1	Exemple												
2	1							abaisser	-	sa	-	carte maîtresse	-	+	-	§abaisser§sa carte maîtresse												
3	2							abaisser	+	ses	-	cartes	-	+	-	§abaisser§ses cartes												
4	3							abaisser	-	son	-	jeu	-	+	-	§abaisser§son jeu												
5	4							abandonner	-	sa	+	famille	-	+	-	§abandonner§sa famille												
6	5							abandonner	-	sa	+	famille	-	+	-	§abandonner§sa famille												
7	6							abandonner	-	son	+	nom de jeune fille	-	+	-	§abandonner§son nom de jeune fille												
8	7							abandonner	-	ses	+	pénates	-	+	-	§abandonner§ses pénates												
9	8							abandonner	-	son	-	poste	-	-	-	§abandonner§son poste												
10	9							abattre	-	ses	-	atouts	-	+	-	§abattre§ses atouts												
11	10							abattre	-	sa	-	carte maîtresse	-	+	-	§abattre§sa carte maîtresse												
12	11							abattre	-	ses	-	cartes	-	+	-	§abattre§ses cartes												
13	12							abattre	-	son	-	jeu	-	+	-	§abattre§son jeu												
14	13							abjurer	-	sa	+	foi	-	+	-	§abjurer§sa foi												
15	14							accélérer	-	ses	-	enjambées	-	-	-	§accélérer§ses enjambées												
16	15							accélérer	+	son	-	train	-	+	-	§accélérer§son train												
17	16							accomplir	-	son	-	devoir	-	+	-	§accomplir§son devoir												
18	17							accroître	-	son	+	capital	-	+	-	§accroître§son capital												
19	18							accroître	-	sa	+	part du marché	-	+	-	§accroître§sa part du marché												
20	19							accuser	-	son	-	âge	-	-	-	§accuser§son âge												
21	20							affûter	-	ses	+	arguments	-	+	-	§affûter§ses arguments												
22	21							aggraver	-	son	-	cas	-	+	-	§aggraver§son cas												
23	22							aimer	-	ses	-	aises	-	-	-	§aimer§ses aises												
24	23							aller	-	son petit	-	bonhomme de chemin	-	-	-	§aller§son petit bonhomme de chemin												
25	24							aller	-	son	-	chemin	-	-	-	§aller§son chemin												
26	25							aller	-	son	-	train	-	-	-	§aller§son train												
27	26							aller	-	son petit	-	train	-	-	-	§aller§son petit train												
28	27							aller	-	son	+	train de sénateur	-	-	-	§aller§son train de sénateur												
29	28							allonger	-	ses	-	enjambées	-	-	-	§allonger§ses enjambées												
30	29							alumer	-	ses	-	quinquets	-	+	-	§alumer§ses quinquets												
31	30							améliorer	-	son	-	temps	-	+	-	§améliorer§son temps												
32	31							amener	-	sa	-	barbaque	-	+	-	§amener§sa barbaque												
33	32							amener	-	sa	-	bidoche	-	+	-	§amener§sa bidoche												
34	33							amener	-	sa	-	graisse	-	+	-	§amener§sa graisse												
35	34							amener	-	sa	-	viande	-	+	-	§amener§sa viande												
36	35							apprendre	-	ses	-	classiques	-	-	-	§apprendre§ses classiques												
37	36							appuyer	-	ses	+	coups	-	+	-	§appuyer§ses coups												

Toutes les entrées d'une table ont en commun la structure de base. Par exemple la table 8 décrit les verbes avec un complément (nominal ou phrastique) introduit par la préposition "de" (Exemple : *Jean se repent de sa conduite*)

Pour chaque lemme d'une table, les colonnes indiquent des propriétés et en particulier des informations sur :

- les réalisations possibles des arguments (catégorie, préposition, etc.);
- les propriétés syntaxiques du verbe ou de ses arguments (pronominalisation, etc.);
- les sous-catégorisations alternatives;

Le lexique-grammaire montre qu'il n'existe pas deux éléments ayant le même comportement syntaxique, d'où:

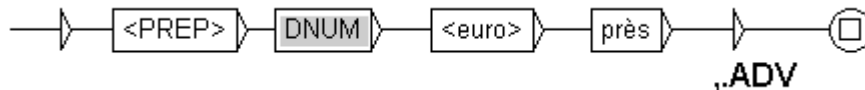
- il est impossible d'établir des règles générales qui expliquent la langue,
- on doit accumuler des descriptions de phénomènes particuliers.

Entités nommées et Lexique-Grammaire

- Nécessité des entités nommées pour rendre opérationnel le lexique-grammaire.
- Besoin du lexique-grammaire pour mettre en relation les entités nommées et affiner l'annotation et l'extraction

Entités nommées => lexique-grammaire

- Génération de toutes les variantes/variations en prenant en compte des variables (le lexique-grammaire ne définit pas ses variables) qui font appel :
 - à un dictionnaire : *à la grande surprise de Nhum*
 - à un sous-graphe :



Entités nommées => lexique-grammaire

Instancier les arguments N0, N1,N2

- Entités Nommées : *Nicolas Sarkozy*
- Entités « non Nommées » (*L'ancien Président de la République*)
- Groupe Nominal incluant une entité nommée (*La femme de Sarkozy*)
 - Prendre en compte la coordination
 - Paul et Marie*
 - Paul et lui*
 - Andrew Stuart et ses collègues*

Lexique-grammaire => entités nommées

- Elargir le contexte : *Quai d'Orsay*
- Restreindre le focus sémantique des compléments essentiels (Watrin 2003), pour établir des patrons d'extraction :

Table 6 : N0 hum V (*La France déclare*)

Table 36DT (verbes dits datifs de structure N0 V N1 à N2)

N0	V	N1	N2	Dnum	Advp
<vendeur>	offrir	<objet>	<acquéreur>	-	-
<vendeur>	payer	<objet>	<acquéreur>	<montant>	-
<acquéreur>	racheter	<objet>	<vendeur>	<montant>	<montant>

Propriétés : N0 humain
 N2 humain
 N2 bénéficiaire

Cas d'ambiguïtés non résolus

- « *Air Liquid* » est une compagnie de gaz.
- « *Air Mini* » qui est un smartphone Apple et non une compagnie aérienne.
- « *LA COMPAGNIE DE BEL AIR : Une parfaite connaissance des marchés visés et une maîtrise de savoir-faire spécifiques* »
Bel air est un camping et non une compagnie aérienne.

Conclusion

- Méthode symbolique intégration future des statistiques (Treecloud)
- Analyse quantitative mais surtout qualitative
- Utilisation et récupération des ressources existantes
- Exploitation approfondie des tables du lexique-grammaire
- Analyseur syntaxique

MERCI