

# Informations Biographiques Extraites à l'aide de Contextes Observés Linguistiquement



## 1 Extraction d'informations biographiques dans les textes de la presse financière

Nous définissons une information biographique comme une relation prédicative à plusieurs arguments dont l'un est nécessairement une entité de la classe d'objet <Personne>.

Les autres actants intervenant dans un événement biographiques appartiennent typiquement à des classes d'objets bien définies telles que les classes <Personne>, <Lieu>, <Date>, <Organisation>, <Secteur d'Activité>, ou <Profession>.

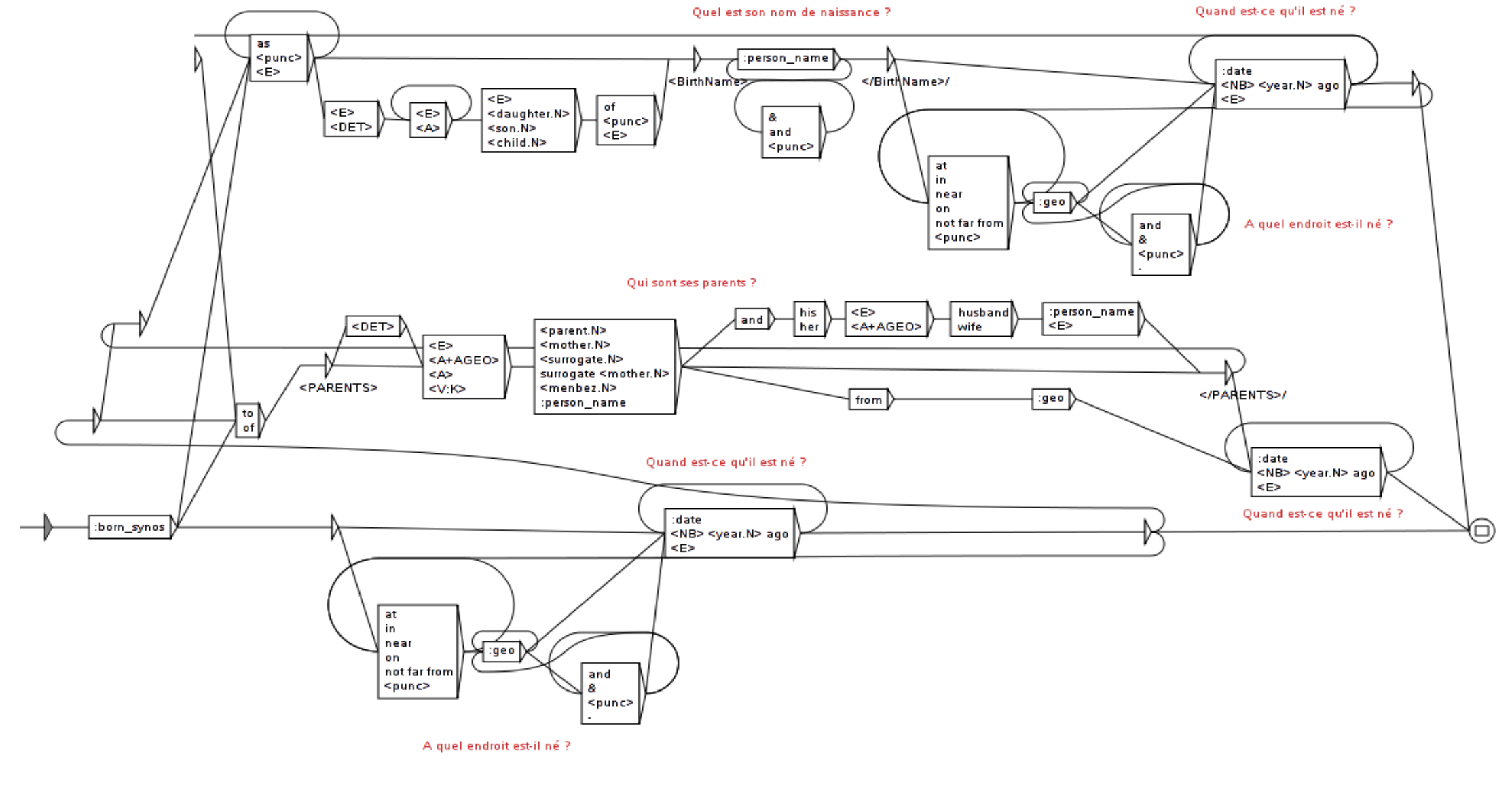
## 2 12 types d'événements biographiques étudiés

- |   |   |
|---|---|
| <b>6 événements personnels :</b><br>- la naissance<br>- l'enfance<br>- la formation<br>- le mariage<br>- le divorce<br>- le décès | <b>6 événements professionnels :</b><br>- l'obtention d'un emploi<br>- l'occupation d'un poste<br>- le licenciement<br>- la succession<br>- la démission<br>- le départ en retraite |
|---|---|

## 3 Hypothèse du sous-langage

Notre hypothèse de départ est que l'ensemble des énoncés décrivant l'un de ces 12 types d'événement biographique se caractérise par un lexique de taille finie et un nombre de schémas de phrases limité.

En ce sens, nous considérons que ces énoncés constituent un sous-langage qu'il nous est possible de décrire de manière exhaustive à l'aide de grammaires locales et de lexiques terminologiques.



## 4 Effectifs des différentes classes d'objets

Classe d'objet	Sous-classe d'objet	Balance sémantique	Nombre d'instances	Exemples pour des instances de classe d'objet
Nomme Proprie	Titre	(Title)	370	Queen, Lord, PhD, Mr
	Prénom	(FirstName)	38 500	Lara, Marie-Luise, Ben
	Nom de Famille	(Surname)	1 250 000	Obama-Smith, Vildic
	Nom de Personne (Complet)	(LongName)	8 300 000	Hennis Nordqvist
Rôle Social	Famille	(Human)	6 400	daughter, son, aunt
	Profession	(JobDescriptor)	45 000	cook, kitchen helper
	Habitant	(Citizen)	600	Auckland, Brooklyn
Secteur d'activité	Matière	(Discipline)	580	art history
	Branche	(Sector)	38 000	life insurance, farming
Organisation	Type d'organisation	(CompanyDescriptor)	23 800	car manufacturer
	Nom d'organisation	(Company)	516 000	Fujitsu Siemens
	Forme juridique d'entreprise	(LegalForm)	115	ltd, inc, plc, AG, GmbH, Ev., S.A., s.a.r.l., LLC
Lieu	Païs et Continent	(Nation) (Continent)	430	South America, France, Germany, Europe
	Ville	(City)	327 400	Gravenhage, Paris
	Date	Mois	(Month)	24
	Jour de semaine	(DayOfWeek)	7	Saturday, Sunday

Plus de 10 millions d'unités lexicales au total dont environ 15% ont été découvertes à l'aide de méthodes de *bootstrapping*.

## 5 Extraction d'informations biographiques et remplissage de formulaires

1st Filipino bishop {S}

Novo Ecijano becomes first Filipino bishop in Los Angeles. {S}

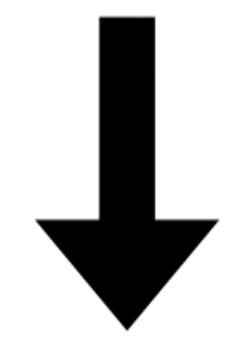
Oscar Azarcon Solis, a 50-year-old Philippine-born prelate, was appointed by the Vatican in early December as the first Filipino-American Catholic bishop in America. {S}

Solis was born in San Jose, Nueva Ecija and once served as parish priest of Cabanatuan City. {S}

Currently a pastor of the St. Joseph Co-Cathedral in Thibodaux, Louisiana, Solis will be helping oversee the Archdiocese of Los Angeles, California, after February. {S}

As bishop of the Titular See of Urcl, Solis will assist Cardinal Roger Mahony administer the second largest Catholic archdiocese in the US with five million members, including several thousand Filipinos. {S}

The appointment of Solis was welcome news to Filipino-American Catholics in North America who are smarting from the scandal arising from a number of sexual molestation charges brought against a number of Philippine-born priests in California. {S}



person name	career	job	organization	location	date
Oscar Azarcon Solis		prelate		Philippines	
Novo Ecijano, Solis		parish priest		Cabanatuan City	
50 (2004)		pastor	St. Joseph Co-Cathedral	Thibodaux, Louisiana	
Philippine (Novo Ecijano)		first Filipino-American Catholic bishop	Vatican	America	December (2003)
San Jose, Nueva Ecija		prelate; parish priest; pastor; bishop		Los Angeles	
Roger Mahony		helping oversee the Archdiocese of Los Angeles	Cardinal Roger Mahony	Los Angeles, California	after February (2004)
cardinal		assist			
		work			
		administer the second largest Catholic archdiocese		US	

## 6 Évaluation

Type d'événement	Nombre d'événements	Nombre d'arguments	Annotations correctes	Erreurs d'annotation	Précision	Rappel
Naissance	291	583	571	0	100 %	97.9 %
Enfance	1	2	2	0	100 %	100 %
Formation	55	222	213	9	95.8 %	89.2 %
Mariage	12	28	27	0	100 %	96.4 %
Divorce	2	2	2	0	100 %	100 %
Décès	122	254	236	2	99.2 %	92.1 %
Obtention d'un emploi	68	287	273	14	94.9 %	74.2 %
Occupation d'un poste	147	448	431	26	94.0 %	78.6 %
Licenciement	4	9	9	0	100 %	100 %
Succession	2	7	5	0	100 %	71.4 %
Démission	3	6	5	0	100 %	83.3 %
Départ en retraite	3	6	6	0	100 %	100 %
<b>Total</b>	<b>710</b>	<b>1854</b>	<b>1780</b>	<b>51</b>	<b>97.1 %</b>	<b>89.8 %</b>

Evaluation effectuée sur un corpus de 150 courtes biographies extraites depuis le site <http://www.biography.com>.

## 7 Conclusion et perspectives

- Extension de la description à de nouveaux types d'événements biographiques
- Analyse de nouveaux genres textuels
- Résolution des coréférences
- Génération automatique de la biographie complète d'une personne en compilant les informations extraites depuis différents documents.

