

Analyse sémantique automatique

Pascal Amsili / Marie Candito
12-13

Semantic role labeling (M Candito)

1

Contenu

- **Sémantique lexicale (en bref)**
 - unité lexicale
 - signifiants et ambiguïté
 - relations lexicales
 - représentation du sens d'une UL
 - rôles sémantiques
- **Sémantique lexicale computationnelle**
 - ressources pour l'anglais
 - réseau lexical : Wordnet
 - classes sémantiques : VerbNet
 - rôles sémantiques : FrameNet, PropBank
 - pour le français?
 - tâches
 - Calcul de similarité lexicale
 - **WSD** : désambiguïssation lexicale (word sense disambiguation)
 - **SRL** : étiquetage de rôles sémantiques (semantic role labeling)
 - **RTE** : Reconnaissance d'inférence textuelle (recognizing textual entailment)
 - Lexical Acquisition
 - acquisition de cadres de sous-catégorisation
 - acquisition de préférences lexicales
 - reconnaissance de composés (multi-word units)

2

Plan du cours

- Introduction
- Rôles sémantiques
 - définition et perspective historique
 - une liste de rôles, et ses inconvénients
 - exemple de défis d'une théorie des rôles sémantiques
 - quelques théories de rôles sémantiques
 - avec ou sans classes sémantiques de prédicats
- Ressources annotées
- Etiquetage en rôles sémantiques
 - définition
 - applications
 - systèmes de SRL
 - campagnes de SRL (shared tasks)

3

Introduction

- L'étiquetage en rôles sémantiques utilise une notion extrêmement débattue en **linguistique** (pure)
 - les « **rôles sémantiques** »
- et est dans le même temps une tâche de **TAL** par excellence
 - cf. il s'agit de repérer dans un texte:

« **qui** a fait **quoi** **où** **quand** **comment** et **pourquoi** ? »

4

Sources

- Palmer, Gildea and Sue, 2010, « Semantic role labeling », Morgan & Claypool
- Computational Linguistics, special issue on Semantic role labelling, 2008
 - Marquez, Carreras, Litkowski et Stevenson, 2008, introduction to the special issue
- Tutoriel SRL de Lluís Marquez à ACL 2009
- Tutoriel SRL de Yih et Toutanova à NAACL 2006

- Kasper, 2008 : *A comparison of thematic role theories*, Marburg University

5

Rôles sémantiques: un concept ancien

- au XX^{ème} siècle, énorme littérature:
 - pionniers : Gruber 65, Fillmore 68 (case relations)
 - Jackendoff 70, 72
 - Dowty 90 (proto-roles) ...

- mais on peut en voir la trace dans la grammaire de Panini (6^{ème} siècle AC !)
 - rôles « Karaka »
 - permettant de classer les participants d'un événement, définition sémantique
 - ces rôles sont des caractéristiques de classes de verbes
 - cas morphologiques
 - règles de correspondances entre cas et rôles

6

Rôles sémantiques

- une même notion sous des termes variés
 - Gruber, Jackendoff « relations thématiques »
 - Fillmore « case relations » ou « deep cases »
 - rôles sémantique, rôle thématique
- rôles sémantiques =
 - **généralisations** sur les arguments d'un prédicat en vue de capturer des régularités dans l'interprétation (sémantique) des représentations syntaxiques
- c'est donc une notion qui relève de **l'interface syntaxe-sémantique**

7

Fonctions grammaticales et rôles

- une même FG peut correspondre à +sieurs rôles
 - *Paul dort*
 - *Paul mange une pomme*
 - *Paul reçoit un colis*
 - *Paul plaît à Marie*
 - *Le tiroir contient des babioles*
- et inversement un même rôle => +sieurs FG
 - cf. les **changements de diathèse**
 - mais aussi
 - **Paul** reçoit un colis de Pierre
 - Pierre donne un colis à **Paul**
 - Pierre fournit **Paul** en cigarettes
- Les rôles sémantiques sont sensés permettre de mieux expliquer le comportement lexical de verbes que la simple sous-catégorisation syntaxique

8

Linking

- il existe cependant **de très fortes régularités** dans la **correspondance entre arguments sémantiques et arguments syntaxiques**
 - « manger » : sujet = mangeur, objet = ce qui est mangé
 - « lire » : sujet = liseur, objet = lu
 - pour bcp de V, le « V-eur » est le sujet, et le « V-é » est l'objet
- une théorie linguistique se doit d'en rendre compte
- => théorie du « **linking** » :
 - modélisation correspondance argts sémantiques / argts syntaxiques
- mais attention
 - l'histoire a montré qu'une simple liste de rôles sémantiques ne suffit pas à expliquer cette correspondance
 - la réalisation syntaxique n'est explicable que par une représentation plus profonde de la sémantique lexicale du verbe
 - en outre, des **irrégularités** existent
 - se rappeler qqch / se souvenir de qqch

9

Rôles sémantiques : structure

- liste à plat de rôles discrets
 - Gruber, Fillmore (Case Grammar), LFG à ses débuts, HPSG...
- hiérarchie de rôles
 - Jackendoff, Fillmore (Construction Grammar), Optimality theory, ...
- rôles généralisés
 - proto-roles de Dowty 89, 91

10

Rôles sémantiques : desiderata

- utilisés dans différentes théories linguistiques
- ces théories visent en général à ce que les rôles respectent les caractéristiques :
 1. définition opératoire, **indépendante** du prédicat
 2. **complétude/unicité** : tout argument de prédicat reçoit exactement un rôle
 3. **distinction** : aucun argument ne reçoit le même rôle qu'un autre
 - ou le même ens. de rôles qu'un autre
- => ces 3 propriétés sont trop fortes,
 - aucune liste de rôles n'a pu être établie qui permette
 - d'assigner des rôles à un nombre représentatif de prédicats d'une langue
 - en respectant ces propriétés

11

Une liste de rôles ≈ consensuelle (Palmer et al. 09, repris de Saeed, 2003)

Role	Description	Examples
Agent	Initiator of action, capable of volition	The batter smashed the pitch into left field. The pilot landed the plane as lightly as a feather.
Patient	Affected by action, undergoes change of state	David trimmed his beard . John broke the window .
Theme	Entity moving, or being "located"	Paola threw the Frisbee . The picture hangs above the fireplace.
Experiencer	Perceives action but not in control	He tasted the delicate flavor of the baby lettuce. Chris noticed the cat slip through the partially open door.
Beneficiary	For whose benefit action is performed	He sliced me a large chunk of prime rib, and I could hardly wait to sit down to start in on it. The Smiths rented an apartment for their son .
Instrument	Intermediary/means used to perform an action	He shot the wounded buffalo with a rifle . The surgeon performed the incision with a scalpel .
Location	Place of object or action	There are some real monsters hiding in the anxiety closet . The band played on the stage .
Source	Starting point	The jet took off from Nairobi . We heard the rumor from a friend .
Goal	Ending point	The ball rolled to the other end of the hall . Laura lectured to the class .

12

De l'impossibilité d'arrêter une liste

- des tests ont été proposés **pour identifier les rôles en contexte**
 - exemple: AGENT : possibilités d'ajouts insistant sur la volition (pour Vinf..., délibérément, exprès...)
 - *Jean a acheté le livre à Paul pour le lire* **#Jean a reçu le livre de Paul pour le lire**

- rôles supplémentaires éventuels?
 - **RECIPIENT**
 - *il a légué sa fortune à **une association*** => ou **GOAL? BENEFICIARY?**
 - **PERCEPT** : entité perçue
 - *Le général a inspecté **les troupes*** *Avez-vous entendu **le tonnerre?***
 - *Le requin a effrayé les baigneurs*
 - **Quel rôle pour :**
 - *J'ai vendu ma moto à Paul **pour 1000 euros*** *Paul a payé **1000 euros** pour une croûte*
 - *le chien a léché **ma main*** => changement d'état ou pas?
 - *les fidèles ont enduit **l'icône** d'huile* => changement d'état ou pas?

- où s'arrêter?
 - => **il n'existe aujourd'hui aucune liste de rôles + critères satisfaisante**

13

Unicité de rôle?

- unicité de rôle pour un argument s'avère une propriété trop forte
 - **Jean a sauté par la fenêtre**
 - AGENT ou THEME ?
 - unicité de rôle pour un prédicat également
 - *Jean a rencontré Paul hier par hasard.*
 - rôle de Jean et Paul différent?
 - *Jean et Paul se sont rencontrés hier par hasard.*
 - remis en cause par différents auteurs
 - Dowty (cf. infra)
 - ou Jackendoff (90) qui propose +sieurs niveaux de rôles
 - rôles de type spatiaux
 - rôles de type « action » (tous les autres...)
- | | | |
|----------------------------------|---------|----------------------|
| <i>Jean a frappé Pierre</i> | | |
| THEME | BUT | <i>thematic tier</i> |
| ACTOR | PATIENT | <i>action tier</i> |
| | | |
| <i>Marie est arrivée à Paris</i> | | |
| THEME | BUT | <i>thematic tier</i> |
| ACTOR | | <i>action tier</i> |
| | | |
| <i>Marie a reçu une lettre</i> | | |
| GOAL | THEME | <i>thematic tier</i> |

14

Agentivité

- l'entité qui a un effet sur une autre n'est pas forcément volontaire, ni consciente
- d'où de nombreuses propositions de raffinements du rôle Agent
- par ex. Cruse, 73: agentivité représentée par 4 propriétés, non mutuellement exclusives
 - **volitif** : implique un acte volontaire
 - *Paul observe la scène*
 - mais aussi (!!):
 - *Le capitaine est mort pour sauver ses camarades*
 - *Le blessé restait immobile pour éviter d'être découvert*
 - **initiative** : initiation d'une action via une commande
 - *le garde mena les prisonniers de l'autre côté de la cour*
 - **agentif**: action d'un objet (vivant, machine, force naturelle) réalisée au moyen de son énergie propre
 - *la machine a broyé l'épave*
 - *la crue a englouti le village*
 - **effectif** : exercice d'une force du fait de sa position ou mouvement
 - *le pilier soutient le bâtiment*
 - *en tombant, l'arbre a écrasé la voiture*

15

Défi pour toute théorie du linking via rôles : les V psychologiques

- V mettant en jeu
 - un « stimulus » et un être conscient, qui reçoit/fait l'expérience de ce stimulus : un « experiercer »
- Ces V se classent en 2 linkings typiques:

Experiercer = sujet Stimulus = objet direct ou oblique	Stimulus = sujet Experiercer = objet direct ou oblique
x aime y	y plaît à x
x craint y	y effraie x
x considère y comme ...	y frappe x comme étant ...

- Comment expliquer ces linkings différents pour apparemment les mêmes rôles ?

16

Défi pour toute théorie du linking via rôles: l'alternance locative

1. *Marie charge des oranges dans le camion*
 2. *Marie charge le camion d'oranges*
 - propriétés aspectuelles varient:
 - camion totalement chargé en (2), pas totalement en (1)
 - sens légèrement différent, représentable par gloses différentes:
 - 1: X cause que Y va sur/dans Z
 - 2: X cause que Z change d'état en déplaçant Y dans/sur Z
- y a-t-il polysémie?
 - psycho-linguistique tend à montrer qu'il s'agit d'un seul V
 - mais alors 1 verbe avec des rôles différents?
 1. *Marie charge [des oranges THEME] [dans le camion LOC]*
 2. *Marie charge [le camion PATIENT] [d'oranges THEME]*

17

Défi pour toute théorie du linking via rôles: Verbes converses

1. *Marie vend une voiture à Jean*
2. *Jean achète une voiture à Marie*
 - même situation
 - comment expliquer les linkings différents?
 - difficile d'arguer pour plus d'agentivité dans l'un ou l'autre des cas
 - cf. le vendeur et l'acheteur ont des actes volontaires à accomplir pour que la transaction est lieu

18

Plan du cours

- Introduction
- Rôles sémantiques
 - définition et perspective historique
 - une liste de rôles, et ses inconvénients
 - exemple de défis d'une théorie des rôles sémantiques
 - quelques théories de rôles sémantiques
 - avec ou sans classes sémantiques de prédicats
- Ressources annotées
- Etiquetage en rôles sémantiques
 - définition
 - applications
 - systèmes de SRL
 - campagnes de SRL (shared tasks)

19

Quelques théories de rôles

- sans classification sémantique des prédicats
 - Jackendoff, 72, 83, 92 : **Lexical Conceptual Structures**
 - Dowty, 91: **Proto-roles**
- avec classification sémantique de prédicats
 - Levin, 93 : classes verbales et alternances
 - Fillmore, 68 ... : **Frame Semantics**

20

LCS: Lexical Conceptual Structures

- liste réduite de foncteurs conceptuels
 - considérés comme des primitives
 - GO, CAUSE, BE ...
 - n'ayant pas vocation à représenter toute la sémantique, mais seulement les concepts clés importants pour l'interface syntaxe-sémantique
- les rôles sont présentés comme les arguments de ces foncteurs

Marie a couru de la mairie à la poste.

GO (Marie, mairie, poste)

Marie emmène Paul de la mairie à la poste.

CAUSE(Marie, GO (Paul, mairie, poste))

21

LCS

- les primitives ont été raffinées par la suite
 - GO_{POSIT}: changement de lieu
 - GO_{POSS}: changement de propriété
 - GO_{IDENT}: changement d'état

L'oiseau s'enfuit de la cage GO_{POSIT}(oiseau, cage, x)

Marie prend une fleur à Jean CAUSE(
Marie,
GO_{POSS}(fleur, Jean, Marie)

Marie vend une fleur à Jean CAUSE(
Marie,
GO_{POSS}(fleur, Marie, Jean))

22

LCS : Linking hiérarchique

- Jackendoff 92 définit
 - une hiérarchie de rôles sémantiques (pour LCS=arguments des primitives conceptuelles)
 - une hiérarchie de fonctions syntaxiques
 - le linking se faisant d'après ces 2 hiérarchies
 - rôle sém le plus haut réalisé syntaxiquement par la position syntaxique la plus haute disponible
 - eg: causateur => toujours sujet
- échec du passage à l'échelle :
 - l'extension à d'autres verbes (en particulier plus abstrait) et à d'autres langues a échoué

23

Les proto-rôles de Dowty

- notion de prototype
 - mot « panthère » : dénote une classe de félins, pelage noir, 4 pattes, court vite, vit dans la jungle...
 - quid de panthères de zoo? de panthères ayant perdu une patte?
 - plus qu'une définition, les caractéristiques ci-dessus constituent le **prototype** de ce qu'est une panthère
- même principe appliqué aux rôles:
 - prototype d'Agent : être conscient causant volontairement un changement d'état d'un autre participant...
 - mais tous les Agents ne vérifient pas forcément cet archétype
- => Dowty préfère décomposer les rôles traditionnels en un ensemble d'implications
 - (propriétés sur l'argument, découlant de la sémantique du prédicat)
 - pas forcément toutes vérifiées

24

Les proto-roles de Dowty (2)

- Dowty nomme « proto-roles » ces ens. d'implications
- un argument donné d'un prédicat donné
 - peut porter un ou plusieurs de ces proto-roles à **des degrés divers**
 - ⇔ satisfait un nb + ou – important d'implications définitoires des proto-roles
- 2 proto-roles principaux pour une théorie du linking:
 - proto-agent => plutôt sujet
 - proto-patient => plutôt objet direct (ou oblique)

25

Proto-agent

- **propriétés définitoires, et exemple de sujets n'exhibant qu'une seule de ces propriétés**
 - a. implication volontaire dans l'événement ou état
 - *Marie ignore Paul délibérément*
 - b. conscience (et/ou perception) de l'évnt ou état
 - *Marie sait / croit / est déçue par ce fait*
 - *Marie voit / craint Paul*
 - c. cause un évt ou chgt d'état pour un autre participant
 - *la misère engendre la délinquance*
 - d. mouvement (relatif à la position d'un autre participant s'il existe)
 - *le pneu déboulonné par l'accident dépassa le rocher*
 - *L'eau remplit le bateau*
 - *Marie est tombée*
 - e. (existe indépendamment de l'événement)
 - *Marie a besoin d'une nouvelle voiture*
- rôles traditionnels typiquement correspondants:
 - agent, experierer, instrument, cause ...

26

Proto-patient

- **propriétés définitives, et exemple d'objets directs ayant ces propriétés**
 - a. subit un changement d'état
 - *Marie a déplacé **le rocher*** (déplacement indéfini, donc pas proto-agent d.)
 - *Marie mange **une pomme*** (satisfait b. et c.)
 - b. thème incrémental
 - *Marie traversa **la rivière** / remplit le verre* (satisfait aussi d.)
 - *Marie mange **une pomme*** (satisfait aussi a.)
 - c. affecté d'une manière causée par un autre participant
 - *Fumer provoque **le cancer du poumon***
 - d. stationnaire relativement au mouvement d'un autre participant
 - *la balle dépassa **la flèche***
 - e. (existence dépendante de l'évt)
 - *Marie a besoin **d'une nouvelle voiture***
 - *Marie construit **une maison** / effaça **ses notes*** (satisfait aussi a.)
- rôles traditionnels typiquement correspondants:
 - patient, theme, percepts ...

27

Linking de Dowty (Dowty 91, p. 576)

- (31) ARGUMENT SELECTION PRINCIPLE: In predicates with grammatical subject and object,¹⁷ the argument for which the predicate entails the greatest number of Proto-Agent properties will be lexicalized as the subject of the predicate; the argument having the greatest number of Proto-Patient entailments will be lexicalized as the direct object.
- (32) COROLLARY 1: If two arguments of a relation have (approximately) equal numbers of entailed Proto-Agent and Proto-Patient properties, then either or both may be lexicalized as the subject (and similarly for objects).
- (33) COROLLARY 2: With a three-place predicate, the nonsubject argument having the greater number of entailed Proto-Patient properties will be lexicalized as the direct object and the nonsubject argument having fewer entailed Proto-Patient properties will be lexicalized as an oblique or prepositional object (and if two nonsubject arguments have approximately equal numbers of entailed P-Patient properties, either or both may be lexicalized as direct object).
- (34) NONDISCRETENESS: Proto-roles, obviously, do not classify arguments exhaustively (some arguments have neither role) or uniquely (some arguments may share the same role) or discretely (some arguments could qualify partially but equally for both proto-roles).

28

Dowty : Verbes converses

- Dowty ne cherche pas à expliquer la différence de sujet entre
 - x vend y à z
 - z vend y à x
 - x loue y à z / z loue y à x
- pour lui, poser des rôles différents pour x et z dans ces exemples est contre-productif, cf. manifeste un manque de généralité
- mais son corollaire 2 prédit quel argument est objet versus a-objet, une fois le sujet fixé

29

Dowty : V psychologiques

Experier = sujet Stimulus = objet direct ou oblique	Stimulus = sujet Experier = objet direct ou oblique
x aime y	y plaît à x
x craint y	y effraie x
x considère y comme ...	y frappe x comme étant ...

- dans les 2 cas, x et y ont le même nb de propriétés proto-a et proto-p
 - lesquelles?
 - mais les versions où le Stimulus est sujet comportent un cht d'état pour l'objet / l'oblique x
 - => ce qui donne une propriété de proto-patient supplémentaire au x, et donc le Stimulus l'emporte pour être sujet

30

Dowty: l'alternance locative

1. *Marie charge des oranges dans le camion*
2. *Marie charge le camion d'oranges*
 - propriétés aspectuelles varient:
 - camion totalement chargé en (2), pas totalement en (1)
 - sens légèrement différent, représentable par gloses différentes:
 - 1: X cause que Y va sur/dans Z
 - 2: X cause que Z change d'état en déplaçant Y dans/sur Z
- Dowty réfute cette distinction location vs theme
 - en (2) camion est thème incrémental => donc objet (corollaire 2)
 - en (1), c'est plutôt les oranges qui sont totalement chargées => thème incrémental => objet

31

Quelques théories de rôles

- sans classification sémantique des prédicats
 - Jackendoff, 72, 83, 92 : Lexical Conceptual Structures
 - Dowty, 91: Proto-roles
- avec classification sémantique de prédicats
 - Levin, 93 : classes verbales et alternances
 - Fillmore, 68 ... : Frame Semantics

32

2 approches pour la classification sémantique de prédicats

- Classification sémantique de prédicats **en fonction de leur comportement syntaxique**
 - => partage de **comportement syntaxique fin**
 - => partage de **propriétés sémantiques grossières** uniquement
 - ressources
 - anglais : Levin, 93 => ressource VerbNet (Kipper)
 - français : LVF Dubois et Dubois-Charlier
- Classification « purement » sémantique
 - => partage d'une **sémantique fine**
 - => au **comportement syntaxique potentiellement varié**
 - Wordnet
 - Frame Semantics
 - on part de situations prototypiques relevant de la connaissance du monde qu'a un locuteur, indispensable à la compréhension d'énoncés
 - regroupement de lexèmes pouvant évoquer le même frame
 - ressource
 - FrameNet

33

Classes sémantiques de Levin (93)

- classification de verbes d'après les alternances de linking qu'ils peuvent montrer
 - 3100 V anglais, 47 classes au 1^{er} niveau, 193 au 2nd et 3^{ème} niveau
 - les alternances définissent une signature syntaxique
 - chaque classe est par ailleurs associée à une sémantique grossière
 - => le postulat est en effet que les alternances sont permises / interdites d'après la sémantique du V

34

Classes sémantiques de Levin

- exemples (extrait de Palmer et al. 09)

Class	Syntactic Frames	Semantic Components
break 45.1	John broke the jar. / The jar broke. / Jars break easily.	change-of-state
cut 21.1	John cut the bread. / *The bread cut. / Bread cuts easily.	change-of-state, recognizable action, sharp instrument
hit 18.1	John hit the wall. / *The wall hit. / *Walls hit easily.	contact, exertion of force

- rem:
 - couverture limitée :
 - 3000 verbes, et seulement leurs sens les plus proéminents
 - théoriquement plus problématique:
 - certaines classes partagent la majeure partie des alternances

35

Ressource dérivée : Verbnet

- Kipper et al. 2000
 - ressource disponible en ligne
 - <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>
- extension des classes de Levin
 - + 1000 verbes
 - + 200 classes
 - ajout de rôles sémantiques explicites (liste de 24 rôles)

36

Pour le français...

- lexique-grammaire
 - classification des V d'après leurs propriétés syntaxiques fines
 - pas uniquement les alternances
 - revendique comme non sémantique
 - dans les faits, on retrouve des généralisations sémantiques
- LVF: Lexique des verbes français
 - Dubois et Dubois-Charlier
 - même idée de classification d'après propriétés syntaxiques
 - mais volonté d'obtenir une classification sémantique
 - 14 classes de niveau 1

37

2 approches pour la classification sémantique de prédicats

- Classification sémantique de prédicats **en fonction de leur comportement syntaxique**
 - => partage de **comportement syntaxique fin**
 - => partage de **propriétés sémantiques grossières** uniquement
 - ressources
 - anglais : Levin, 93 => ressource VerbNet (Kipper)
 - français : LVF Dubois et Dubois-Charlier
- Classification « purement » sémantique
 - => partage d'une **sémantique fine**
 - => au **comportement syntaxique potentiellement varié**
 - Wordnet
 - Frame Semantics
 - on part de situations prototypiques relevant de la connaissance du monde qu'a un locuteur, indispensable à la compréhension d'énoncés
 - regroupement de lexèmes pouvant évoquer le même frame
 - ressource
 - FrameNet

38

Frame Semantics

- Fillmore, 76 et les 30 années suivantes
- sémantique empirique mettant en avant la continuité entre langue et connaissance/ expérience du monde
 - sens de « *pension alimentaire* »
 - référence à 2 personnes ayant été mariées, ne l'étant plus...
 - sens du mot « *petit-déjeuner* » :
 - fait appel aux coutumes d'un pays
 - sens **prototypique**: collation du matin
 - mais aussi dormir jusqu'à 13h et petit-déjeuner à 14h
 - nuit blanche et petit-déjeuner à 6h
 - etc...

39

Frame Semantics (2)

- frame = situation prototypique, évoquée par lexèmes
 - *vendre, vente, acheter, achat, payer, facturer* : font appel à une même situation « transaction commerciale »
 - comprendre ces mots suppose de connaître la coutume de transaction commerciale
- chaque frame peut être décrite par une définition et une description précise de la sémantique des participants à la situation
 - leur **rôle sémantique**
 - appelé « élément de frame » pour insister sur le fait qu'ils sont constitutifs de la définition d'une frame
 - exemple:
 - **INGESTION** : An Ingestor consumes food or drink (Ingestibles), which entails putting the Ingestibles in the mouth for delivery to the digestive system. This may include the use of an Instrument.

40

Frame Semantics (3)

- NB: La Frame semantics ne propose pas de théorie du linking
 - mais volonté de confrontation empirique aux données, à grande échelle
- => projet **FrameNet**
 - +de 1000 frames, +de 10000 couples lemme/frame
 - +de **150000 occurrences de prédicat annotées pour frame + rôles**
 - les rôles y sont spécifiques à une frame
 - mais des généralisation sont faites via des relations entre frames et entre rôles
 - exemples:
 - frame INGESTION =hérite_de=> frame INGEST_SUBSTANCE
 - rôles reliés Ingestor/Ingestor , Ingestibles/Substance
 - frame KILLING =is_causative_of=> frame DEATH
 - role Victim de KILLING lié à role Protagonist de DEATH

41

FrameNet : exemple de frame

Commitment

[Lexical Unit Index](#)

Definition:

A **Speaker** makes a commitment to an **Addressee** to carry out some future action. This may be an action desirable (as with promise) or not desirable (as with threaten) to the **Addressee**. Some of the words in this frame allow an **Addressee** to be expressed.

You **PROMISED** **me** **you'd come to my graduation.**

Other words cannot normally occur with an **Addressee**.

Kim **UNDERTOOK** **to finish the project by June.**

FEs:

Core:

Addressee [Add]
Semantic Type: Sentient

The Speaker's commitment can be made to an Addressee. With those words which allow this frame element to be expressed, Addressee usually occurs as a PP Complement headed by to or the (Indirect) Object of verbs: Kim made a **PROMISE** **to me**. Kim **PROMISED** **me** that everything would be okay.

Medium [Medium]

As with other frames in the Communication domain, the **MEDIUM** of communication may be expressed. Medium is the physical entity or channel used to transmit the Message. Bob **THREATENED** the workers **in Spanish**.

Message [Msg]
Semantic Type: Message

An expression of the commitment made by the Speaker expresses the frame element Message. Message is expressed as a finite or non-finite clausal Complement or an NP Object: I **SWORE** **that I would never make the same mistake again**. The owner finally **CONSENTED** **to sell the business**. They **THREATENED** **revenge**.

Speaker [Spkr]
Semantic Type: Sentient

The Speaker is the person who commits him/herself to do something. Speaker is usually the External Argument of predicative uses of the target or the Genitive modifier of a noun: **Jo** made a **PROMISE** to me. **Jo** **PROMISED** me she would cook dinner. **Jo's** **PROMISE** to cook dinner came to nothing.

Topic [Top]

Topic expressions are not very frequent in this frame. They can occur with noun targets. Jo made endless **PROMISES** **about working harder at school**.

42

Exemple : Frame Commitment: relations avec d'autres frames unités lexicales associées

FE Core set(s):

{Message, Topic}, {Medium, Speaker}

Frame-frame Relations:

Inherits from:

Is Inherited by:

Perspective on:

Is Perspectivized in:

Uses: [Communication](#)

Is Used by: [Going back on a commitment](#), [Make agreement on action](#)

Subframe of:

Has Subframe(s):

Precedes:

Is Preceded by:

Is Inchoative of:

Is Causative of:

See also:

Lexical Units:

commit.v, commitment.n, consent.v, covenant.n, covenant.v, oath.n, pledge.n, pledge.v, promise.n, promise.v, swear.v, threat.n, threaten.v, undertake.v, undertaking.n, volunteer.v, vow.n, vow.v

FrameNet: Les V converses

- notion de **perspective**:
 - certains lexèmes « profilent » (mettent en avant)
certains participants
- exemple acheter/vendre = 2 frames en
relation de perspective
 - Commerce_sell
 - Commerce_buy
 - mêmes rôles Buyer, Seller, Goods, Money ...

FrameNet: L'alternance locative

- représentée par 2 frames
- Filling
 - V à alternance locative: *load (charger), pile (*entasser), spray (asperger)*
 - ou pas : *fill (remplir), cover (couvrir)*
- Placing
 - V à alternance locative: *load (charger), pile (entasser)*
 - ou pas: *place (placer), inject (injecter) ...*
- => *clairement pas satisfaisant, cf. aucun lien entre les 2 « sens » de « to load »*
 - étonnant sachant que Fillmore est un des premiers à avoir travaillé sur cette alternance (Fillmore, 68, The case for case)

45

FrameNet : Les V psychologiques

- *Experiencer_obj*
 - captiver, effrayer, ensorceler, calmer, reconforter ...
- *Experiencer_focus*
 - craindre, aimer, redouter, haïr ...
- seul le rôle Experiencer est partagé par les 2 frames
 - X=Stimulus effraie Y=Experiencer
 - Y=Experiencer craint X=Content
- dommage!

46

Plan du cours

- Introduction
- Rôles sémantiques
 - définition et perspective historique
 - une liste de rôles, et ses inconvénients
 - exemple de défis d'une théorie des rôles sémantiques
 - quelques théories de rôles sémantiques
 - avec ou sans classes sémantiques de prédicats
- Ressources annotées
- Etiquetage en rôles sémantiques
 - définition
 - applications
 - systèmes de SRL
 - campagnes de SRL (shared tasks)

47

Ressources annotées

- annotées = annotation en rôles sémantiques sur corpus
- **FrameNet** :
 - Parmi les ressources citées supra, seul FrameNet comporte des annotations sur corpus
 - font partie du projet depuis son origine
- **PropBank** :
 - annotation en rôles,
 - indépendamment d'une théorie linguistique particulière
 - bien que lié à VerbNet

48

Annotations FrameNet

- 2 types d'annotations :
 - annotations lexicographiques
 - annotations « full-text »
- Annotations lexicographiques
 - au cours de la constitution d'un frame,
 - recherche des lexèmes pouvant évoquer ce frame
 - NB: lexèmes V, N, Adj, Adv, Prep, Conj
 - pour chaque lexème, recherche dans le corpus BNC d'exemples avec ce sens
 - et annotation en frame + rôles
 - NB: la recherche d'exemples avec un but affiché de couvrir exhaustivement les réalisations syntaxiques possibles des rôles d'un frame, pour un lexème donné
 - NB: le corpus ainsi obtenu **n'est pas représentatif de la distribution normale**
 - des sens d'un prédicat
 - des réalisations syntaxiques des rôles d'un frame, pour un prédicat donné

49

FrameNet: annotations lexicographiques frame = Commitment, lu= to threaten

- 500-vpto
 1. You do n't forget **someone who THREATENS to kill you** .DNI
 2. **The greying gunman** pulled the weapon out of a carrier bag and **THREATENED to shoot a cashier unless she handed over money** .DNI
 3. I knew that I had n't been kidding myself when he turned up one evening looking exhausted , miserable and defeated , to tell me **his wife** had **THREATENED to start divorce proceedings** .DNI
 4. **Angry students at Margaret Thatcher 's old Oxford college** are **THREATENING to sue** over a decision to admit men .DNI
- 530-np-vpto
 1. **He THREATENED tough action** to control council spending -- even if it meant painful limits on Tory authorities .DNI
 2. Already , in November , **de Gaulle** had had to **THREATEN resignation** to put an end to Communist demands for certain key cabinet posts .DNI
 3. **He** also **THREATENED the use of his veto** to force negotiations to a conclusion .DNI
- 570-np-ppwith
 1. **Defies** , who had given up **THREATENING Daak with a court-martial for assault** , voiced the same fear .
 2. **They THREATENED her** with a knife and bound her hands and feet with a rope .INI
 3. Minor disturbances , such as **THREATENING the postman** with a knife when , for the twentieth time , her expectancy of a letter was disappointed , were passed over , but the soap powder was a little frightening .CNIINI
 4. The students followed but **he THREATENED them** with a knife before running off .INI
 5. **They** demanded money and **THREATENED the 22-year-old female assistant** with a piece of scaffolding .INI
 6. Always grumbling , **he** had **THREATENED them with all kinds of dreadful punishment** if he had caught them walking in St Andrew 's churchyard or sheltering in the porch .
 7. **Clients** frequently **THREATEN licensed dealers with exposure to the Press** , and this has proved a relatively efficacious way of getting bona fide complaints heeded .
 8. **They** decide to **THREATEN the old man with violence** .
- 620-np-ppother
 1. ` Do you know if **anybody** ever **THREATENED her** in any way ? "INI
- 650-np-pother
 1. After his arrest **he THREATENED officers** by drawing his finger across his neck to suggest **he would one day kill them** .
 2. **Her parents THREATENED a civil suit for ` emotional distress "** .DNI
 3. **He THREATENED his crew** with having their tongues cut out if they would not swear that they had sailed along the coast of the Indies (they had , in fact , been sailing past Cuba) .
- 670-pass-by
- 680-pass
 1. **My family 's** being **THREATENED** .CNIINI
 2. Towards the end of April matters came to a head when **parents** were **THREATENED with prosecution** .CNI

50

FrameNet : annotations full-text

- tentative d'annotation de texte continu
 - par opposition à des exemples choisis
 - <https://framenet.icsi.berkeley.edu/fndrupal/index.php?q=fulltextIndex>
- création pour l'occasion de frames supplémentaires
 - une partie de ces données annotées ont servi de données de développement / données d'évaluation à SemEval 2007, tâche 19
 - cf. infra

51

FrameNet: annotation full-text

NTI - Iran_Biological

[Lexical Unit Index](#) [Frame Index](#)

1. Overview

2. *Iran is often ACCUSED_{Judgment_communication} in Western CIRCLES_{Aggregate_of} SECRETLY_{Secrecy_status} DEVELOPING_{Cause_to_make_progress} an OFFENSIVE_{Attack} BIOLOGICAL_{Weapon} WARFARE_{Hostile_encounter} (BW_{Hostile_encounter}) PROGRAM_{Project} . In a 1996 REPORT_{Statement} to the US Senate , the Central Intelligence Agency CLAIMED_{Statement} that `` Iran has HAD_{Possession} a BIOLOGICAL_{Weapon} WARFARE_{Hostile_encounter} PROGRAM_{Project} SINCE_{Time_vector} the early 1980s . Currently the PROGRAM_{Project} is in its RESEARCH_{Research} and DEVELOPMENT_{Cause_to_make_progress} stages , but we BELIEVE_{Certainty} Iran HOLDS_{Containing} some STOCKS_{Store_of} BW_{Hostile_encounter} AGENTS_{Active_substance} and WEAPONS_{Weapon} ... ''*

3. *Iran has RATIFIED_{Ratification} the Biological and Toxin Weapons Convention (BWC) and publicly DECRIED_{Judgment_communication} all FORMS_{Type_of} weapons of mass destruction (WMD) . At the same time , it does HAVE_{Possession} a very SOPHISTICATED_{Stage_of_progress} BIOTECHNOLOGY_{Artifact} INFRASTRUCTURE_{Infrastructure} that INCLUDES_{Inclusion} LEADING_{First_rank} RESEARCH_{Research} FACILITIES_{Locale_by_use} and TRAINED_{Education_teaching} PERSONNEL_{Employing} . It has also PARTICIPATED_{Participation} in several scientific EXCHANGES_{Discussion} with FOREIGN_{Foreign_or_domestic_country} EXPERTS_{Expertise} .*

[Clear Sentences](#) [Turn Colors On](#)

[X] [EvalueeIran] is [Frequencyoften] ACCUSED^{Target} [Communicatorin Western circles] [Reasonof secretly developing an offensive biological warfare (BW) program] .

[X] [CommunicatorIran] has ratified the Biological and Toxin Weapons Convention (BWC) and [Mannerpublicly] DECRIED^{Target} [Evalueeall forms of weapons of mass destruction (WMD)] [ReasonINI]

52

FrameNet : autres langues

- les généralisations obtenues dans FrameNet sont dans une certaine mesure faites indépendamment de la syntaxe de l'anglais
 - contrairement par ex. aux classes de Levin
 - => la structure FrameNet (frames + relations entre frames) peut jouer le rôle d'une structure conceptuelle, compatible avec plusieurs langues
- constitution automatique de ressources FrameNet pour d'autres langues
 - par **projection** à partir de l'anglais
 - constitution de **lexiques** framenet pour d'autres langues
 - projection via dictionnaires électroniques
 - pour le français : Mouton et al. 2010
 - => lexique framenet-fr (bruité)
 - projection via alignements de mots obtenus sur corpus parallèles
 - thèse de Sebastian Padò, 2007 : Cross-lingual annotation projection models for Role-Semantic Information
 - application à l'allemand et le français
 - constitution de **données annotées**
 - également, thèse de Sebastian Padò, 2007
 - résultat bcp plus hasardeux
 - corpus parallèle
 - apprentissage d'un analyseur framenet, sur les données annotées pour l'anglais
 - application sur partie anglaise d'un corpus parallèle
 - projection à la partie langue cible du corpus parallèle
 - en suivant les liens d'alignement

53

FrameNet : autres langues

- constitution manuelle et semi-automatique
 - projets lexicographiques
 - espagnol, japonais, suédois, brésilien ...
 - projets orienté corpus, sur arbres syntaxiques
 - allemand : projet SALSA
 - annotation exhaustive via framenet des sens de 5000 verbes
 - tels que trouvés dans le treebank Tiger
 - => préservation de la distribution des sens
 - français : projet ASFALDA (2012-2015)
 - annotation sur FrenchTreebank
 - annotation exhaustive via FrameNet
 - de concepts (« transaction commerciale », « communication »)...
 - de verbes moyennement fréquents (en couvrant tous leurs sens)

54

Propbank

- Palmer, Gildea and Kingsbury, The Proposition Bank: A Corpus Annotated with Semantic Roles, Computational Linguistics Journal, 31:1, 2005
 - <http://verbs.colorado.edu/~mpalmer/projects/ace.html>
- arbres syntagmatiques du PennTreebank annotés avec
 - sens des occurrences de verbes
 - lien avec classe VerbNet quand possible
 - rôles sémantiques des arguments de ces verbes
- rôles sémantiques
 - **spécifiques à un sens de verbe**
 - 6 rôles pour arguments sous-catégorisés : arg0, arg1, ... arg6
 - 11 rôles pour les ajouts : ArgM-TMP, ArgM-MNR, ArgM-LOC ...⁵⁵

PropBank: exemple : *to decline*

Frameset: **decline.01** "go down incrementally"

Arg1: entity going down

Arg2: amount gone down by, EXT

Arg3: start point

Arg4: end point

Ex: ..._[Arg1 its net income] *declining* _[Arg2-EXT 42%] _[Arg4 to \$121 million] _[ArgM-TMP in the first 9 months of 1989]. (wsj_0067)

Frameset: **decline.02** "demure, reject"

Arg0: agent

Arg1: rejected thing

Ex: _[Arg0 A spokesman_i] *declined* _[Arg1 *trace*_i to elaborate] (wsj_0038)

PropBank: très orienté vers la syntaxe

- annotation sur arbres syntagmatiques
 - but affiché: servir de données d'entraînements pour extraire les relations prédicats-arguments à partir d'une analyse syntaxique
- numérotation des arguments sémantiques en fonction de la sous-catégorisation canonique

PropBank	
<i>buy</i>	<i>sell</i>
Arg0: buyer	Arg0: seller
Arg1: thing bought	Arg1: thing sold
Arg2: seller	Arg2: buyer
Arg3: price paid	Arg3: price paid
Arg4: benefactive	Arg4: benefactive

- rem: la glose associée au rôle a un but explicatif, mais n'a pas de statut théorique

57

PropBank : extensions

- autres langues :
 - le caractère orienté syntaxe rend impossible la réutilisation de la ressource propbank pour d'autres langues
 - mais la méthodologie est réutilisable
 - => de fait il existe un Chinese Propbank
 - <http://www.cis.upenn.edu/~chinese/>
- extension aux noms prédicatifs
 - NomBank
 - <http://nlp.cs.nyu.edu/meyers/NomBank.html>

58

PropBank: peu de généralisations

- les verbes ne sont pas organisés en classes
- généralisations via les rôles associés aux ajouts
 - Time, Manner, Cause, Location...
- rôles arg0 et arg1: généralisation à la **Dowty**
 - For a particular verb, Arg0 is generally the argument exhibiting features of a prototypical Agent (Dowty, 1991) while Arg1 is a prototypical Patient or Theme. No consistent generalizations can be made across verbs for the higher numbered arguments, though an effort was made to consistently define roles across members of VerbNet classes. (Palmer et al., 2005, p.4)
- généralisation sur les différentes **diathèses** pour un même sens de V :
 - ...[A0 the company] to ... offer [A1 a 15% to 20% stake] [A2 to the public]
 - ...[A0 Sotheby's] ... offered [A2 the Dorrance heirs] [A1 a money-back guarantee]
 - ...[A1 an amendment] offered [A0 by Rep. Peter DeFazio] ...
 - ...[A2 Subcontractors] will be offered [A1 a settlement] ...

59

Ressources annotées: pros and cons

- framenet
 - + généralisations sémantiques, qui s'affranchissent en partie de la syntaxe
 - => intérêt applicatif
 - => utilisabilité pour plusieurs langues
 - - couverture des frames totalement insuffisante
 - - non représentativité des exemples lexicographiques
 - - incohérences manifestes
 - différences ds le niveau de granularité des frames
 - incohérences dans les relations entre frames
- propbank
 - + taille conséquente de texte continu annoté
 - mais limité au Wall Street Journal
 - - manque de généralisations
 - avec une annotation à la propbank, on est encore loin de savoir « qui a fait quoi quand où ... »
 - conceptuellement + simple, et donc = la référence de données supervisées pour la tâche de SRL

60