

Un corpus de verbes annoté sémantiquement avec les tables du Lexique-Grammaire

Myriam RAKHO

Laboratoire d'Informatique de l'Institut Gaspard Monge
Université Paris-Est Marne-la-Vallée

Description de la tâche

Un mot polysémique et ses différents sens possibles

comprendre **A** **1** Contenir, renfermer en soi. **2** Faire entrer dans un tout. **3** Saisir le sens de. **4** Se représenter, se faire une idée de. **5** Se rendre compte de, que. **6** Faire preuve de compréhension, de tolérance envers. **B** **1** Pouvoir être compris. **2** Bien s'entendre, en parlant de personnes.
[*Dictionnaire* HACHETTE 2009]

Description de la tâche

Un mot polysémique et ses différents sens possibles

comprendre **A** **1** Contenir, renfermer en soi. **2** Faire entrer dans un tout. **3** Saisir le sens de. **4** Se représenter, se faire une idée de. **5** Se rendre compte de, que. **6** Faire preuve de compréhension, de tolérance envers. **B** **1** Pouvoir être compris. **2** Bien s'entendre, en parlant de personnes.
[Dictionnaire HACHETTE 2009]

Des occurrences de ce mot dans leur contexte :

*Vous devez **comprendre** que nous ne vous laisserons pas faire cela.*

*Nous **comprenons** votre inquiétude*

*Le prochain dialogue sur les droits de l'homme **comprendra** des séminaires sur...*

Description de la tâche

Un mot polysémique et ses différents sens possibles

comprendre **A** **1** Contenir, renfermer en soi. **2** Faire entrer dans un tout. **3** Saisir le sens de. **4** Se représenter, se faire une idée de. **5** Se rendre compte de, que. **6** Faire preuve de compréhension, de tolérance envers. **B** **1** Pouvoir être compris. **2** Bien s'entendre, en parlant de personnes.
[Dictionnaire HACHETTE 2009]

Des occurrences de ce mot dans leur contexte :

*Vous devez **comprendre** que nous ne vous laisserons pas faire cela.*

*Nous **comprenons** votre inquiétude*

*Le prochain dialogue sur les droits de l'homme **comprendra** des séminaires sur...*



WSD = Associer à chaque occurrence l'étiquette du sens qui lui correspond

L'approche supervisée

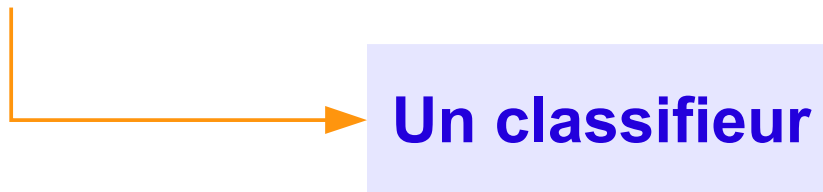
Un corpus de référence

SENS	CONTEXTE
4	Nous ne comprenons pas pourquoi la Commission entreprend...
6	Je comprends très bien votre préoccupation.
1	Les textes des traités ne comprennent pas de dispositions spécifiques.

L'approche supervisée

Un corpus de référence

SENS	CONTEXTE
4	Nous ne comprenons pas pourquoi la Commission entreprend...
6	Je comprends très bien votre préoccupation.
1	Les textes des traités ne comprennent pas de dispositions spécifiques.



L'approche supervisée

Un corpus de référence

SENS	CONTEXTE
4	Nous ne comprenons pas pourquoi la Commission entreprend...
6	Je comprends très bien votre préoccupation.
1	Les textes des traités ne comprennent pas de dispositions spécifiques.



Un corpus de test

CONTEXTE

*Vous devez **comprendre** que nous ne vous laisserons pas faire cela.*
*Nous **comprenons** votre inquiétude*
*Le prochain dialogue sur les droits de l'homme **comprendra** des séminaires sur..*

Les étiquettes de sens (1)

- ▶ Dans l'approche multilingue :
 - **les traductions** en langue cible extraites d'un corpus bilingue

Les étiquettes de sens (1)

- Dans l'approche multilingue :
 - **les traductions** en langue cible extraites d'un corpus bilingue

*Je **comprends** les craintes de l'honorable député*
*I **recognise** the concern of the honourable Member*

*Les responsabilités **comprennent**, déjà à cet égard, les relations avec la CC*
*Here responsibilities **include** relations with the Court of Auditors*

*Nous **comprenons** vos inquiétudes*
*We **understand** your concerns*
etc.



Sens : *understand ; include ; realise ; understanding ; appreciate ; see ; contain ; aware ;
comprise ; grasp ; involve ; realize ; cover ; comprehend ; recognise ; gather ; sympathise ; mean ;
hear ; get ; consist ; accept ; interpret ; impress ; incorporate ; encompass ; etc.*

Les étiquettes de sens (2)

involve	<i>Le principe de précaution ne comprend pas la politisation de la science</i> <i>The precautionary principle does not involve the politicisation of science</i>
imply	<i>Je crois vraiment que la réforme, avec tout ce qu'elle comprend, facilitera ...</i> <i>I do believe that reform with all it implies will very strongly facilitate ...</i>
contain	<i>...un récit des méfaits commis comprenant l'énumération des scandales ...</i> <i>It contains scandals ...</i>

→ 1 sens - plusieurs traductions = sens trop fins

Les étiquettes de sens (2)

involve	<i>Le principe de précaution ne comprend pas la politisation de la science</i> <i>The precautionary principle does not involve the politicisation of science</i>
imply	<i>Je crois vraiment que la réforme, avec tout ce qu'elle comprend, facilitera ...</i> <i>I do believe that reform with all it implies will very strongly facilitate ...</i>
contain	<i>...un récit des méfaits commis comprenant l'énumération des scandales ...</i> <i>It contains scandals ...</i>

→ 1 sens - plusieurs traductions = sens trop fins

understand	<i>Les personnes commettant des fraudes doivent comprendre qu'elles seront poursuivies; People need to understand that if they commit fraud...</i>
understand	<i>Nous comprenons votre inquiétude; We understand your concerns</i>
understand	<i>Vous n'y comprenez rien; You do not understand the first thing of it</i>

→ 3 sens - 1 même traduction = sens trop grossiers

Les étiquettes de sens (2)

► Dans notre approche :

→ les emplois des mots dans le Lexique-Grammaire

[Boons, Guillet et Leclère, 1976a, 1976b] [Gross 1975] [Guillet et Leclère, 1992]

6

ID	N0 =: Nhum	<ENT>V	Aux := avoir	Aux := être	N1 =: Qu Pind	N1 =: Qu Psubj	<OPT>
73	+	<i>comprendre</i>	+	-	+	-	<i>Max a compris qu'Ida était coupable à cet indice</i>

12

ID	N0 =: Nhum	<ENT>V	Aux := avoir	Aux := être	N1 =: Qu Pind	N1 =: Nhum	<OPT>
17	+	<i>comprendre</i>	+	-	-	+	<i>Max comprend qu'Ida ne vienne pas</i>

Les étiquettes de sens (2)

comprendre

table	entrée	exemple
6	73	Max a compris qu'Ida était coupable à cet indice
12	17	Max comprend qu'Ida ne vienne pas
32R3	178	Max comprend (cette langue + l'anglais)
38R	54	Max a compris cette remarque comme une plaisanterie
c7	49	\$ne comprendre \$ rien
10	46	Faire ce travail comprend pour Max de tout nettoyer
32NM	26	Ce livre comprend dix chapitres
38LR	48	Le garçon comprend le service dans la note

(total : 24 emplois, dans le Lexique-Grammaire dont 8 libres et 16 figés)



Sens : *V_6_73 ; V_12_17 ; V_32R3_18 ; V_38R_54 ; V_31H_124 ; C_c7_49 ;
V_10_46 ; V_32NM_26 ; V_38LR_48*

Les étiquettes de sens (2)

- V_10_46** *Le principe de précaution ne **comprend** pas la politisation de la science*
*The precautionary principle does not **involve** the politicisation of science*
- V_10_46** *Je crois vraiment que la réforme, avec tout ce qu'elle **comprend**, facilitera ...*
*I do believe that reform with all it **implies** will very strongly facilitate ...*
- V_10_46** *...un récit des méfaits commis **comprenant** l'énumération des scandales ...*
*It **contains** scandals ...*

- 1 sens - plusieurs traductions = sens trop fins
- 1 sens - 1 entrée LG

Les étiquettes de sens (2)

- V_10_46** *Le principe de précaution ne **comprend** pas la politisation de la science*
*The precautionary principle does not **involve** the politicisation of science*
- V_10_46** *Je crois vraiment que la réforme, avec tout ce qu'elle **comprend**, facilitera ...*
*I do believe that reform with all it **implies** will very strongly facilitate ...*
- V_10_46** *...un récit des méfaits commis **comprenant** l'énumération des scandales ...*
*It **contains** scandals ...*

- 1 sens - plusieurs traductions = sens trop fins
- 1 sens - 1 entrée LG

- V_6_73** *Les personnes commettant des fraudes doivent **comprendre** qu'elles seront poursuivies; People need to **understand** that if they commit fraud...*
- V_12_17** *Nous **comprenons** votre inquiétude; We **understand** your concerns*
- C_c7_49** *Vous n'y **comprenez** rien; You do not **understand** the first thing of it*

- 3 sens - 1 même traduction = sens trop grossiers
- 3 sens - 3 entrées LG

Le corpus annoté

Données

- Sélection des verbes polysémiques
- Constitution de leur corpus

Annotations

- Traductions
- Emplois du Lexique-Grammaire
- Lexies
- Classes de lexies

Exemple d'utilisation

- Désambiguïsation sémantique

Partie 1 – Le corpus annoté

Le corpus annoté

Données

Sélection des verbes polysémiques
Constitution de leur corpus

Annotations

Traductions
Emplois du Lexique-Grammaire
Lexies
Classes de lexies

Exemple d'utilisation

Désambiguïsation sémantique

Sélection des verbes polysémiques

Liste constituée lors de la campagne ARCADE pour l'évaluation des systèmes d'alignement automatique multilingue au niveau des mots

Sélection des verbes polysémiques

Liste constituée lors de la campagne ARCADE pour l'évaluation des systèmes d'alignement automatique multilingue au niveau des mots

Procédure de sélection :

- **Corpus** : le *Journal Officiel de la Communauté européenne* (JOC)
(5 langues : allemand, anglais, espagnol, français, italien ; environ 1 million de mots par langue ; corpus aligné au niveau des phrases)

Sélection des verbes polysémiques

Liste constituée lors de la campagne ARCADE pour l'évaluation des systèmes d'alignement automatique multilingue au niveau des mots

Procédure de sélection :

- **Corpus** : le *Journal Officiel de la Communauté européenne* (JOC)
(5 langues : allemand, anglais, espagnol, français, italien ; environ 1 million de mots par langue ; corpus aligné au niveau des phrases)
- 1. élimination des mots polycatégoriels selon le dictionnaire Multext
- 2. sélection des 200 noms, 200 verbes et 200 adjectifs les plus fréquents
- 3. alignement des contextes en français qui les contiennent avec leur traduction dans les 4 autres langues du corpus : fr-en, fr-all, fr-ita, fr-esp
- 4. présentation de ces couples de contextes à 6 juges humains qui devaient cocher, pour chaque couple, le ou les sens du mot concerné dans le contexte en question ou « je ne sais pas ».
- 5. extraction des 20 mots les plus polysémiques de chaque catégorie

Extraction des corpus des verbes

Le corpus utilisé : EuroParl-Intersection

(version SemEval-2010 du corpus EuroParl, 6 langues : allemand, anglais, espagnol, français, italien, néerlandais ; 884 603 paragraphes par langue, 10 à 50 M de mots par langue ; paragraphes alignés dans les 6 langues)

→ Extraction des contextes en français (langue source) qui contiennent l'un des 20 verbes étudiés

Extraction des corpus des verbes

Le corpus utilisé : EuroParl-Intersection

(version SemEval-2010 du corpus EuroParl, 6 langues : allemand, anglais, espagnol, français, italien, néerlandais ; 884 603 paragraphes par langue, 10 à 50 M de mots par langue ; paragraphes alignés dans les 6 langues)

→ Extraction des contextes en français (langue source) qui contiennent l'un des 20 verbes étudiés

<i>arrêter</i> (2104)	<i>couvrir</i> (1286)	<i>ouvrir</i> (2657)	<i>présenter</i> (11810)
<i>comprendre</i> (12260)	<i>entrer</i> (2354)	<i>parvenir</i> (7470)	<i>rendre</i> (6774)
<i>conclure</i> (4652)	<i>exercer</i> (1864)	<i>passer</i> (6489)	<i>répondre</i> (7526)
<i>conduire</i> (2115)	<i>importer</i> (3058)	<i>porter</i> (3342)	<i>tirer</i> (2165)
<i>connaître</i> (5813)	<i>mettre</i> (8731)	<i>poursuivre</i> (5364)	<i>venir</i> (8198)

Nombre de contextes trouvés dans le corpus EuroParl-Intersection pour chaque verbe
(5000 contextes par mot en moyenne)

Le corpus annoté

Données

- Sélection des verbes polysémiques
- Constitution de leur corpus

Annotations

- Traductions
- Emplois du Lexique-Grammaire
- Lexies
- Classes de lexies

Exemple d'utilisation

- Désambiguïsation sémantique

Les traductions : exemple

SENS	CONTEXTE
include	<i>Les responsabilités comprennent, déjà à cet égard, les relations avec la CC</i>
	<i>Here responsibilities include relations with the Court of Auditors</i>
involve	<i>Le principe de précaution ne comprend pas la politisation de la science</i>
	<i>The precautionary principle does not involve the politicisation of science</i>
understand	<i>Nous comprenons votre inquiétude</i>
	<i>We understand your concerns</i>
understand	<i>Vous n'y comprenez rien</i>
	<i>You do not understand the first thing of it</i>

Les traductions : annotation

1. alignement bi-directionnel au niveau des mots avec Giza++ :
→ 2 alignements : français-anglais et anglais-français

puis extraction automatique des mots en anglais qui ont été alignés avec le verbe polysémique dans les deux alignements

FR : *Les personnes commettant des fraudes doivent **comprendre** qu'elles seront poursuivies*

EN : *People need to **understand** that if they commit fraud they will be prosecuted*

français	→	anglais :
personne		people
commettre		commit
fraude		fraud
devoir		need
comprendre		understand
être		be
poursuivre		prosecute

anglais	→	français :
people		personne
need		devoir
understand		comprendre
commit		commettre
fraud		fraude
be		être
prosecute		poursuivre

Les traductions : annotation

2. validation manuelle des alignements

→ le mot polysémique est traduit par une expression multi-mots (*mwe*) en anglais :

français	→	anglais :	anglais	→	français :	
- <i>conclure</i>		<i>closure</i>	<i>bring</i>		<i>conclure</i>	=> <i>conclure = bring closure</i>
- <i>arrêter</i>		<i>justice</i>	<i>bring</i>		<i>arrêter</i>	=> <i>arrêter = bring to justice</i>

→ le verbe polysémique appartient à une mwe traduite par un seul mot en anglais :

français	→	anglais :	anglais	→	français :	
- <i>porter</i>		<i>damage</i>	<i>damage</i>		<i>atteinte</i>	=> <i>porter atteinte = damage</i>
- <i>porter</i>		<i>compromise</i>	<i>compromise</i>		<i>atteinte</i>	=> <i>porter atteinte = compromise</i>
- <i>rendre</i>		<i>publicise</i>	<i>publicise</i>		<i>public</i>	=> <i>rendre public = publicise</i>

Les traductions : exemple

comprendre (183 traductions)

{ *understand* (**5065**) ; *include* (679) ; *realise* (260) ; *appreciate* (200) ; *contain* (135) ; *comprise* (62) ; *grasp* (52) ; *realize* (41) ; *involve* (40) ; *comprehend* (37) ; *cover* (34) ; *recognise* (33) ; *gather* (32) ; *sympathise* (30) ; *consist* (21) ; *get* (21) ; *hear* (21) ; *accept* (20) ; *interpret* (20) ; *impress* (18) ; *encompass* (16) ; *incorporate* (16) ; *recognize* (11) ; *embrace* (7) ; *sense* (7) ; *bringhometo* (6) ; *imagine* (6) ; ... ; *amountto* (1) ; *becomeapparent* (1) ; *extend* (1) ; *inclusive* (1) ; *lie* (1) ; *makeplain* (1) ; *relateto* (1) ; *surround* (1) ; *together*(1) ; *too* (**1**) }

Les traductions : statistiques

verbe	#contextes	#traductions
<i>arrêter</i>	2033	150
<i>comprendre</i>	8240	183
<i>conclure</i>	3488	79
<i>conduire</i>	2114	96
<i>connaitre</i>	5786	158
<i>couvrir</i>	2183	85
<i>entrer</i>	2325	107
<i>exercer</i>	1851	86
<i>importer</i>	2778	71
<i>ouvrir</i>	2656	127
<i>parvenir</i>	7469	152
<i>porter</i>	3301	219
<i>poursuivre</i>	5354	154
<i>rendre</i>	6731	177
<i>tirer</i>	2163	102
<i>venir</i>	7369	120
Moyenne	4115	129

Les emplois du Lexique-Grammaire : annotation

- choix entre les emplois libres et figés des verbes dans les tables du Lexique-Grammaire
- Étiquettes de sens : $\{V|C\}_{table_entrée}$
V = emploi libre, C = emploi figé, *table* = nom d'une table, *entrée* = entrée dans la table

Les emplois du Lexique-Grammaire : annotation

- choix entre les emplois libres et figés des verbes dans les tables du Lexique-Grammaire
- Étiquettes de sens : $\{V|C\}_{table_entrée}$
V = emploi libre, C = emploi figé, *table* = nom d'une table, *entrée* = entrée dans la table

comprendre (8 emplois)

étiquette	table	entrée	exemple
V_6_73	6	73	Max a compris qu'Ida était coupable à cet indice
V_12_17	12	17	Max comprend qu'Ida ne vienne pas
V_32R3_178	32R3	178	Max comprend (cette langue + l'anglais)
V_38R_54	38R	54	Max a compris cette remarque comme une plaisanterie
C_c7_49	c7	49	\$ne comprendre\$ rien
V_10_46	10	46	Faire ce travail comprend pour Max de tout nettoyer
V_32NM_26	32NM	26	Ce livre comprend dix chapitres
V_38LR_48	38LR	48	Le garçon comprend le service dans la note

(total : 24 emplois, dans le Lexique-Grammaire dont 8 libres et 16 figés)

Les emplois du Lexique-Grammaire : exemple

comprendre (8 emplois)

{ V_6_73 (**4560**); V_12_17 (2495); V_10_46 (541); V_32NM_26 (458); V_38LR_48 (50);
V_38R_54 (47); C_c7_49 (46); V_32R3_178 (**43**) }

SENS	CONTEXTE
V_10_46	Les responsabilités comprennent , déjà à cet égard, les relations avec la CC
V_10_46	Le principe de précaution ne comprend pas la politisation de la science
V_12_17	Nous comprenons votre inquiétude
C_c7_49	Vous n'y comprenez rien

Les emplois du Lexique-Grammaire : annotation

- un seul emploi possible par contexte
en cas d'ambiguïté : **priorité au sens proprement linguistique**
par rapport au sens pour le locuteur

Les emplois du Lexique-Grammaire : annotation

- un seul emploi possible par contexte
en cas d'ambiguïté : **priorité au sens proprement linguistique**
par rapport au sens pour le locuteur

*Je ne **comprends** pas lorsque vous affirmez qu ' il n ' y aura pas de transfert de compétences . Si vous étendez le vote_à_la_majorité et que , par conséquent , vous supprimez le veto national , vous transférez donc des compétences de l ' État nation vers les institutions_européennes .*

- sens proprement linguistique : V_6_73
- sens implicite : V_12_17
= *Je ne comprends pas que vous disiez qu'il n'y aura pas de transfert de compétences*
- sens assigné : **V_6_73**

Les emplois du Lexique-Grammaire : annotation

- un seul emploi possible par contexte
en cas d'ambiguïté : **priorité au sens proprement linguistique**
par rapport au sens pour le locuteur

*L ' amendement 14 ne vas pas vraiment dans ce sens et je ne peux par conséquent pas le soutenir . Quant à l' amendement 13 , l ' agitation de rigueur a vu le jour . Bref , vous **comprendrez** que je ne peux pas non plus soutenir l ' amendement 13 . J ' espère que les collègues en entreverront la logique .*

- sens proprement linguistique : V_6_73
- sens véhiculé par le locuteur : V_12_17
= *Vous comprendrez que je ne puisse pas non plus soutenir l' amendement 13*
- sens assigné : **V_6_73**

Les emplois du Lexique-Grammaire : statistiques

verbe	#contextes	#traductions	#entrées LG
<i>arrêter</i>	2033	150	12
<i>comprendre</i>	8240	183	8
<i>conclure</i>	3488	79	5
<i>conduire</i>	2114	96	10
<i>connaitre</i>	5786	158	14
<i>couvrir</i>	2183	85	16
<i>entrer</i>	2325	107	6
<i>exercer</i>	1851	86	4
<i>importer</i>	2778	71	5
<i>ouvrir</i>	2656	127	17
<i>parvenir</i>	7469	152	3
<i>porter</i>	3301	219	20
<i>poursuivre</i>	5354	154	5
<i>rendre</i>	6731	177	14
<i>tirer</i>	2163	102	19
<i>venir</i>	7369	120	12
Moyenne	4115	129	10

Les lexies : annotation

- Concaténation, par contexte, des deux premières étiquettes de sens : les traductions et les emplois du Lexique-Grammaire
- Étiquettes de sens : *emploi_LG#traduction*

SENS	CONTEXTE
<i>V_10_46#include</i>	Les responsabilités comprennent , déjà à cet égard, les relations avec la CC
<i>V_10_46#involve</i>	Le principe de précaution ne comprend pas la politisation de la science
<i>V_12_17#understand</i>	Nous comprenons votre inquiétude
<i>C_c7_49#understand</i>	Vous n'y comprenez rien

Les lexies : annotation

- Concaténation, par contexte, des deux premières étiquettes de sens : les traductions et les emplois du Lexique-Grammaire
- Étiquettes de sens : *emploi_LG#traduction*

SENS	CONTEXTE
<i>V_10_46#include</i>	Les responsabilités comprennent , déjà à cet égard, les relations avec la CC
<i>V_10_46#involve</i>	Le principe de précaution ne comprend pas la politisation de la science
<i>V_12_17#understand</i>	Nous comprenons votre inquiétude
<i>C_c7_49#understand</i>	Vous n'y comprenez rien

comprendre (183 traductions, 8 emplois, 308 lexies)

{ V_6_73#understand (**3008**); V_12_17#understand (1956); **V_10_46#include** (362); **V_32NM_26#include** (254); V_6_73#realise (241); V_6_73#appreciate (118); V_6_73#see (90); V_32NM26#contain (82); V_12_17#appreciate (81); V_12_17#see(58); V_10_46#contain (52); V_6_73#grasp (46); V_32R3_178#understand (41); V_6_73#know (40); V_6_73#realize (40); V_32NM_26#comprise(37); C_c7_49#understand (36); **V_38LR_48#include** (32); etc. }

Les lexies : statistiques

verbe	#contextes	#traductions	#entrées LG	#lexies
<i>arrêter</i>	2033	150	12	242
<i>comprendre</i>	8240	183	8	308
<i>conclure</i>	3488	79	5	122
<i>conduire</i>	2114	96	10	145
<i>connaitre</i>	5786	158	14	238
<i>couvrir</i>	2183	85	16	128
<i>entrer</i>	2325	107	6	189
<i>exercer</i>	1851	86	4	120
<i>importer</i>	2778	71	5	101
<i>ouvrir</i>	2656	127	17	253
<i>parvenir</i>	7469	152	3	185
<i>porter</i>	3301	219	20	494
<i>poursuivre</i>	5354	154	5	219
<i>rendre</i>	6731	177	14	347
<i>tirer</i>	2163	102	19	160
<i>venir</i>	7369	120	12	306
Moyenne	4115	129	10	222

Les classes de lexies : annotation

- Classes des lexies associées à un même emploi LG
- Étiquettes de sens : *{/lexie+}*
Chaque contexte a pour étiquette la classe de lexies à laquelle appartient la lexie qui lui était assignée précédemment.

CLASSE DE LEXIES	CONTEXTE
<i>{V_10_46#include ; V_10_46#involve ; V_10_46#imply ; etc.}</i>	<i>Les responsabilités comprennent, déjà à cet égard, les relations avec la CC</i>
<i>{V_10_46#include ; V_10_46#involve ; V_10_46#imply ; etc.}</i>	<i>Le principe de précaution ne comprend pas la politisation de la science</i>
<i>{V_12_17#understand ; V_12_17#appreciate ; V_12_17#sympathise ; etc.}</i>	<i>Nous comprenons votre inquiétude</i>

Conclusion

TRAD	LG	LEXIE	CLASSE DE LEXIES	CONTEXTE
include	V_10_46	V_10_46#include	{V_10_46#include ; V_10_46#involve ; V_10_46#imply ; etc.}	Les responsabilités comprennent , déjà à cet égard, les relations avec la CC
involve	V_10_46	V_10_46#involve	{V_10_46#include ; V_10_46#involve ; V_10_46#imply ; etc.}	Le principe de précaution ne comprend pas la politisation de la science
understand	V_12_17	V_12_17#understand	{V_12_17#understand ; V_12_17#appreciate ; V_12_17#sympathise ; etc.}	Nous comprenons votre inquiétude
understand	C_c7_49	C_c7_49#understand	C_c7_49#understand	Vous n'y comprenez rien

Partie 2 – Exemple d'utilisation

Les représentations des contextes

- **syn-all :**
cooccurents syntaxiques des verbes (gouverneur et dépendants) toutes relations syntaxiques confondues
- **syn-srel-\$REL :**
cooccurents syntaxiques (gouverneur et dépendants) par relations syntaxique
(*syn-srel-SUJ*, *syn-srel-OBJ*, *syn-srel-AUX_TPS*, etc.)
- **ngb :**
voisins (noms, verbes, adjectifs et adverbes)

Analyse syntaxique

Analyse syntaxique avec l'analyseur en dépendances Bonsai-v2.3

[Candito et Crabbé, 2010]

*Les personnes commettant des fraudes doivent **comprendre** qu ' elles seront poursuivies .*

rank	wordform	lemma	cpos	fpos	g m	n p s	t	grank	srel
1	Les	le	D	DET	- -	p - def	-	2	det
2	personnes	personne	N	NC	f -	p - neg	-	6	suj
3	commettant	commettant	A	ADJ	- -	- - qual	-	2	mod
4	des	de	P+D	P+D	- -	- - def	-	2	dep
5	fraudes	fraude	N	NC	f -	p - c	-	4	obj
6	doivent	devoir	V	V	- ind	p 3 -	pst	0	obj
7	comprendre	comprendre	V	VINF	- inf	- - -	-	6	obj
8	qu'	que	C	CS	- -	- - s	-	7	obj
9	elles	cln	CL	CLS	- -	- - suj	-	11	suj
10	seront	être	V	V	- ind	p 3 -	fut	11	aux_pass
11	poursuivies	poursuivre	V	VPP	f part	p - -	past	8	obj
21	.	.	PONCT	PONCT	- -	- - s	-	6	ponct

Légende : *rank*=position ; *wordform*=forme fléchiée ; *lemma*=lemme ; *cpos*=étiquette grammaticale grossière ; *fpos*=étiquette grammaticale fine ; *g*=genre ; *m*=mode ; *n*=nombre ; *p*=personne ; *s*=sémantique ; *t*=temps ; *grank*=position du gouverneur syntaxique ; *srel*=relation syntaxique avec le gouverneur

La représentation *syn-all*

- contenu : cooccurents syntaxiques (gouverneur et dépendants)
- vecteur de descripteurs : *relation_syntaxique#caractéristique=valeur*

*Les personnes commettant des fraudes doivent **comprendre** qu ' elles seront poursuivies .*

rank	worform	lemma	cpos	fpos	g	m	n	p	s	t	grank	srel
1	Les	le	D	DET	-	-	p	-	def	-	2	det
2	personnes	personne	N	NC	f	-	p	-	neg	-	6	subj
3	commettant	commettant	A	ADJ	-	-	-	-	qual	-	2	mod
4	des	de	P+D	P+D	-	-	-	-	def	-	2	dep
5	fraudes	fraude	N	NC	f	-	p	-	c	-	4	obj
6	doivent	devoir	V	V	-	ind	p	3	-	pst	0	obj
7	comprendre	comprendre	V	VINF	-	inf	-	-	-	-	6	obj
8	qu'	que	C	CS	-	-	-	-	s	-	7	obj
9	elles	cln	CL	CLS	-	-	-	-	subj	-	11	subj
10	seront	être	V	V	-	ind	p	3	-	fut	11	aux_pass
11	poursuivies	poursuivre	V	VPP	f	part	p	-	-	past	8	obj
21	.	.	PONCT	PONCT	-	-	-	-	s	-	6	ponct

Légende : *la cible* ; *son gouverneur syntaxique* ; *ses dépendants syntaxiques*

La représentation *syn-all*

- contenu : cooccurrents syntaxiques (gouverneur et dépendants)
- vecteur de descripteurs : *relation_syntaxique#caractéristique=valeur*

*Les personnes commettant des fraudes doivent **comprendre** qu ' elles seront poursuivies .*

SUJ#wordform=personnes ; SUJ#lemma=personne ; SUJ#cpos=N ;
SUJ#fpos=NC ; SUJ#g=f ; SUJ#n=p
OBJ#wordform=doivent ; OBJ#lemma=devoir ; OBJ#cpos=V ; OBJ#fpos=V ;
OBJ#m=ind ; OBJ#n=p ; OBJ#p=3 ; OBJ#t=pst ;
OBJ#wordform=qu ; OBJ#lemma=que ; OBJ#cpos=V ; OBJ#fpos=V ;
OBJ#m=ind ; OBJ#s=s
OBJ#wordform=poursuivies ; OBJ#lemma=poursuivre ; OBJ#cpos=V ;
OBJ#fpos=VPP ; OBJ#g=f ; OBJ#m=part ; OBJ#n=p ; OBJ#t=past ;

La représentation ***syn-srel-\$REL***

- contenu : cooccurents syntaxiques par relation syntaxique
- séparation des cooccurents de *syn-all* par relation syntaxique

*Les personnes commettant des fraudes doivent **comprendre** qu ' elles seront poursuivies .*

syn-srel-OBJ :

OBJ#wordform=doivent ; OBJ#lemma=devoir ; OBJ#cpos=V ; OBJ#fpos=V ;
OBJ#m=ind ; OBJ#n=p ; OBJ#p=3 ; OBJ#t=pst ;

OBJ#wordform=qu ; OBJ#lemma=que ; OBJ#cpos=V ; OBJ#fpos=V ;
OBJ#m=ind ; OBJ#s=s

OBJ#wordform=poursuivies ; OBJ#lemma=poursuivre ; OBJ#cpos=V ;
OBJ#fpos=VPP ; OBJ#g=f ; OBJ#m=part ; OBJ#n=p ; OBJ#t=past ;

syn-srel-SUJ

SUJ#wordform=personnes ; SUJ#lemma=personne ; SUJ#cpos=N ;
SUJ#fpos=NC ; SUJ#g=f ; SUJ#n=p

La représentation *ngb*

- contenu : voisins (noms, verbes, adjectifs et adverbes)
- vecteur de descripteurs : *caractéristique=valeur*

*Les personnes commettant des fraudes doivent **comprendre** qu ' elles seront poursuivies .*

rank	worform	lemma	cpos	fpos	g	m	n	p	s	t	grank	srel
1	Les	le	D	DET	-	-	p	-	def	-	2	det
2	personnes	personne	N	NC	f	-	p	-	neg	-	6	subj
3	commettant	commettant	A	ADJ	-	-	-	-	qual	-	2	mod
4	des	de	P+D	P+D	-	-	-	-	def	-	2	dep
5	fraudes	fraude	N	NC	f	-	p	-	c	-	4	obj
6	doivent	devoir	V	V	-	ind	p	3	-	pst	0	obj
7	comprendre	comprendre	V	VINF	-	inf	-	-	-	-	6	obj
8	qu'	que	C	CS	-	-	-	-	s	-	7	obj
9	elles	cln	CL	CLS	-	-	-	-	subj	-	11	subj
10	seront	être	V	V	-	ind	p	3	-	fut	11	aux_pass
11	poursuivies	poursuivre	V	VPP	f	part	p	-	-	past	8	obj
21	.	.	PONCT	PONCT	-	-	-	-	s	-	6	ponct

Légende : *la cible* ; *ses voisins (N, V, A, ADV)*

La représentation *ngb*

- contenu : voisins (noms, verbes, adjectifs et adverbes)
- vecteur de descripteurs : *caractéristique=valeur*

*Les personnes commettant des fraudes doivent **comprendre** qu ' elles seront poursuivies .*

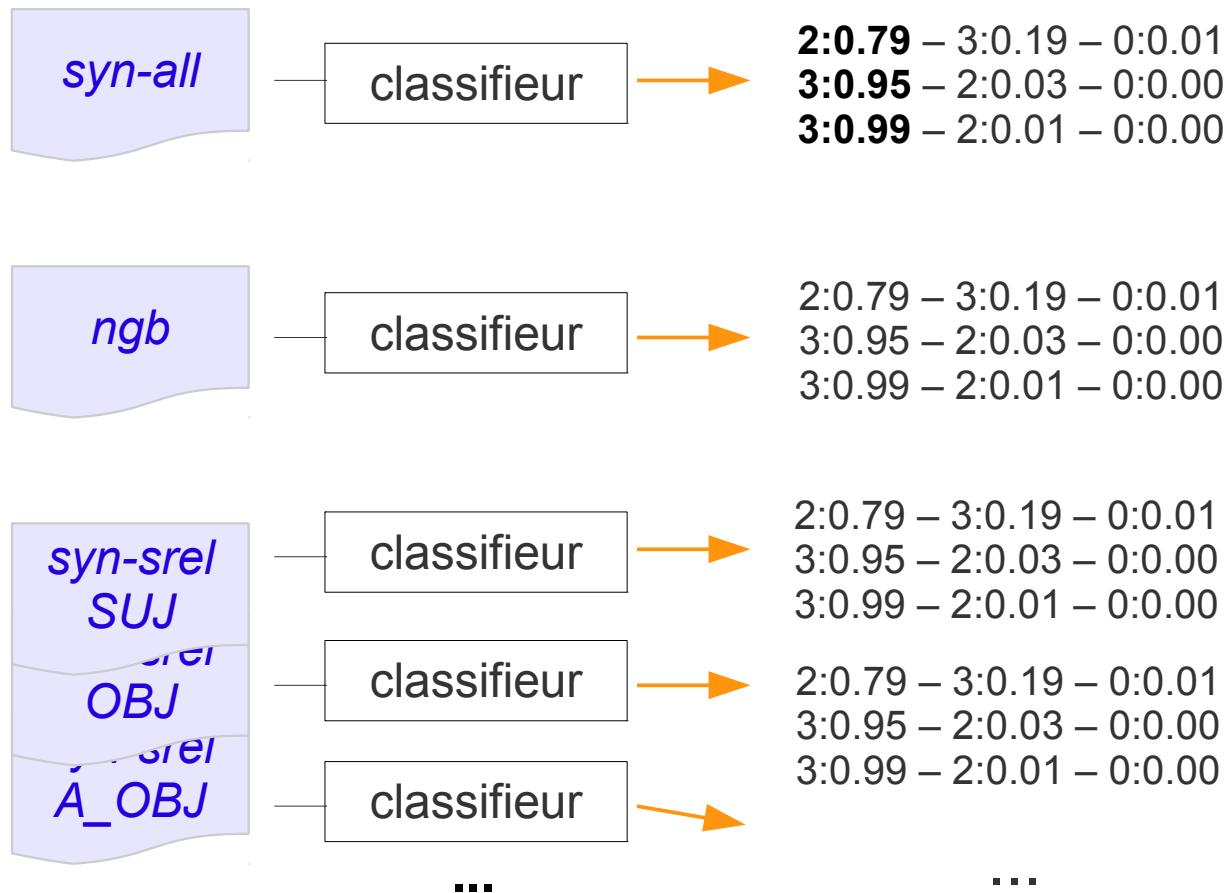
*wordform=personnes ; lemma=personne ; cpos=N ; fpos=NC ; g=f ; n=p
wordform=commettant ; lemma=commettre ; cpos=A ; fpos=ADJ ; s=qual ;
wordform=fraudes ; lemma=fraude ; cpos=N ; fpos=NC ; g=f ; n=p ; s=c
wordform=doivent ; lemma=devoir ; cpos=V ; fpos=V ; m=ind ; n=p ; p=3 ; t=pst ;
wordform=seront ; lemma=être ; cpos=V ; fpos=V ; m=ind ; n=p ; p=3 ; t=fut ;
wordform=poursuivies ; lemma=poursuivre ; cpos=V ; fpos=VPP ; g=f ; m=part ;
n=p ; t=past ;*

Liste des représentations des contextes d'un mot

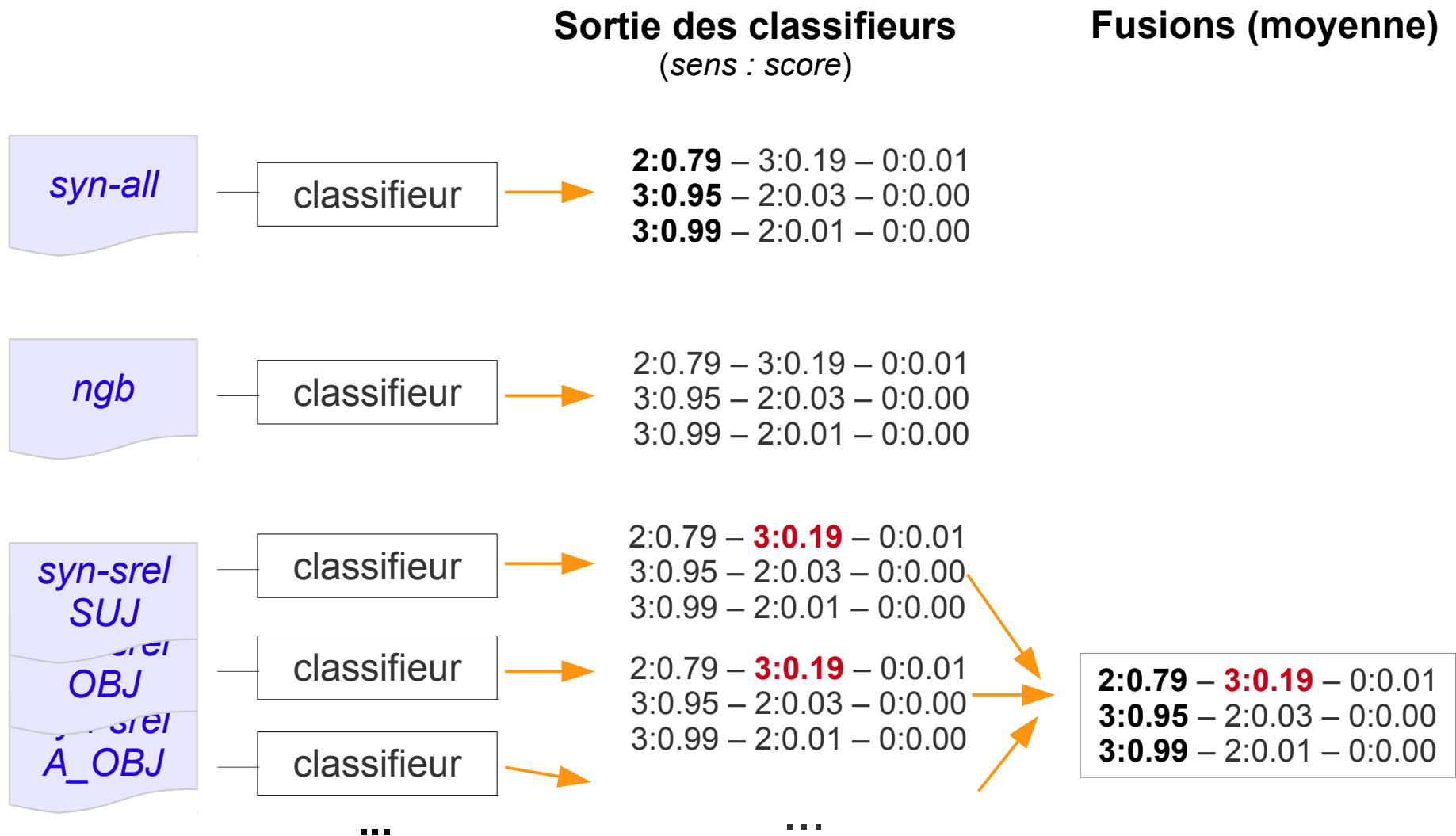
	REPRÉSENTATION	CONTENU
1	<i>ngb</i>	les voisins (noms, verbes, adjectifs et adverbes)
2	<i>syn-all</i>	les cooccurrents syntaxiques toutes relations syntaxiques confondues
3	<i>syn-srel-suj</i>	le sujet du verbe (ou de son auxiliaire)
4	<i>syn-srel-obj</i>	le(s) complément(s)
5	<i>syn-srel-de_obj</i>	compléments d'objet indirects (non locatifs) introduits par la préposition <i>de</i>
6	<i>syn-srel-a_obj</i>	compléments d'objet indirects (non locatifs) introduits par la préposition <i>à</i>
7	<i>syn-srel-p_obj</i>	autres compléments d'objet indirects
8	<i>syn-srel-ats</i>	attributs du sujet
9	<i>syn-srel-ato</i>	attributs de l'objet
10	<i>syn-srel-mod</i>	modifieurs (pour les noms, les relatives en sont exclues)
11	<i>syn-srel-mod_rel</i>	relatives adnominales (pour les noms)
12	<i>syn-srel-aux_tps</i>	auxiliaires de temps
13	<i>syn-srel-aux_pass</i>	auxiliaires du passif
14	<i>syn-srel-aux_caus</i>	verbes causatifs (forme verbale composée d'un causatif suivi d'un infinitif)
15	<i>syn-srel-aff</i>	clitiques figés
16	<i>syn-srel-coord</i>	relation portée par un coordonnant
17	<i>syn-srel-dep_coord</i>	relation portée par un coordonné
18	<i>syn-srel-arg</i>	utilisé dans le cas de prépositions « liées » (pour les noms)
19	<i>syn-srel-det</i>	déterminants (pour les noms)
20	<i>syn-srel-dep</i>	dépendants prépositionnels (sans distinction argument/ajout) (pour les noms)

Les classifieurs

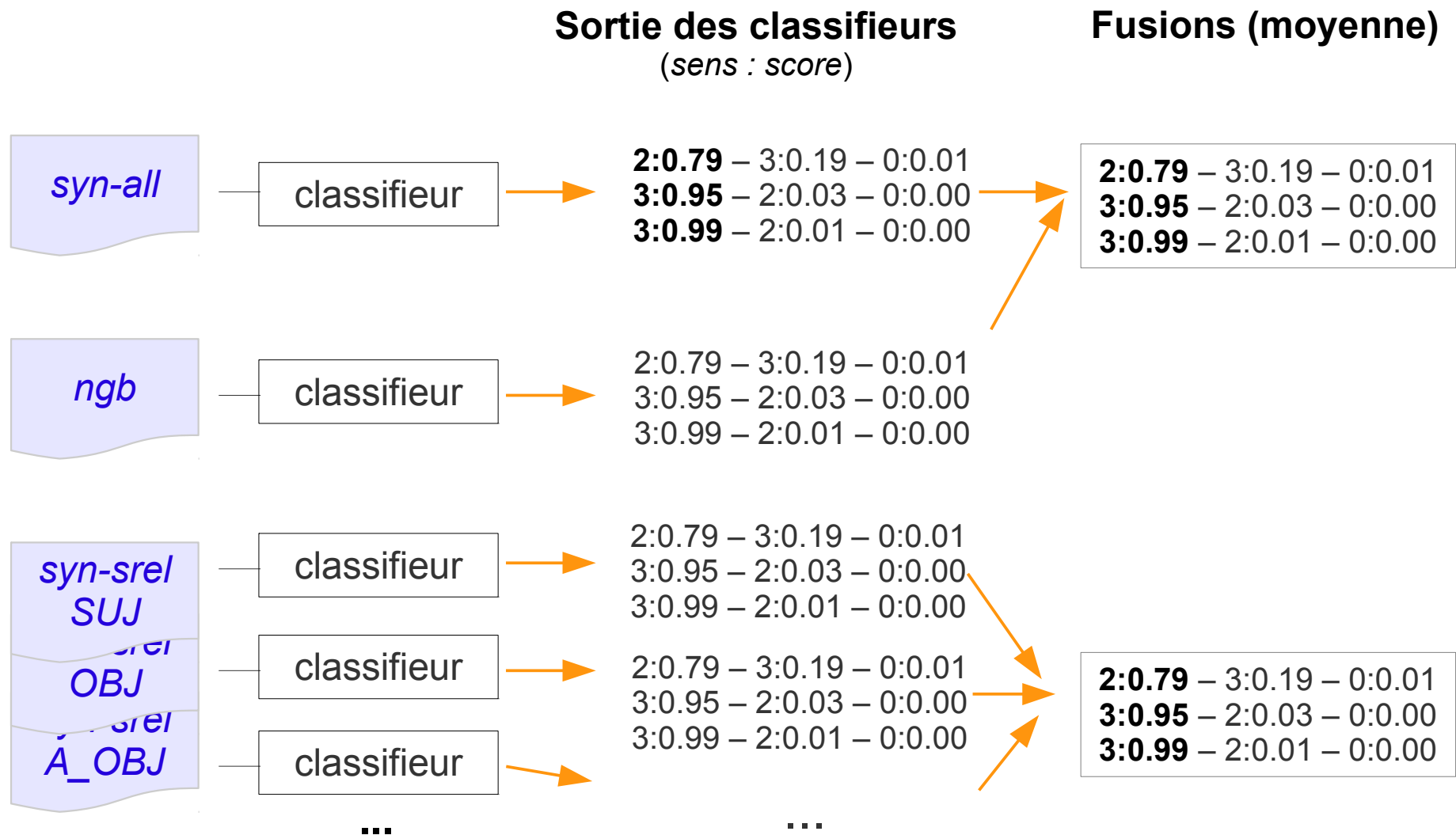
Sortie des classifieurs (sens : score)



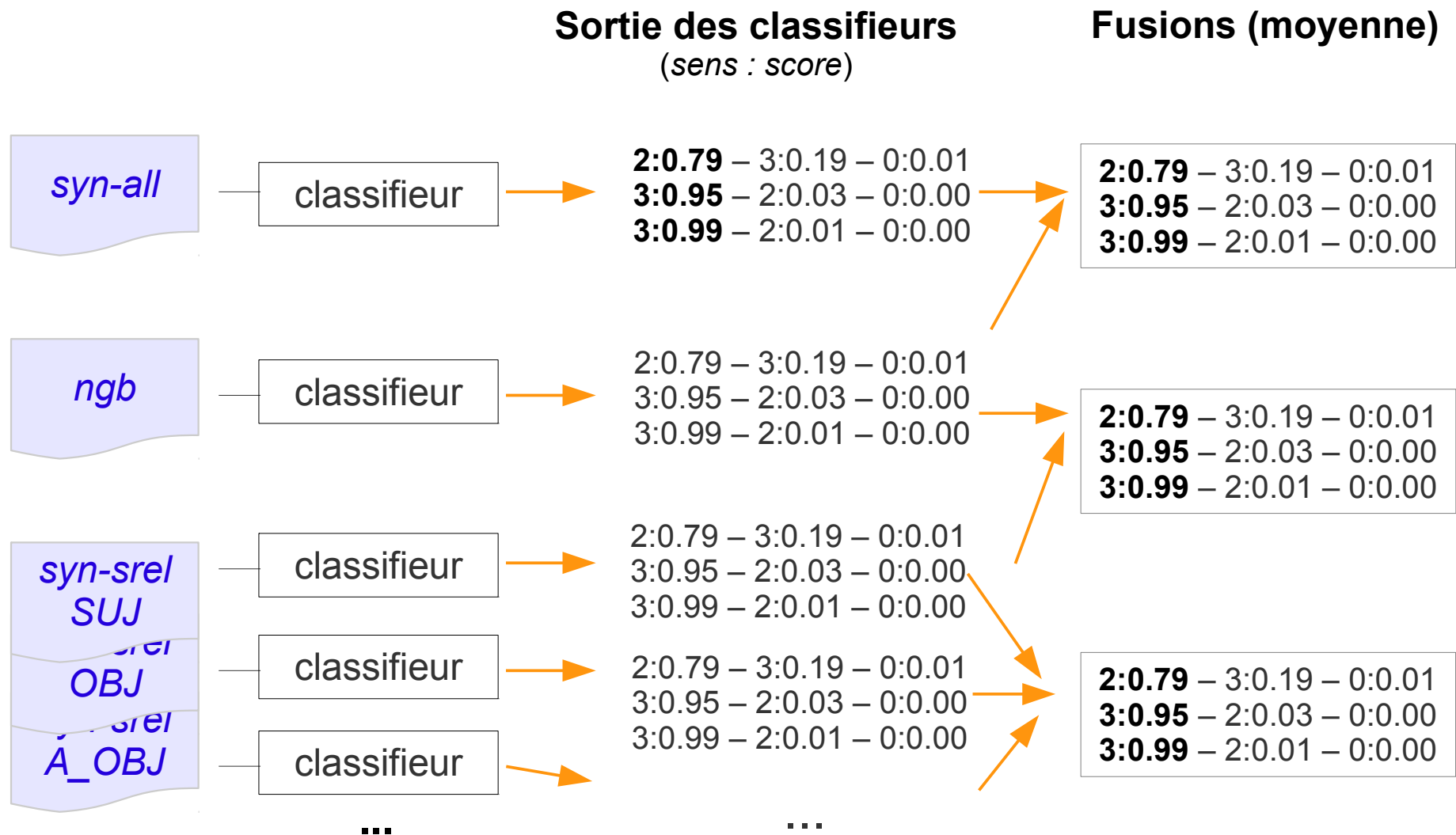
Les classifieurs



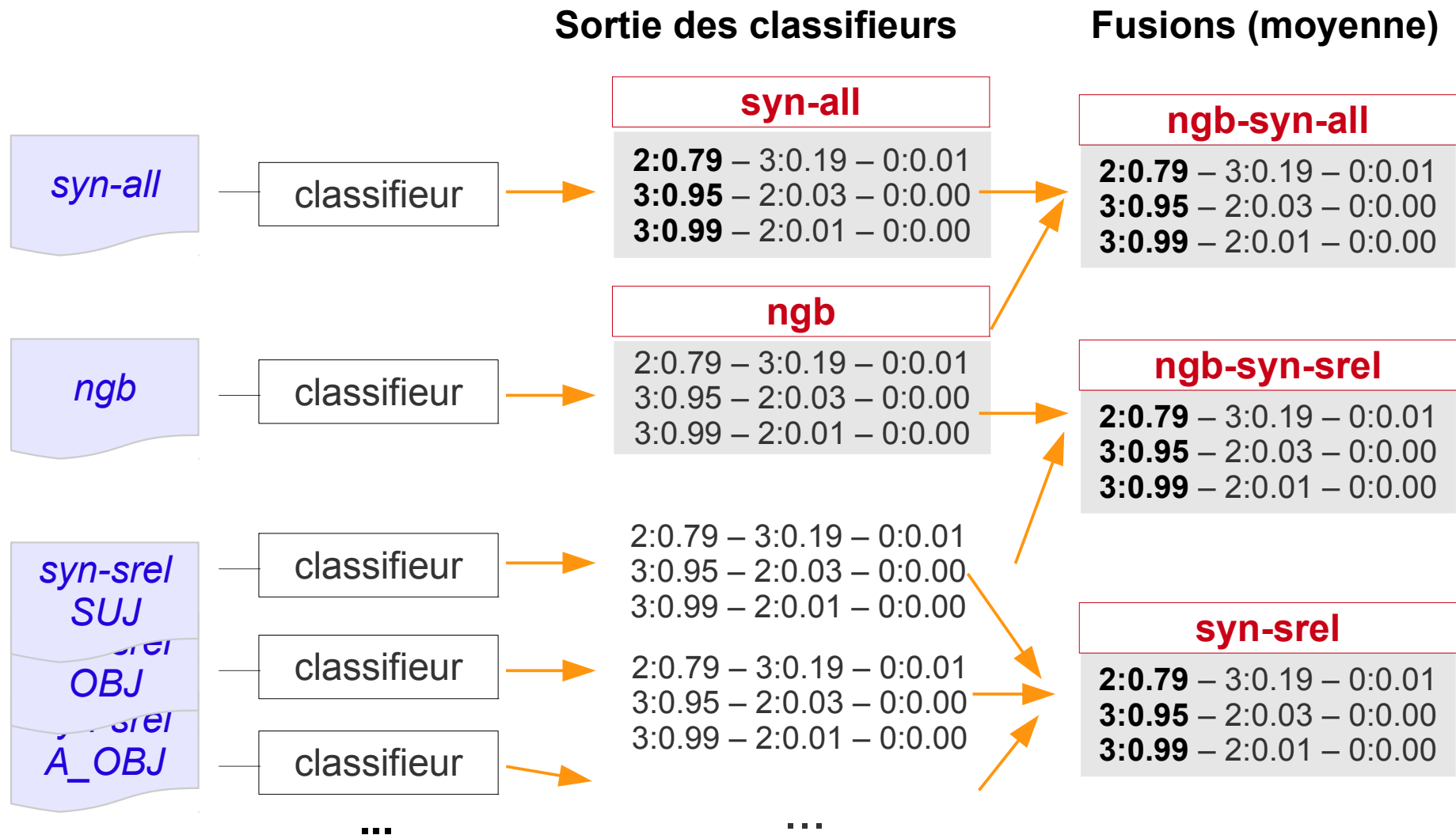
Les classifieurs



Les classifieurs



Les classifieurs



Mesures d'évaluation

Précision = nombre de prédictions correctes / nombre de prédictions

Taux d'erreur = $1 - \text{Précision}$

Évaluation des représentations (précision)

		entropie	ngb	ngb-syn-all	ngb-syn-srel	syn-all	syn-srel
Emplois LG	<i>arrêter</i>	-2.4456	63.50	78.50	99.00	66.00	99.00
	<i>comprendre</i>	-1.6525	90.50	92.94	94.52	86.72	94.52
	<i>importer</i>	-1.0553	98.56	94.24	99.28	91.01	99.28
Traductions	<i>arrêter</i>	-4.7112	37.00	39.21	52.42	37.44	52.42
	<i>comprendre</i>	-2.9051	59.57	60.76	80.14	60.40	80.14
	<i>importer</i>	-2.9645	67.38	73.48	86.38	70.61	86.38
Lexies	<i>arrêter</i>	-5.6398	18.75	21.67	47.50	21.25	47.50
	<i>comprendre</i>	-3.9419	53.71	56.96	67.52	47.68	67.52
	<i>importer</i>	-3.1450	69.04	70.82	77.94	63.70	77.94

→ meilleures performances : ***syn-srel***

Évaluation des représentations (précision)

		entropie	ngb	ngb-syn-all	ngb-syn-srel	syn-all	syn-srel
Emplois LG	<i>arrêter</i>	-2.4456	63.50	78.50	99.00	66.00	99.00
	<i>comprendre</i>	-1.6525	90.50	92.94	94.52	86.72	94.52
	<i>importer</i>	-1.0553	98.56	94.24	99.28	91.01	99.28
Traductions	<i>arrêter</i>	-4.7112	37.00	39.21	52.42	37.44	52.42
	<i>comprendre</i>	-2.9051	59.57	60.76	80.14	60.40	80.14
	<i>importer</i>	-2.9645	67.38	73.48	86.38	70.61	86.38
Lexies	<i>arrêter</i>	-5.6398	18.75	21.67	47.50	21.25	47.50
	<i>comprendre</i>	-3.9419	53.71	56.96	67.52	47.68	67.52
	<i>importer</i>	-3.1450	69.04	70.82	77.94	63.70	77.94

→ meilleures performances : **syn-srel** et **ngb-syn-srel**

Évaluation des représentations (précision)

		entropie	ngb	ngb-syn-all	ngb-syn-srel	syn-all	syn-srel
Emplois LG	<i>arrêter</i>	-2.4456	63.50	78.50	99.00	66.00	99.00
	<i>comprendre</i>	-1.6525	90.50	92.94	94.52	86.72	94.52
	<i>importer</i>	-1.0553	98.56	94.24	99.28	91.01	99.28
Traductions	<i>arrêter</i>	-4.7112	37.00	39.21	52.42	37.44	52.42
	<i>comprendre</i>	-2.9051	59.57	60.76	80.14	60.40	80.14
	<i>importer</i>	-2.9645	67.38	73.48	86.38	70.61	86.38
Lexies	<i>arrêter</i>	-5.6398	18.75	21.67	47.50	21.25	47.50
	<i>comprendre</i>	-3.9419	53.71	56.96	67.52	47.68	67.52
	<i>importer</i>	-3.1450	69.04	70.82	77.94	63.70	77.94

- meilleures performances : **syn-srel** et **ngb-syn-srel**
- performances les plus faibles : **ngb** et **syn-all**

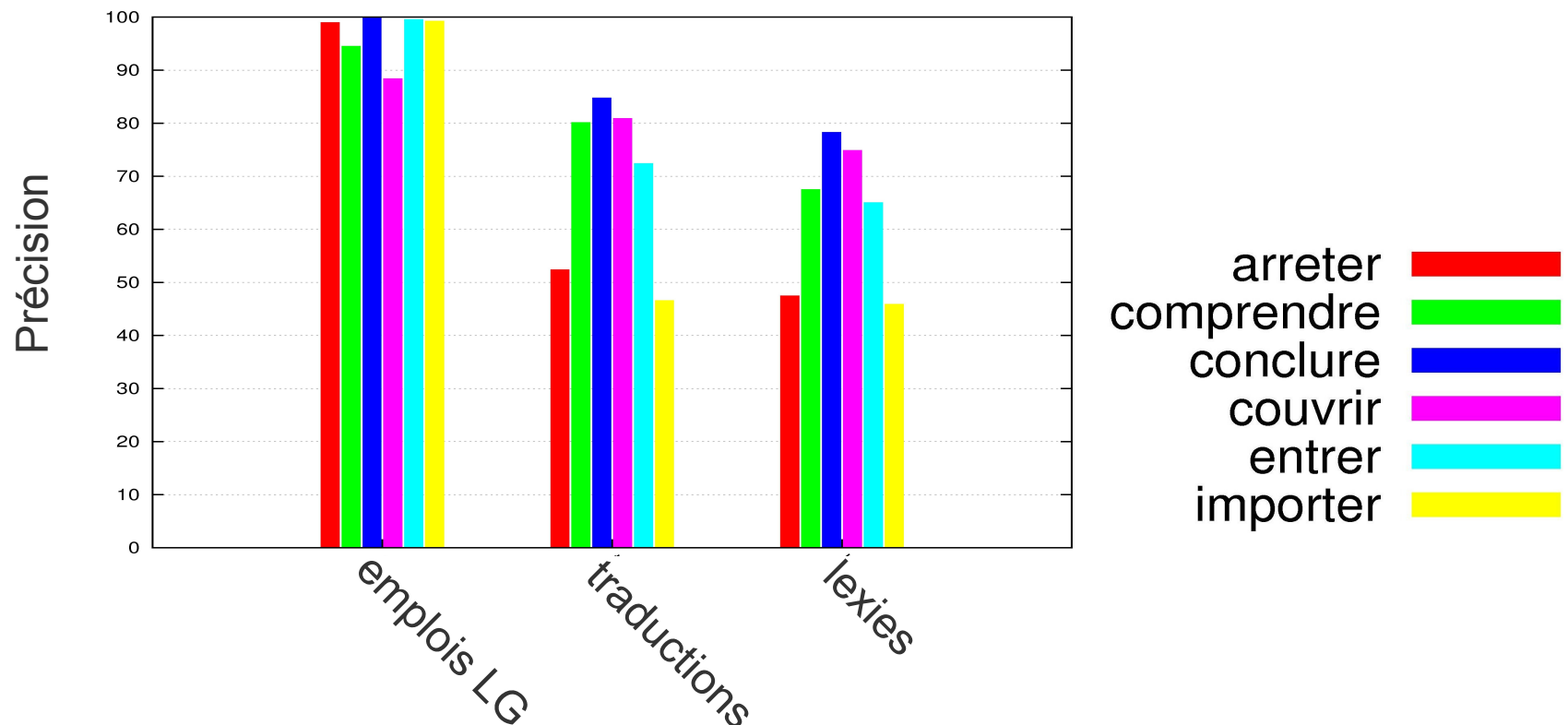
Évaluation des représentations (précision)

		entropie	ngb	ngb-syn-all	ngb-syn-srel	syn-all	syn-srel
Emplois LG	<i>arrêter</i>	-2.4456	63.50	78.50	99.00	66.00	99.00
	<i>comprendre</i>	-1.6525	90.50	92.94	94.52	86.72	94.52
	<i>importer</i>	-1.0553	98.56	94.24	99.28	91.01	99.28
Traductions	<i>arrêter</i>	-4.7112	37.00	39.21	52.42	37.44	52.42
	<i>comprendre</i>	-2.9051	59.57	60.76	80.14	60.40	80.14
	<i>importer</i>	-2.9645	67.38	73.48	86.38	70.61	86.38
Lexies	<i>arrêter</i>	-5.6398	18.75	21.67	47.50	21.25	47.50
	<i>comprendre</i>	-3.9419	53.71	56.96	67.52	47.68	67.52
	<i>importer</i>	-3.1450	69.04	70.82	77.94	63.70	77.94

- meilleures performances : **syn-srel** et **ngb-syn-srel**
- performances les plus faibles : **ngb** et **syn-all**
- grande différence de précision entre **syn-srel** et **syn-all**

Évaluation des étiquettes de sens

**Précision des classifieurs en fonction
des étiquettes de sens utilisées**
(représentation *syn-srel*)



→ Emplois LG > Traductions > Lexies

Conclusion

Perspectives

- compléter le corpus avec :
 - plus d'exemples pour les emplois LG sous-représentés
 - des exemples des emplois non représentés dans EuroParl
- construire un corpus annoté avec le Lexique-Grammaire pour :
 - 20 noms polysémiques
 - 20 adjectifs polysémiques

Bibliographie

LE LEXIQUE-GRAMMAIRE :

- BOONS, J.-P., GUILLET, A. et LECLÈRE, Ch. 1976a. *La structure des phrases simples en français. Constructions intransitives*. Genève, Droz
- BOONS, J.-P., GUILLET, A. et LECLÈRE, Ch. 1976b. *La structure des phrases simples en français. Classes de constructions transitives*. Rapport de Recherches du LADL n°6, Paris : Université Paris 7.
- GROSS, M. 1975. *Méthodes en syntaxe*. Paris : Hermann.
- GUILLET, A. et LECLÈRE, Ch. 1992. *La structure des phrases simples en français. Les constructions transitives locatives*. Genève, Droz.

Bibliographie

RESSOURCES ET OUTILS :

Collectif. 2009. *Dictionnaire HACHETTE*. Hachette Éducation.

CANDITO, M.-H. et CRABBÉ, B. Benchmarking of Statistical Dependency Parsers for French. *Proceedings of COLING'2010 (poster session)*, Beijing, Chine.

CHIH-CHUNG, Ch. and CHICH-JEN, L. 2011. LIBSVM : a library for support vector machines. *ACM Transaction on Intelligent Systems and Technology*, 2:3, pp. 21-27.

CRABBE, B. et CANDITO, M.-H. 2008. Analyse syntaxique du français : des constituants aux dépendances. *Proceedings of TALN 2009*, Senlis, France.

KOEHN, P. 2005. Europarl : A Parallel Corpus for Statistical Machine Translation. *Proceedings of MT Summit X*, pp. 79-86, Phuket, Thailand.

OCH, F. J. et NEY, H.. 2003. Improved Statistical Alignment Models. *Proceedings of the 38-th Annual Meeting for Computational Linguistics*, pp. 440-447, Hong-Kong, Chine.

Fin !