

est l'annotation morphosyntaxique, où le but est de trouver la nature de chaque mot. Prenons la phrase « Yoann Dupont fait une thèse à Paris 3. » :

Yoann	Dupont	fait	une	thèse	à	Paris	3	.
↓	↓	↓	↓	↓	↓	↓	↓	↓
nom-p	nom-p	verbe	dét	nom-c	prép	nom-p	nom-p	ponct

FIGURE 3.1 – exemple d'étiquetage morphosyntaxique

Il est également possible de simuler l'annotation des données par groupes. Pour ce faire, différents marqueurs sont concaténés à l'étiquette : B (Begin) marquant le début d'un groupe nominal, I (In) marquant l'appartenance à un groupe précédemment commencé et O (Out) marque tout ce qui n'appartient à aucun groupe. Il n'y a pas de marqueur de fin explicite car la fin d'un groupe se déduit soit par le début d'un autre groupe soit par une arrivée sur un non-groupe. Ce schéma d'annotation est régulièrement utilisé car il s'agit du plus simple capable de représenter de façon exacte les groupes. Dans l'exemple suivant, nous effectuons un étiquetage en chunks :

Yoann	Dupont	fait	une	thèse	à	Paris	3	.
↓	↓	↓	↓	↓	↓	↓	↓	↓
B-NP	I-NP	B-VN	B-NP	I-NP	B-PP	I-PP	I-PP	O

FIGURE 3.2 – exemple d'étiquetage morphosyntaxique

L'annotation en entités nommées suit alors le même principe et la phrase peut être annotée de la façon suivante :

Yoann	Dupont	fait	une	thèse	à	Paris	3	.
↓	↓	↓	↓	↓	↓	↓	↓	↓
B-Personne	I-Personne	O	O	O	O	B-Organisation	I-Organisation	O

FIGURE 3.3 – exemple d'étiquetage morphosyntaxique



3.2 Systèmes à base de règles

3.2.1 Les outils Luxid ®

Expert System France dispose d'outils permettant de gérer des annotateurs à base de règles, appelés cartouches de connaissance, de leur création à la validation de leur