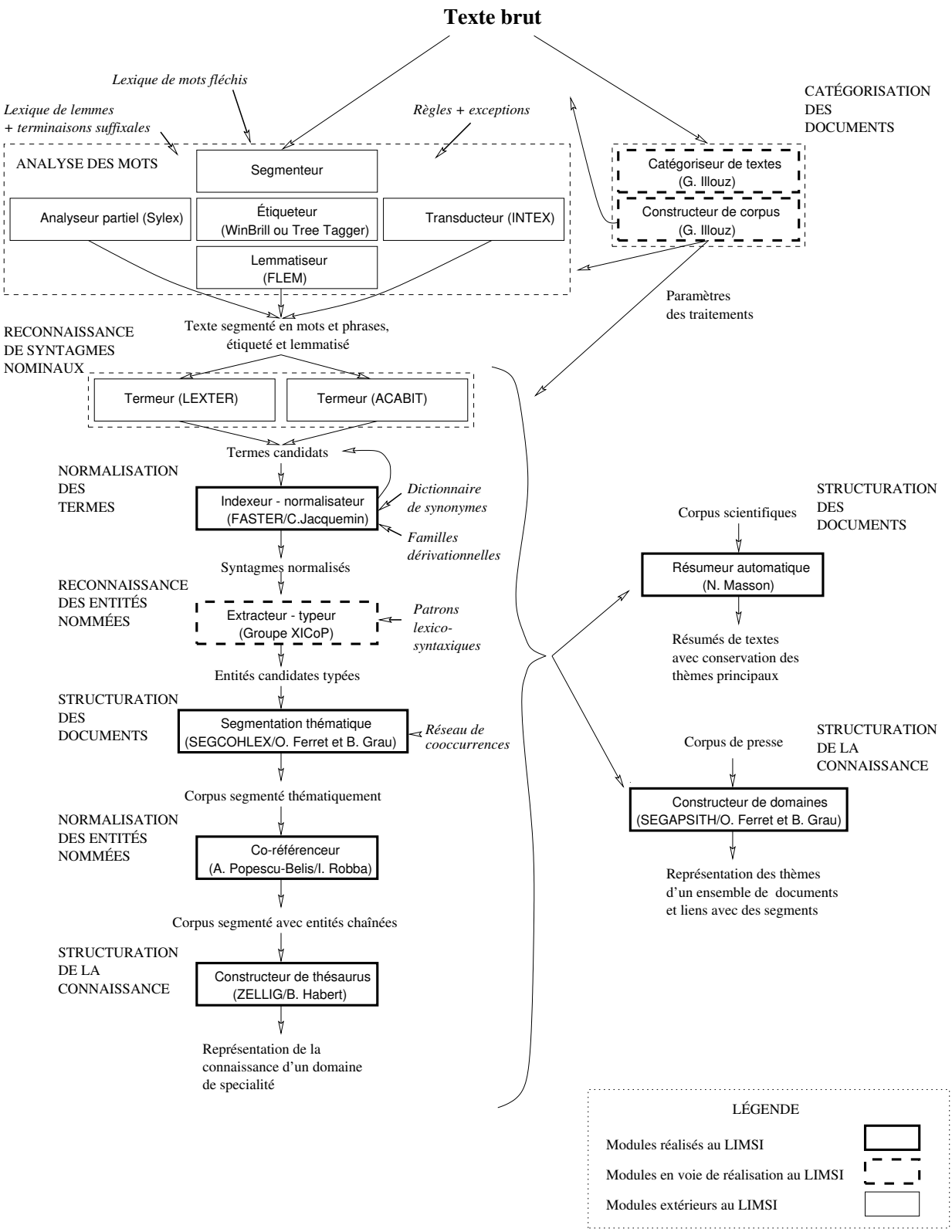
Texte brut
Lexique de mots fléchis
Lexique de lemmes
+ terminaisons suffixales
Règles + exceptions
CATÉGORISATION
DES
DOCUMENTS
ANALYSE DES MOTS
Segmenteur
Analyseur partiel (Sylex)
Étiqueteur
(WinBrill ou Tree Tagger)
Transducteur (INTEX)
Lemmatiseur
(FLEM)
Catégoriseur de textes
(G. Illouz)
Constructeur de corpus
(G. Illouz)
Paramètres
des traitements
RECONNAISSANCE
DE SYNTAGMES
NOMINAUX
Texte segmenté en mots et phrases,
étiqueté et lemmatisé
Termeur (LEXTER)
Termeur (ACABIT)
Termes candidats
NORMALISATION
DES
TERMES
Indexeur - normalisateur
(FASTER/C.Jacquemin)
Dictionnaire
de synonymes
Familles
dérivationnelles
Syntagmes normalisés
RECONNAISSANCE
DES ENTITÉS
NOMMÉES
Extracteur - typeur
(Groupe XICoP)
Patrons
lexico-
syntaxiques
Entités candidates typées
STRUCTURATION
DES
DOCUMENTS
Segmentation thématique
(SEGCOHLEX/O. Ferret et B. Grau)
Réseau de
cooccurrences
Corpus segmenté thématiquement
NORMALISATION
DES ENTITÉS
NOMMÉES
Co-référenceur
(A. Popescu-Belis/I. Robba)
Corpus segmenté avec entités chaînées
STRUCTURATION
DE LA
CONNAISSANCE
Constructeur de thésaurus
(ZELLIG/B. Habert)
Représentation de la
connaissance d'un domaine
de specialité
Corpus scientifiques
STRUCTURATION
DES
DOCUMENTS
Résumeur automatique
(N. Masson)
Résumés de textes
avec conservation des
thèmes principaux
Corpus de presse
STRUCTURATION
DE LA
CONNAISSANCE
Constructeur de domaines
(SEGAPSITH/O. Ferret et B. Grau)
Représentation des thèmes
d'un ensemble de documents
et liens avec des segments
LÉGENDE
Modules réalisés au LIMSI
Modules en voie de réalisation au LIMSI
Modules extérieurs au LIMSI