

Structuration et balisage sémantique des définitions du Trésor de la Langue Française informatisé

Le projet Definiens vise la conversion des définitions du Trésor de la Langue Française informatisé sous une forme structurée. Il n'existe pas à l'heure actuelle pour le français de base de données lexicales, libre d'accès et à large couverture, proposant pour chaque unité lexicale décrite une définition explicitement structurée. Definiens entreprend de combler ce manque en indiquant pour chaque définition que compte le TLFi (environ 270 000):

1. la structure en composantes définitionnelles de type genre prochain et différences spécifiques ;
2. le rôle joué par chaque composante dans la caractérisation du sens de l'unité lexicale définie.

La définition de BROUETTE, illustrée ci-dessous, sera ainsi analysée de la façon suivante : la composante centrale (CC) de la définition sera étiquetée *véhicule* et ses composantes périphériques (CP) seront étiquetés par des rôles indiquant les propriétés principales de ce véhicule, en l'occurrence ses *parties caractéristiques* et sa *fonction*.

*BROUETTE (acception B.1) *= *Véhicule à une roue et à deux brancards servant au transport des matériaux.

BROUETTE (acception B.1)

<PARAPH>

<CC étiquette=*véhicule*>Véhicule</CC>

<CP rôle=*parties caractéristiques*>à une roue et à deux brancards</CP>

<CP rôle=*fonction*>servant au transport des matériaux</CP>

</PARAPH>

Je me concentrerai principalement, au cours de cet exposé, sur la méthode employée pour l'analyse semi-automatique des définitions à l'aide du logiciel MACAON.