

Chapitre 2

WordNet

# Chapitre 2: WordNet

## 2.1 Introduction

---

Dans ce chapitre nous allons décrire WordNet et nous donnons quelques interfaces utilisateurs permettant la consultation de ce dernier.

## 2.2 Description de WordNet

---

**WordNet** est une base de données lexicale développée par des linguistes de l'université de Princeton pour la langue anglaise. Des versions de WordNet pour d'autres langues existent, cependant la version anglaise est la plus complète à ce jour. La base de données ainsi que des outils pour sa manipulation sont disponibles pour téléchargement gratuit.

WordNet repose sur la notion de **synset** (synonym set), un groupe de mots interchangeables, dénotant un sens ou un usage particulier. Par exemple dans la version 2.1 de WordNet le nom commun anglais car est décrit à l'aide de cinq synsets (voir figure 2.1):

The noun car has 5 senses (first 3 from tagged texts)

1. (598) car, auto, automobile, machine, motorcar -- (a motor vehicle with four wheels; usually propelled by an internal combustion engine; "he needs a car to get to work")
2. (24) car, railcar, railway car, railroad car -- (a wheeled vehicle adapted to the rails of railroad; "three cars had jumped the rails")
3. (1) cable car, car -- (a conveyance for passengers or freight on a cable railway; "they took a cable car to the top of the mountain")
4. car, gondola -- (the compartment that is suspended from an airship and that carries personnel and the cargo and the power plant)
5. car, elevator car -- (where passengers ride up and down; "the car was on the top floor")

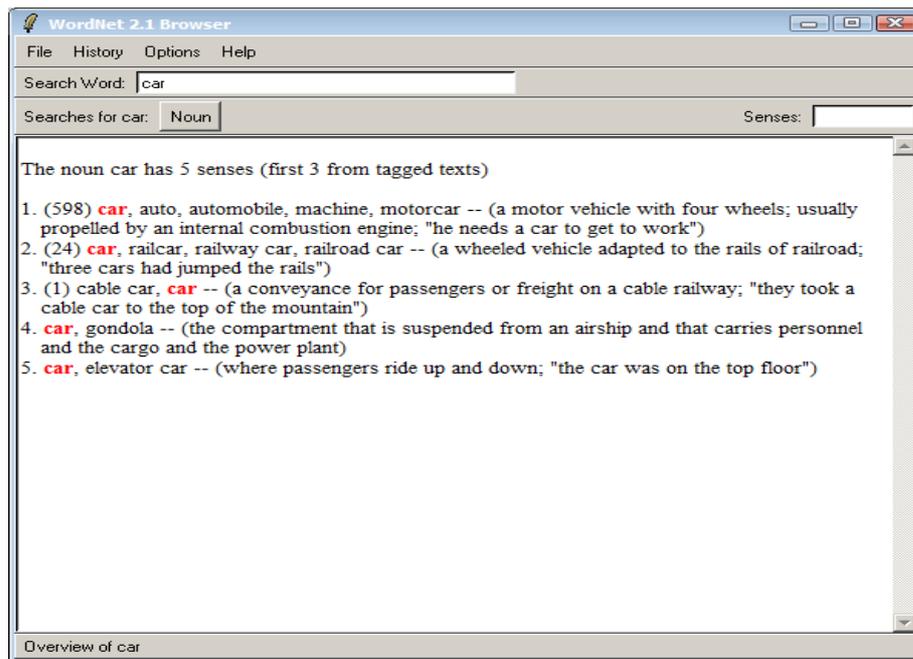


Figure 2.1 : le nom « car » dans WordNet

Nous pouvons trouver dans WordNet plusieurs informations (voir figure 2.2) par exemple les synonymes (par ordre de fréquence), hyponymes, meronymes...

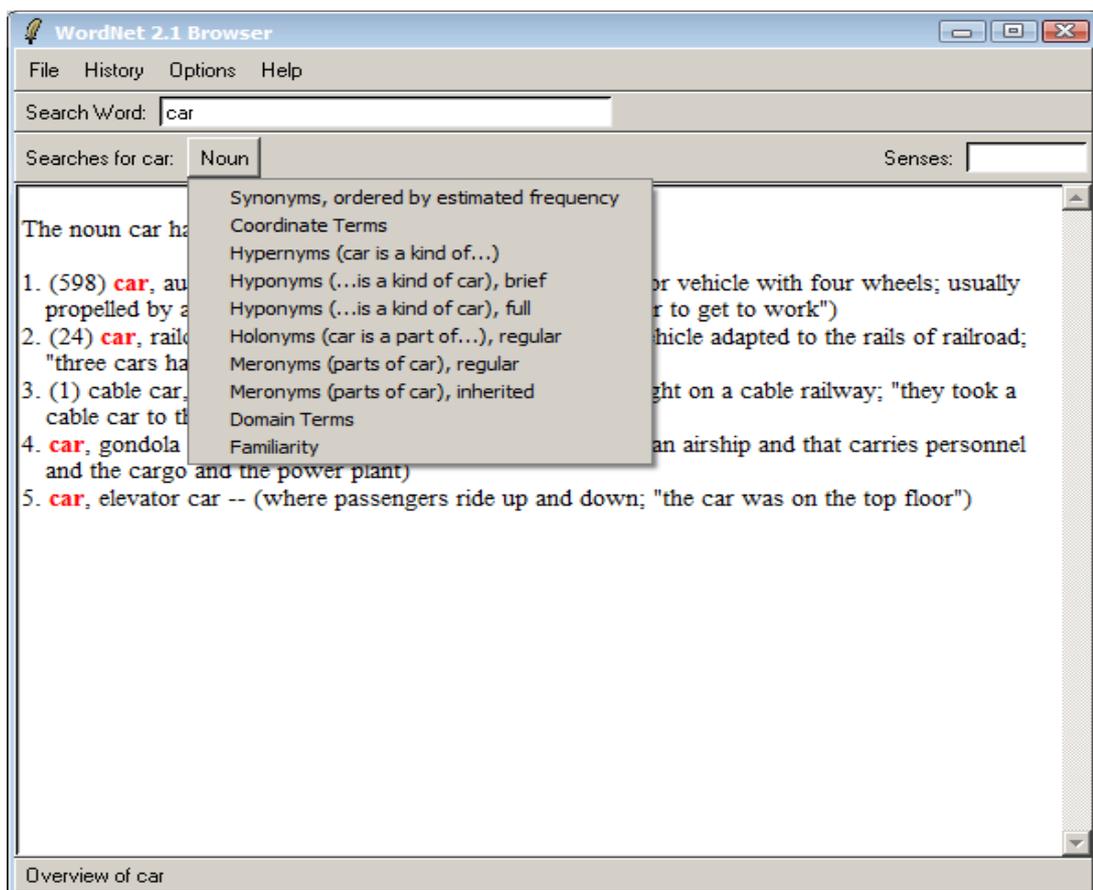


Figure 2.2 Plusieurs informations dans WordNet

La figure 2.3 donne les synonymes (par ordre de fréquence) de « car ».

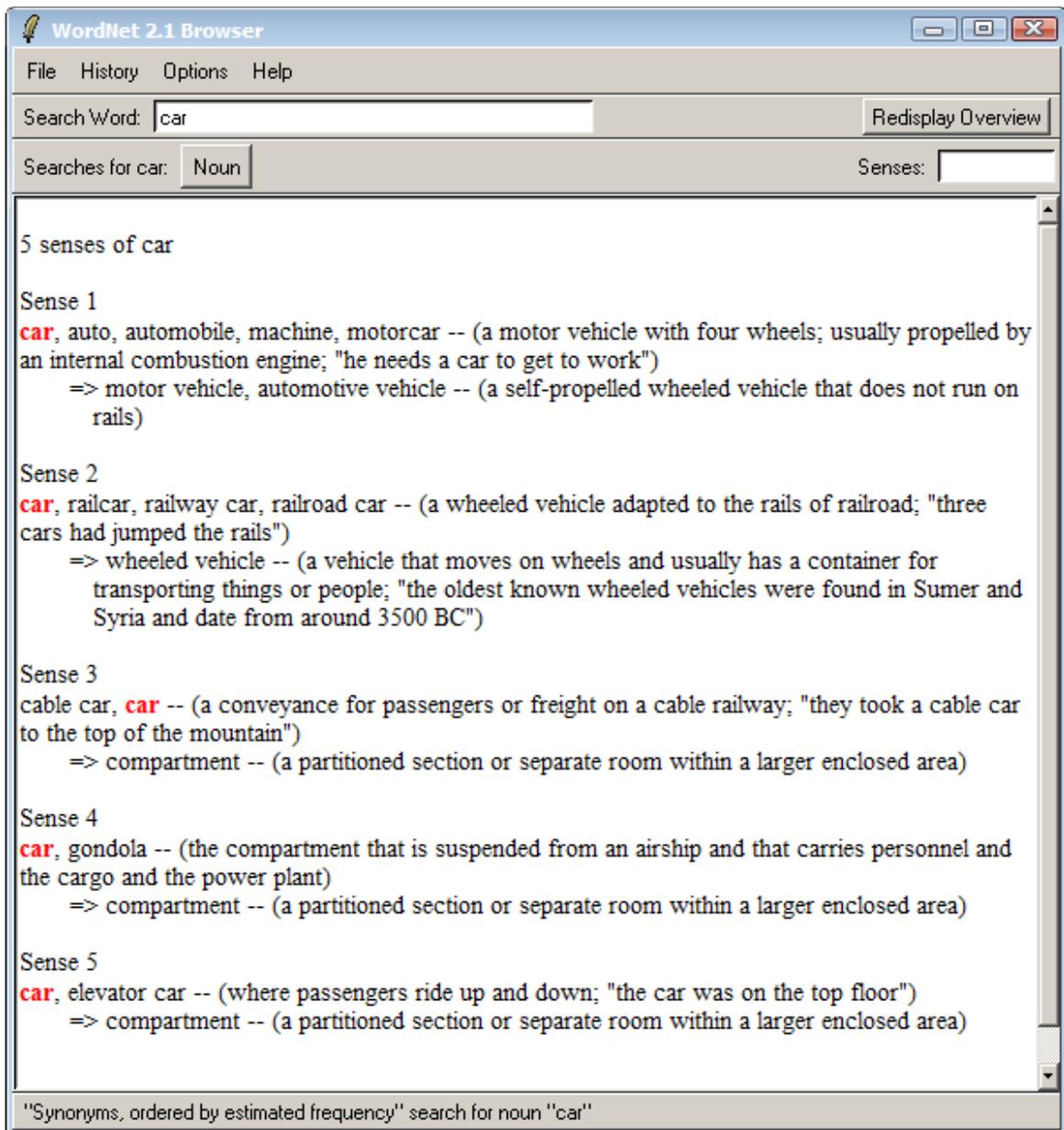


Figure 2.3 Les synonymes (par ordre de fréquence) de « car ».

WordNet répertorie une grande variété de relations sémantiques permettant d'organiser le sens des mots, par exemple à partir du sens le plus commun du mot car (correspondant au synset 1. car, auto...) la relation d'hyponymie définit un arbre de concepts (voir figure 2.4).

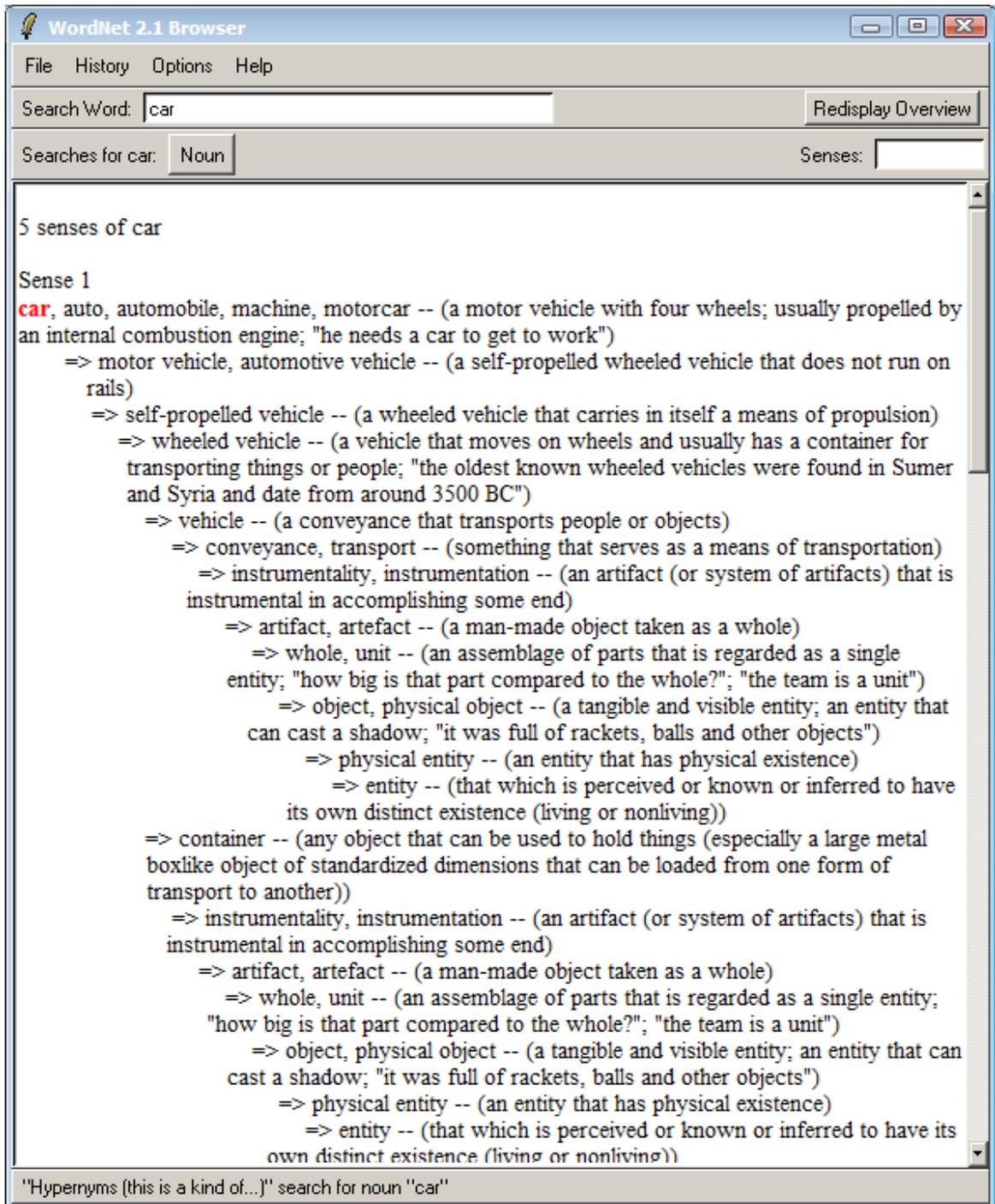


Figure 2.4 La relation d'hyponymie à partir de « car »

Nous pouvons voir aussi la relation inverse de l'hyponymie, l'hyponymie (voir figure 2.5).

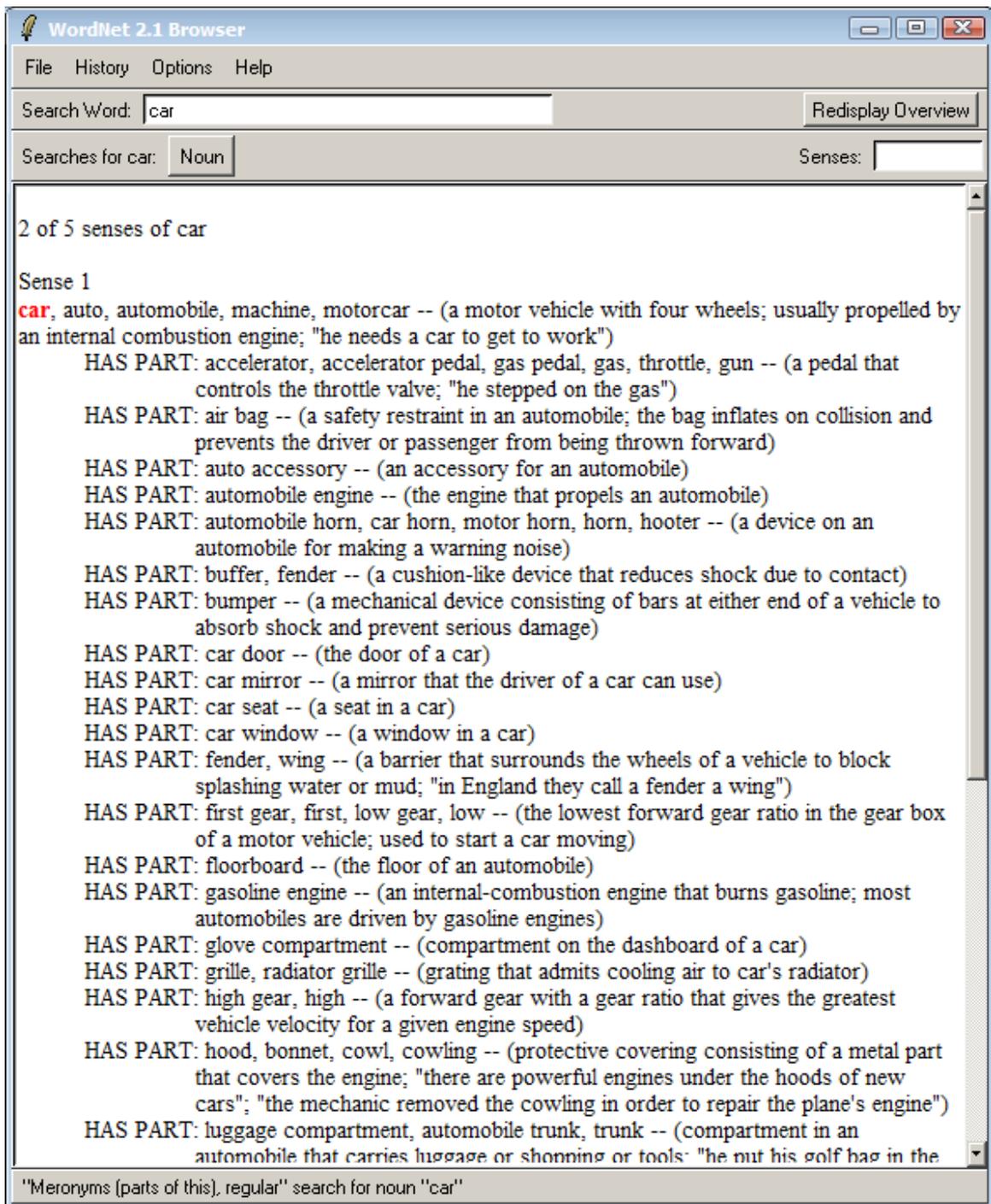


Figure 2.5 La relation l'hyponymie à partir de « car »

Il faut noter que WordNet se compose de quatre grandes catégories lexicales : les noms, les verbes, les adjectifs et les adverbes.

Il existe des projets similaires à WordNet anglais pour d'autres langues, la Global Wordnet Association (GWA<sup>1</sup>) recense une centaine de "Wordnet" dans différentes langues à travers le monde.

Nous pouvons voir dans les caractéristiques suivantes (voir figure 2.6)

Wordnet	Lang	Synsets	Words	Senses	Core	Licence	Data	Citation
<a href="#">Albanet</a>	<a href="#">als</a>	4,676	5,990	9,602	31%	<a href="#">CC BY 3.0</a>	<a href="#">als.zip</a>	<a href="#">cite:als</a>
<a href="#">Arabic WordNet (AWN)</a>	<a href="#">arb</a>	10,165	14,595	21,751	48%	<a href="#">CC BY SA 3.0</a>	<a href="#">arb.zip</a>	<a href="#">cite:arb</a>
<a href="#">Chinese Wordnet (Taiwan)</a>	<a href="#">cmn</a>	4,913	3,206	8,069	28%	<a href="#">wordnet</a>	<a href="#">cmn.zip</a>	<a href="#">cite:cmn</a>
<a href="#">Dan.Net</a>	<a href="#">dan</a>	4,476	4,468	5,859	81%	<a href="#">wordnet</a>	<a href="#">dan.zip</a>	<a href="#">cite:dan</a>
<a href="#">Princeton WordNet</a>	<a href="#">eng</a>	117,659	148,730	206,978	100%	<a href="#">wordnet</a>	<a href="#">eng.zip</a>	<a href="#">cite:eng</a>
<a href="#">Persian Wordnet</a>	<a href="#">fas</a>	17,759	17,560	30,461	41%	<a href="#">Free to use</a>	<a href="#">fas.zip</a>	<a href="#">cite:fas</a>
<a href="#">FinnWordNet</a>	<a href="#">fin</a>	116,763	129,839	189,227	100%	<a href="#">CC BY 3.0</a>	<a href="#">fin.zip</a>	<a href="#">cite:fin</a>
<a href="#">WOLF (Wordnet Libre du Français)</a>	<a href="#">fra</a>	59,091	55,373	102,671	92%	<a href="#">CeCILL-C</a>	<a href="#">fra.zip</a>	<a href="#">cite:fra</a>
<a href="#">Hebrew Wordnet</a>	<a href="#">heb</a>	5,448	5,325	6,872	27%	<a href="#">wordnet</a>	<a href="#">heb.zip</a>	<a href="#">cite:heb</a>
<a href="#">MultiWordNet</a>	<a href="#">ita</a>	34,728	40,343	61,558	83%	<a href="#">CC BY 3.0</a>	<a href="#">ita.zip</a>	<a href="#">cite:ita</a>
<a href="#">Japanese Wordnet</a>	<a href="#">jpn</a>	57,179	91,959	158,064	95%	<a href="#">wordnet</a>	<a href="#">jpn.zip</a>	<a href="#">cite:jpn</a>
<a href="#">Multilingual Central Repository</a>	<a href="#">cat</a>	45,826	46,531	70,622	81%	<a href="#">CC BY 3.0</a>	<a href="#">cat.zip</a>	<a href="#">cite:cat</a>
<a href="#">Multilingual Central Repository</a>	<a href="#">eus</a>	29,413	26,240	48,934	71%	<a href="#">CC BY-NC-SA 3.0</a>	<a href="#">eus.zip</a>	<a href="#">cite:eus</a>
<a href="#">Multilingual Central Repository</a>	<a href="#">glg</a>	19,312	23,124	27,138	36%	<a href="#">CC BY 3.0</a>	<a href="#">glg.zip</a>	<a href="#">cite:glg</a>

Figure 2.6 Les WordNet disponibles [CASTA 2013]

WordNet compte actuellement 115 000 synsets et plus de 200 000 mots [WIKI 2013].

<sup>1</sup> [http://www.globalwordnet.org/gwa/wordnet\\_table.html](http://www.globalwordnet.org/gwa/wordnet_table.html)

## 2.3 Les relations dans WordNet

---

Dans WordNet, le réseau de mots est exprimé par plusieurs relations sémantiques, on trouve parmi elles [BAZI, 2002]:

Relation **Synonymie** : le synset (synonym set), représente un ensemble de mots qui sont interchangeables dans un contexte donné. C'est le terme générique utilisé pour désigner une classe englobant des instances de classes plus spécifiques.

Relation **Hyponymie** : c'est le terme spécifique utilisé pour désigner un membre d'une classe (relation inverse de Hypernymie). **X** est un hyponyme de **Y** si **X** est un type de (kind of) **Y**.

Relation **Holonymie** : le nom de la classe globale dont les noms méronymes font partie. **Y** est un holonyme de **X** si **X** est une partie de (is a part of) **Y**.

Relation **Méronymie** : Le nom d'une partie constituante (part of), substance de (substance of) ou membre (member of) d'une autre classe (relation inverse de l'holonymie). **X** est un méronyme de **Y** si **X** est une partie de **Y**. exemple : {avion}

Has\_meronym {{porte}, {moteur}} ; {moteur} Has\_meronym {{hélice}, {réacteur}}.

WordNet possède aussi quelques autres relations qui sont moins utilisées en pratique. C'est le cas par exemple de la relation **antonym** pour exprimer les sens opposés pour les synsets, de la relation **entailment** pour les verbes : Un verbe **X entails** (nécessite) **Y** si **X** ne peut être fait à moins que **Y** ne le soit ou n'ait été déjà fait. Ou encore de la relation **troponym** pour les verbes : **X** est un troponyme de **Y** si **to X** est pareil que **to Y** dans une certaine façon.

WordNet peut être utilisé de deux façons, soit on utilisant un navigateur soit par programmation directe (il existe des API dans divers langages de programmation).

## 2.4 Exemples d'interfaces pour la consultation de WordNet

---

### a) Le site [WORD, 2013]

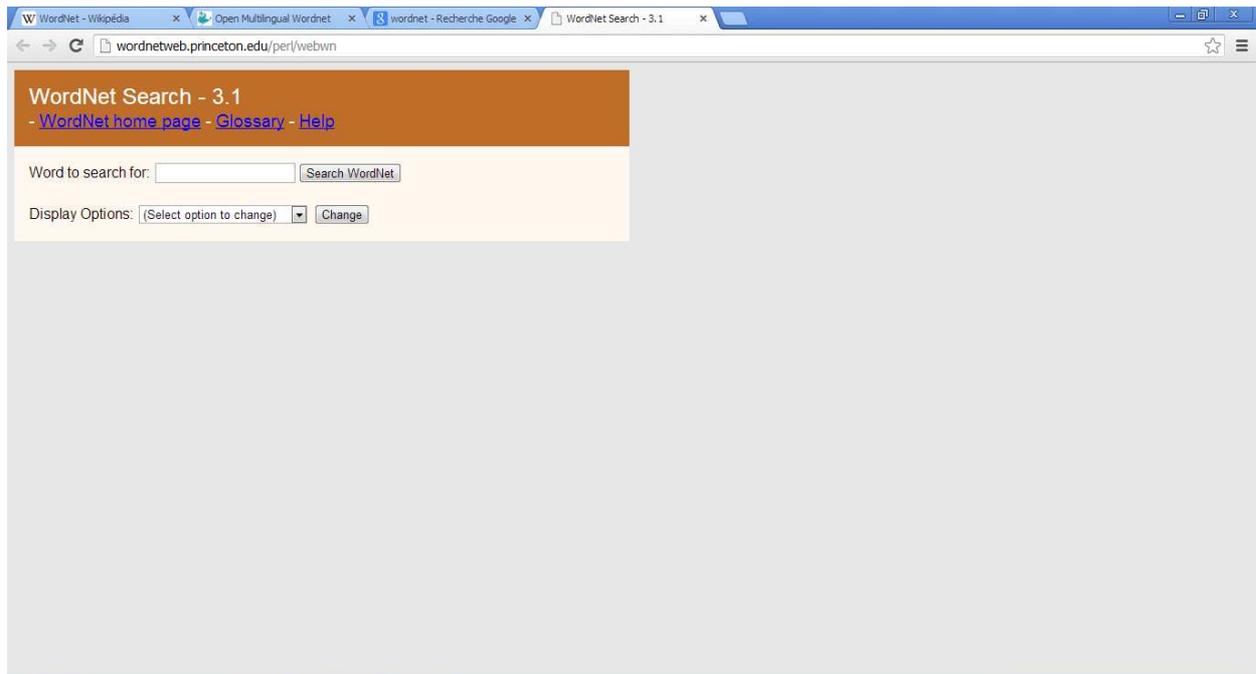


Figure 2.7 Interface de consultation de WordNet [WORD, 2013]

### b) Le site [VANC, 2013]

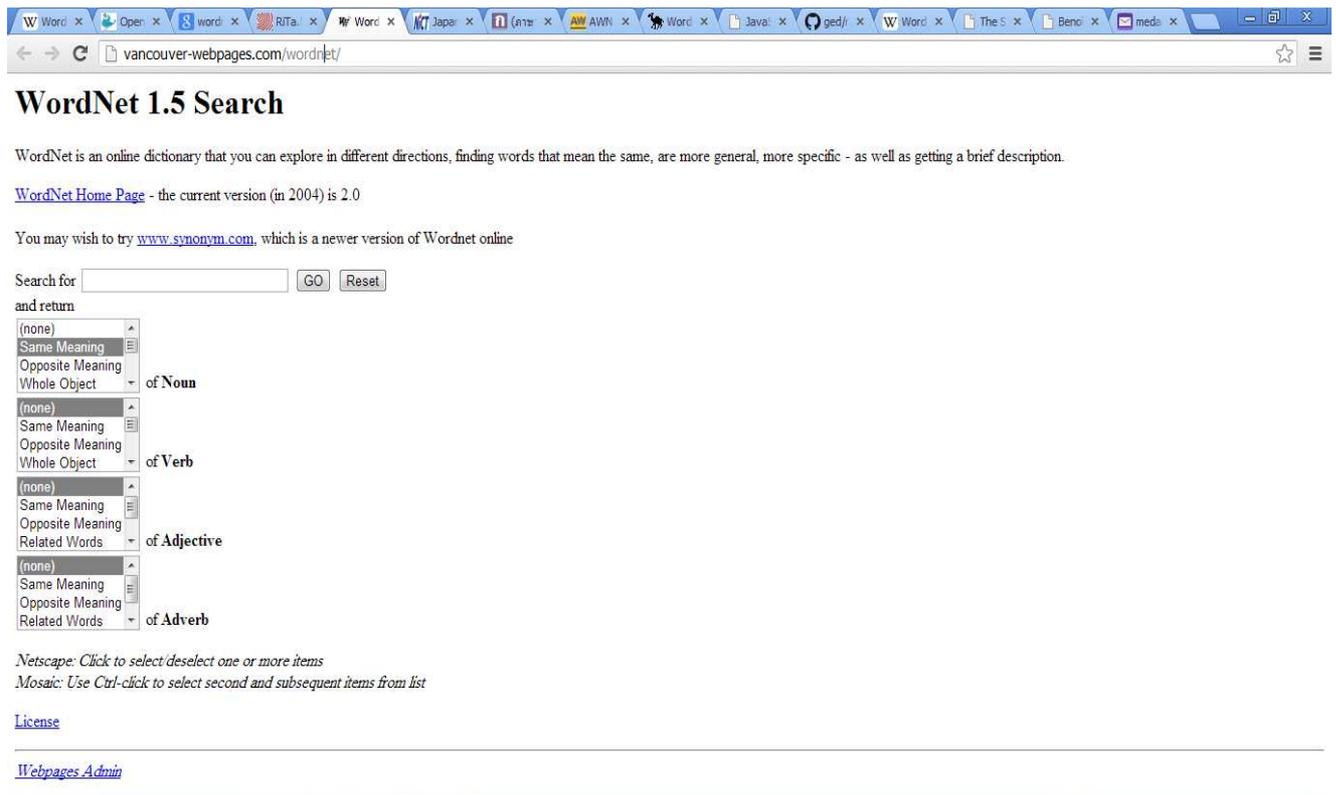


Figure 2.8 Interface de consultation de WordNet [VANC, 2013]

c) Le site [NLPW, 2013]

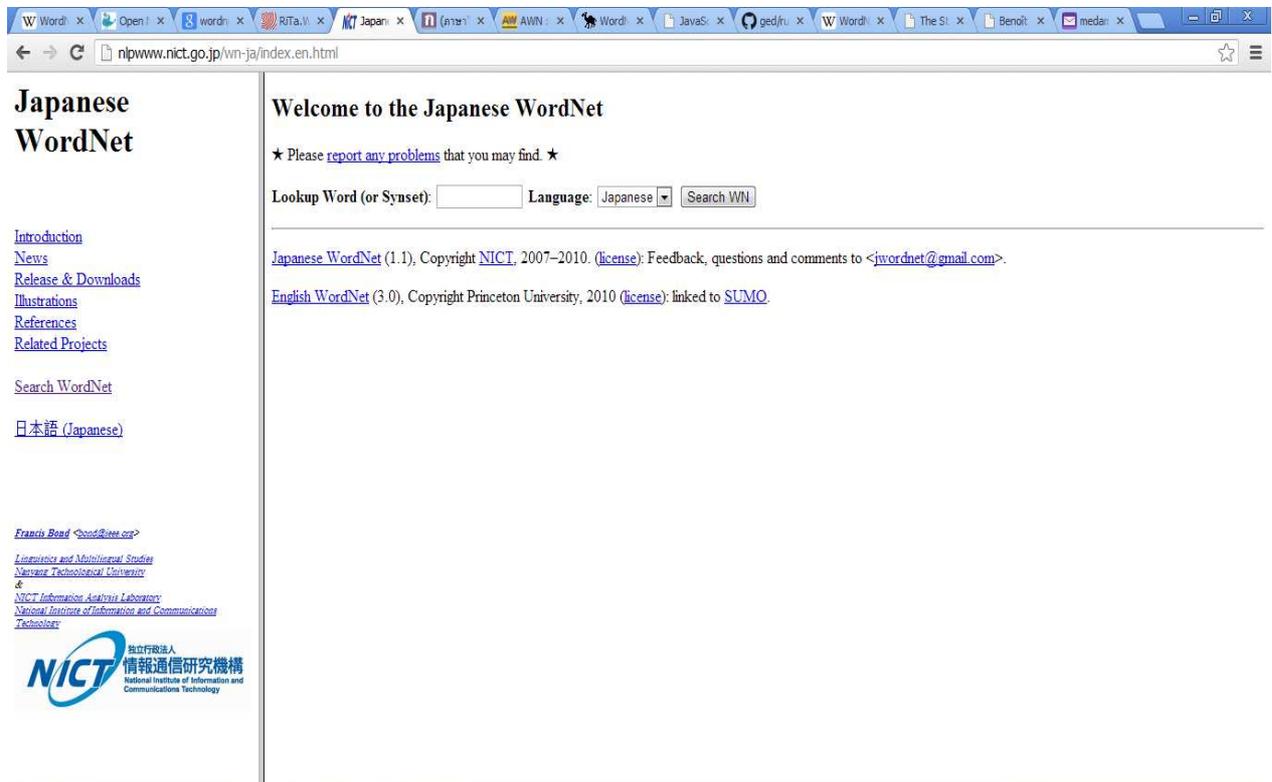


Figure 2.9 Interface de consultation de WordNet [NLPW, 2013]

d) Le site [ASIAN, 2013]

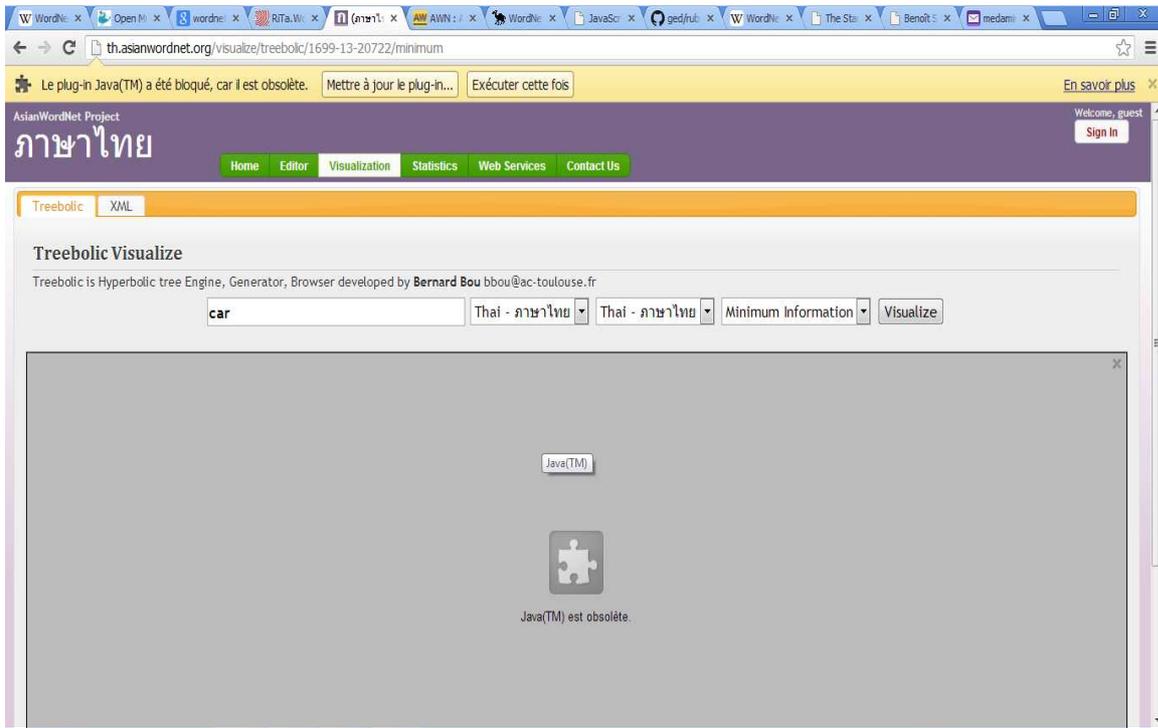


Figure 2.10 Interface de consultation de WordNet [ASIAN, 2013]

e) Le site [KYLE, 2013]

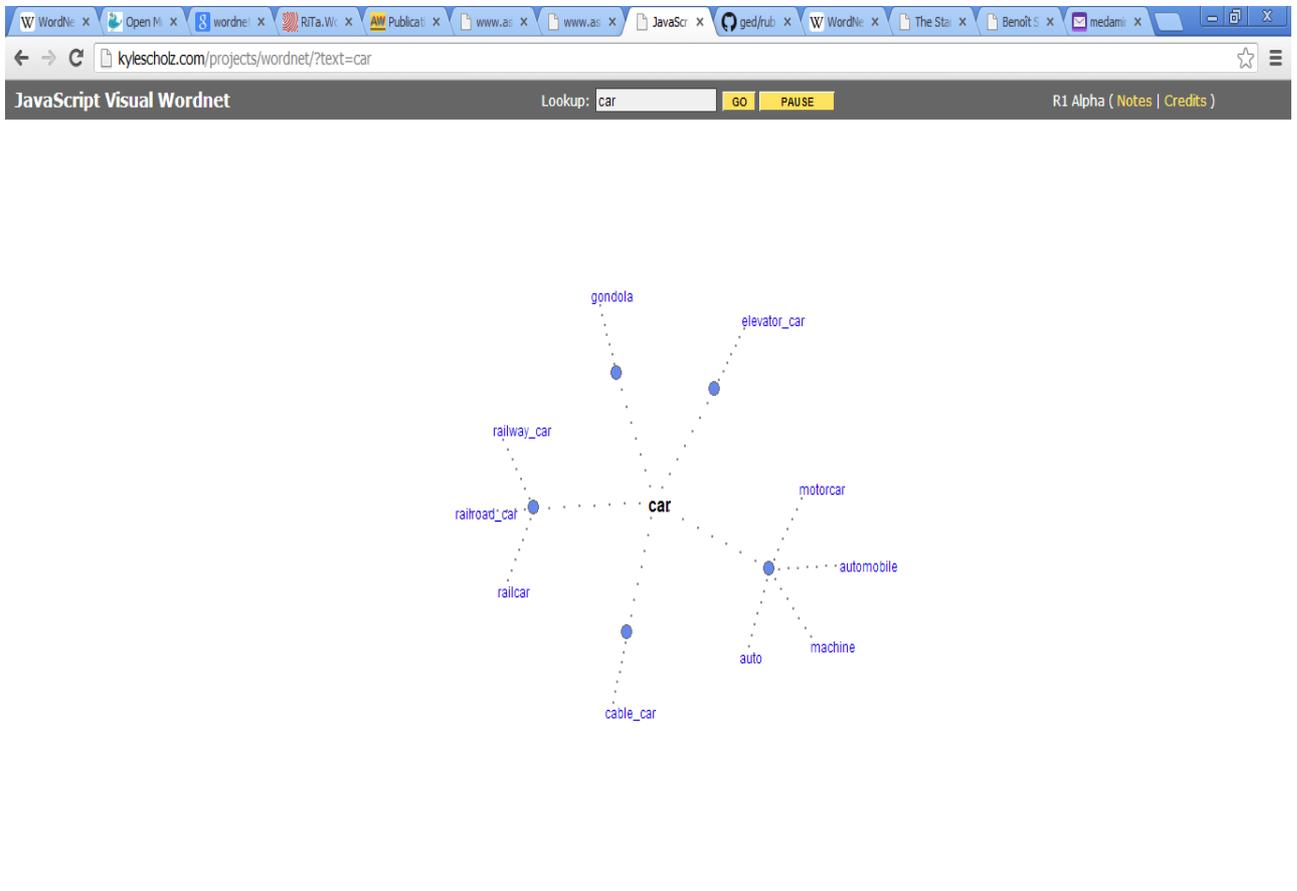


Figure 2.11 Interface de consultation de WordNet [KYLE, 2013]

f) Le site [CFIL, 2013]



Figure 2.12 Interface de consultation de WordNet [CFIL, 2013]

## 2.5 Conclusion

Dans ce chapitre nous avons fait une description de WordNet et nous avons donné des exemples d'implémentations des outils pour sa consultation. Dans le chapitre suivant nous allons détailler la structure du WordNet et décrire notre outil pour la construction d'un WordNet propre à la langue Arabe.