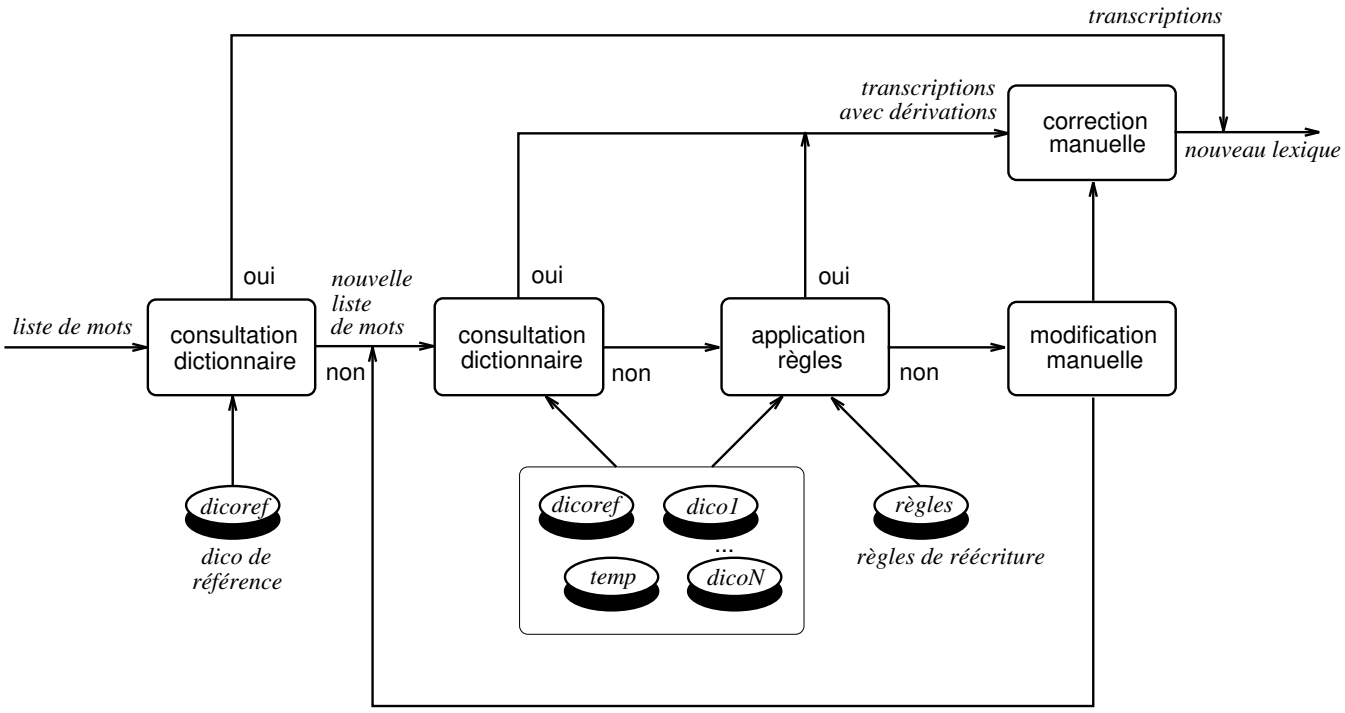

FIG. 1 - Schéma de la génération des prononciations d'un nouveau lexique. Dans cet outil, on infère la prononciation inconnue d'un mot, à partir de mots voisins dans différents dictionnaires, en appliquant des règles sur les suffixes et les préfixes. Dans le cas où plusieurs transcriptions sont possibles, l'expert valide le bon choix, après consultation de la manière dont ces choix ont été obtenus.

	Lexicon		
Test set	*20k*	*40k*	*65k*
Wall Street Journal	2.5	0.8	0.4
Frankfurter Rundschau	10.2	6.8	4.9
Le Monde	5.3	2.4	1.7

TAB. 1 - Taux de mots hors vocabulaire (% MHV) sur des corpus de journaux en anglo-américain (*Wall Street Journal*), français (*Le Monde*) and et allemand (*Frankfurter Rundschau*) en utilisant des lexiques de 20k, 40k, and 65k mots. le taux élevé de MHV en allemand est principalement dû à la composition de mots, ainsi que marginalement aux déclinaisons. Ce taux peut être réduit en utilisant une décomposition morphologique.

INTEREST	`IntrIst In{t}XIst`
CIVILIZATION	`sIvL[xY]zeSxn`
EXCUSE	`Ekskyu[sz]`
ANPFLANZUNG	`?An{p}flAn{t}sUG`
INSTALLIEREN	`?In[sS]tAlirXn`
MONTAGE	`mOntaZX montagX`

FIG. 2 - Exemples de prononciations multiples dans les lexiques anglo-américain et allemand, où {} désigne un phone optionnel, et [] un choix entre plusieurs phones possibles.