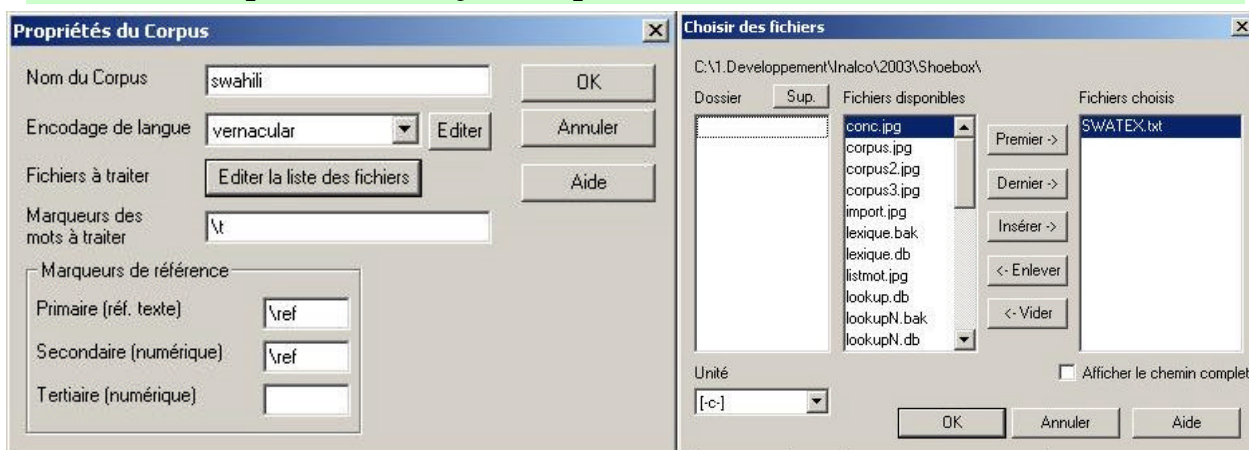


Création de la liste des mots d'un corpus

Il faut commencer par définir un corpus par dans Projet, Corpus et lui donner un nom. Le corpus est un ensemble de fichiers Shoebox pouvant être localisés à différents endroits sur le disque (et pas forcément ouverts dans le projet en cours). Pour ce corpus, il faut définir le nom des champs (séparés par un espace, s'ils sont plusieurs) d'où seront extraits les mots, ainsi que l'encodage de langue associé. Enfin on donnera le nom des champs devant servir à localiser les mots trouvés dans les différents textes. Dans le cas de phrases référencées par un champ \ref contenant une partie littérale suivie d'une partie numérique, comme par exemple \ref Art1 001 (phrase 1 du texte Art1), on choisira \ref comme marqueur de références primaire et secondaire. Si les références sont de type \id Art1 et \ref 001, on choisira \id comme marqueur de référence primaire et \ref comme marqueur de référence secondaire.

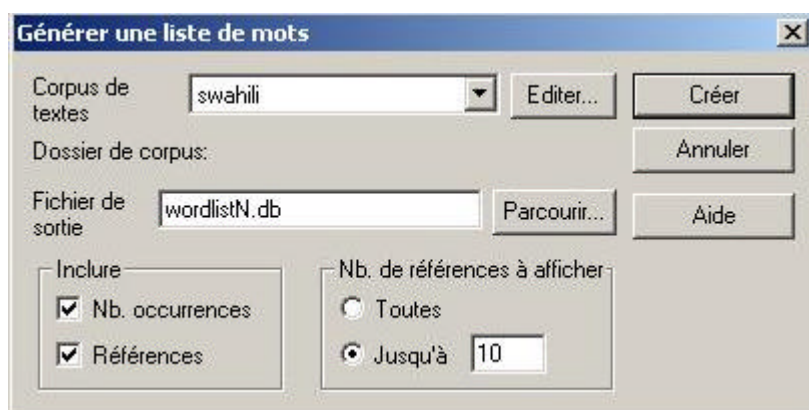
Définir le Corpus - dans Projet, Corpus

Editer la liste des fichiers à traiter



Générer la liste des mots du Corpus - dans Outils, Liste des mots

Ici, une base de données nommée "wordlistN.db" des mots du corpus "swahili" (précédemment créé) sera créée avec 10 références maximum par mots.



Liste des mots (base de données wordlistN.db)

Mot	Compte	Références
abdul	1	art1 013
adha	1	art1 103
afghanistan	3	art1 058; art1 083; art1 094
afrika	11	art1 011; art1 012; art1 013; art1 015; art1 016; art1 017; art1 018
agawanywe	2	art1 049; art1 052
aina	1	art1 063
aitwe	1	art1 058
akachukua	1	art1 031
akamkabidhi	1	art1 050
akasema	2	art1 049; art1 094
akatambua	1	art1 050
akatwe	3	art1 048; art1 049; art1 052
akielezea	1	art1 082
akifafanua	1	art1 072
akihusisha	1	art1 061

Extrait du fichier WordlistN.db

_sh v3.0 400 Liste de mots- Default

\w abdul	\cz 000001
\c 1	\r art1 063
\cz 000001	
\r art1 013	\w aitwe
	\c 1
\w adha	\cz 000001
\c 1	\r art1 058
\cz 000001	
\r art1 103	\w akachukua
	\c 1
\w afghanistan	\cz 000001
\c 3	\r art1 031
\cz 000003	
\r art1 058; art1 083; art1 094	\w akamkabidhi
	\c 1
\w afrika	\cz 000001
\c 11	\r art1 050
\cz 000011	
\r art1 011; art1 012; art1 013; art1 015; art1 016;	\w akasema
art1 017; art1 018	\c 2
\r art1 024; art1 033	\cz 000002
	\r art1 049; art1 094
\w agawanywe	
\c 2	\w akatambua
\cz 000002	\c 1
\r art1 049; art1 052	\cz 000001
	\r art1 050
\w aina	
\c 1	

Recherche de concordances - dans Outils, Concordance

Recherche des occurrences de "ny" dans le Corpus "swaili"

Générer une liste de concordance

Corpus de textes: swahili [Editer] [Rechercher]

Rechercher: ny [Annuler]

Correspondance:

- A l'interne
- A l'initiale
- En finale
- Totale

Casse distinctive

Nb. de correspondances à afficher:

- Toutes
- Jusqu'à 1000

Dossier de corpus: C:\1.Developpement\Inalco\2003\Shoebbox\

Fichier de sortie: lookup.db [Parcourir...]

[Aide]

Création de la liste des concordances de "ny" (base de données lookupN.db) et tri par rapport au champ Cible

Référence	Avant	Cible	Après
art1 049	kutapeli aliruka akasema akatwe	agawanywe	
art1 052	ule aliyetaka mtoto akatwe panga	agawanywe	
art1 060	dini wala itikadi, wanamapinduzi	hawafanyi	ugaidi bila kufikiri kiasi cha kutost
art1 056	Kuna matumaini kwamba ziara	iliyofanywa	na wajumbe wa Baraza la Usalar
art1 096	thman anasema ugaidi unaweza	kufanyika	na mtu binafsi, kikundi cha watu l
art1 096	undi cha watu lakini pia unaweza	kufanyika	na taifa na inapotokea hivyo, ugai
art1 073	181 (110) kukubali	kugawanywa	kwa Palestina na kuundwa kwa n
art1 075	isababisha kukubaliwa na mataifa	kugawanywa	kwa ardhi ya Wapalestina, kwa k
art1 001	"Siogopi mnachotaka	kunifanyia	kwani naamini kwa namna yoyote
art1 002	Lakini nasikitika kwa	kunifanyia	hivyo sasa, mtawatesa wananchi
art1 022	limu katika Azimio hilo ilikuwa ni	kuwafanya	Watanзания watambue kuwa war
art1 085	ka la Kijasusi la Marekani lilianza	kuwakusanya	watu kwa utekelezaji, Osama akiv
art1 086	nyingi kufanikisha vita hiyo kwa	kuwakusanya	watu wa kupigana kutoka nchi mt
art1 008	gine, imeyumba hadi kutumbukia	kwenye	vita vya wao kwa wao.
art1 009	abinafsi huku mamilionni ya watu	kwenye	nchi hiyo, wakibaki maskini.