

Méthodologie DILAF

Transformations du fichier hausa_extrait_v6.xml

Transformation	Rechercher	Remplacer par	Fichier résultat
balise <article>	<code>^<text:span text:style-name="T1">(.*?)</text:p>\$</code>	<code><article>\$1</article></code>	hausa_extrait_v7.xml
balisage entrée	<code><article>([^\>]*)</text:span></text:span></code>	<code><article><kalma>\$1</kalma></code>	hausa_extrait_v8.xml
suppression espaces inutiles qui suivent l'entrée	<code></kalma><text:span text:style-name="f_5f_standard"><text:span text:style-name="T1"> +<text:tab/></text:span></text:span></code>	<code></kalma></code>	hausa_extrait_v9.xml
balisage phonétique	<code><text:span text:style-name="Phonetic_20_form"><text:span text:style-name="T1">\ [([^\>]*) \]</text:span></text:span></code>	<code><furici>\$1</furici></code>	hausa_extrait_v10.xml
suppression espaces inutiles qui suivent la phonétique	<code></furici><text:span text:style-name="T1">+</text:span></code>	<code></furici></code>	hausa_extrait_v11.xml
balisage définition	<code><text:span text:style-name="Definition_20__28_v_29_"><text:span text:style-name="T1">([^\>]*)</text:span></text:span></code>	<code><ma_ana>\$1</ma_ana></code>	hausa_extrait_v12.xml
suppression espaces inutiles qui suivent la définition	<code></ma_ana><text:span text:style-name="T1">+</text:span></code>	<code></ma_ana></code>	hausa_extrait_v13.xml

balisage exemple	<text:span text:style-name="Example_20__28_v_29_"><text:span text:style-name="T1">([^<] *)</text:span></text:span>	<misali>\$1</misali>	hausa_extrait_v14.xml
suppression espaces inutiles qui suivent l'exemple	</misali><text:span text:style-name="T1">+</text:span>	</misali>	hausa_extrait_v15.xml
balisage équivalent français	<text:span text:style-name="f_5f_label"><text:span text:style-name="T1">Far\.:+</text:span></text:span><text:span text:style-name="Notes_20__28_questions_29_"><text:span text:style-name="T1">([^<] +)</text:span></text:span>	<makwatanci>\$1</makwatanci>	hausa_extrait_v16.xml
balisage classe	<text:span text:style-name="Part_20_of_20_speech"><text:span text:style-name="T1">([^<] +)</text:span></text:span>	<nau_i>\$1</nau_i>	hausa_extrait_v17.xml
suppression espaces inutiles qui suivent la classe	</nau_i><text:span text:style-name="T1">+</text:span>	</nau_i>	hausa_extrait_v18.xml
rattachement du point qui suit la définition	</ma_ana><text:span text:style-name="f_5f_standard"><text:span text:style-name="T1">\. </text:span></text:span>	.\</ma_ana>	hausa_extrait_v19.xml
suppression espaces inutiles qui suivent la définition	</ma_ana><text:span text:style-name="T1">+</text:span>	</ma_ana>	hausa_extrait_v20.xml
balisage genre	<text:span text:style-name="f_5f_label"><text:span text:style-name="T1">Jin.:</text:span></text:span><text:span text:style-name="Paradigm"><text:span text:style-name="T1">([^<] +)</text:span></text:span>	<jinsi>\$1</jinsi>	hausa_extrait_v21.xml

rattachement du point qui suit le genre	</jinsi><text:span text:style-name="f_5f_standard"><text:span text:style-name="T1">\.</text:span></text:span>	.</jinsi>	hausa_extrait_v22.xml
suppression espaces inutiles qui suivent la définition	</jinsi><text:span text:style-name="T1">+</text:span>	</jinsi>	hausa_extrait_v23.xml
Insertion de l'identifiant de notices numérotées	<article><kalma>([<]+)</kalma><text:span text:style-name="Homonym_20_number"><text:span text:style-name="T1">([<]+)</text:span></text:span>	<article id="\$1\$2"><kalma>\$1</kalma>>	hausa_extrait_v24.xml
Insertion de l'identifiant de notices non numérotées	<article><kalma>([<]+)</kalma>	<article id="\$11"><kalma>\$1</kalma>	hausa_extrait_v25.xml
Rétablissement des apostrophes	'	'	hausa_extrait_v26.xml
normalisation des noms des identifiants	id="([<]*) ['] ([<]*) "	id="\$1_\$2"	hausa_extrait_v27.xml
suppression espaces inutiles qui suivent le numéro de notice	</kalma><text:span text:style-name="f_5f_standard"><text:span text:style-name="T1"> +<text:tab/></text:span></text:span>	</kalma>	hausa_extrait_v28.xml
balisage variante	<text:span text:style-name="f_5f_label"><text:span text:style-name="T1">Yare: </text:span></text:span><text:span text:style-name="f_5f_vernacular"><text:span text:style-name="T1">([<]+)</text:span></text:span>	<yare>\$1</yare>	hausa_extrait_v29.xml
rattachement du point qui suit la variante	</yare><text:span text:style-name="f_5f_standard"><text:span text:style-name="T1">\.</text:span></text:span>	.</yare>	hausa_extrait_v30.xml

suppression espaces inutiles qui suivent la variante	<code></yare><text:span text:style-name="T1"> +</text:span></code>	<code></yare></code>	hausa_extrait_v31.xml
remplacement de caractères	-	-	hausa_extrait_v32.xml
balisage pluriel	<code><text:span text:style- name="f_5f_label"><text:span text:style- name="T1">Jam.: </text:span></text:span><text:span text:style- name="f_5f_vernacular"><text:span text:style- name="T1">([<]+)</text:span></text:span></code>	<code><jam_i>\$1</jam_i></code>	hausa_extrait_v33.xml
rattachement du point qui suit le pluriel	<code></jam_i><text:span text:style- name="f_5f_standard"><text:span text:style- name="T1">\.</text:span></text:span></code>	<code>.</jam_i></code>	hausa_extrait_v34.xml
suppression espaces inutiles qui suivent le pluriel	<code></jam_i><text:span text:style-name="T1"> +</text:span></code>	<code></jam_i></code>	hausa_extrait_v35.xml
balisage d'un sens numéroté avec insertion d'un identifiant	<code><article id="([<]+)">(.*<text:span text:style- name="Sense_20_number"><text:span text:style- name="T1">([0-9]+)\.\) *</text:span></text:span>(<ma_ana>[<]+</ma_ana><misali>[<] +</misali><jinsi>[<]+</jinsi><jam_i>[<] +</jam_i><makwatanci>[<]+</makwatanci>)</code>	<code><article id="\$1">\$2<rukunin_ma_ana id="\$1_\$3">\$4</rukunin_ma_ ana></code>	hausa_extrait_v36.xml
suppression espaces inutiles qui suivent le bloc sémantique	<code></rukunin_ma_ana><text:span text:style-name="T1"> </text:span></code>	<code></rukunin_ma_ana></code>	hausa_extrait_v37.xml

balisage sens numéroté avec insertion d'un identifiant	<pre><article id="([^\<]+)">(<kalma>[^\<] +</kalma><furici>[^\<]+</furici><nau_i>[^\<] +</nau_i><rukunin_ma_ana id="([^\<]+)"><ma_ana>[^\<] +</ma_ana><misali>[^\<]+</misali><jinsi>[^\<] +</jinsi><jam_i>[^\<]+</jam_i><makwatanci>[^\<] +</makwatanci></rukunin_ma_ana>)<text:span text:style-name="Sense_20_number"><text:span text:style-name="T1">([0-9]+).\) *</text:span></t ext:span>(<ma_ana>[^\<]+</ma_ana><misali>[^\<] +</misali><jinsi>[^\<]+</jinsi><makwatanci>[^\<] +</makwatanci>)</pre>	<pre><article id="\$1">\$2<rukunin_ma_ana id="\$1_\$3">\$4</rukunin_ma_ ana></pre>	hausa_extrait_v38.xml
balisage information morphologique	<pre><text:span text:style- name="f_5f_label"><text:span text:style- name="T1">Morph.: </text:span></text:span><text:span text:style- name="Morphology"><text:span text:style- name="T1">([^\<]+)</text:span></text:span></pre>	<pre><siffolin_kalma>\$1</siffol in_kalma></pre>	hausa_extrait_v39.xml
rattachement du point qui suit l'information morphologique	<pre></siffolin_kalma><text:span text:style- name="f_5f_standard"><text:span text:style- name="T1">\. </text:span></text:span></pre>	<pre>.\</siffolin_kalma></pre>	hausa_extrait_v40.xml
suppression espaces inutiles qui suivent l'information morphologique	<pre></siffolin_kalma><text:span text:style-name="T1"> </text:span></pre>	<pre></siffolin_kalma></pre>	hausa_extrait_v41.xml
suppression espaces inutiles qui suivent la phonétique	<pre></furici><text:span text:style- name="Notes_20__28_questions_29_"><text:span text:style-name="T1"> +</text:span></text:span></pre>	<pre></furici></pre>	hausa_extrait_v42.xml
suppression espaces inutiles qui suivent la classe	<pre></nau_i><text:span text:style- name="Notes_20__28_questions_29_"><text:span text:style-name="T1"> +</text:span></text:span></pre>	<pre></nau_i></pre>	hausa_extrait_v43.xml

suppression espaces inutiles qui suivent la définition	</ma_ana><text:span text:style-name="Notes_20__28_questions_29_"><text:span text:style-name="T1"> +</text:span></text:span>	</ma_ana>	hausa_extrait_v44.xml
suppression espaces inutiles qui suivent l'exemple	</misali><text:span text:style-name="Notes_20__28_questions_29_"><text:span text:style-name="T1"> +</text:span></text:span>	</misali>	hausa_extrait_v45.xml
balisage féminin	<text:span text:style-name="f_5f_label"><text:span text:style-name="T1">Sg.: </text:span></text:span><text:span text:style-name="f_5f_vernacular"><text:span text:style-name="T1">([<+>)</text:span></text:span>	<mace>\$1</mace>	hausa_extrait_v46.xml
rattachement du point qui suit le féminin	</mace><text:span text:style-name="f_5f_standard"><text:span text:style-name="T1">\.</text:span></text:span>	.</mace>	hausa_extrait_v47.xml
suppression espaces inutiles qui suivent le féminin	</mace><text:span text:style-name="Notes_20__28_questions_29_"><text:span text:style-name="T1"> +</text:span></text:span>	</mace>	hausa_extrait_v48.xml
suppression espaces inutiles qui suivent le pluriel	</jam_i><text:span text:style-name="Notes_20__28_questions_29_"><text:span text:style-name="T1"> +</text:span></text:span>	</jam_i>	hausa_extrait_v49.xml
suppression espaces inutiles qui suivent le féminin	</mace><text:span text:style-name="T1">+</text:span>	</mace>	hausa_extrait_v50.xml
suppression espaces inutiles qui suivent le genre	</jinsi><text:span text:style-name="Notes_20__28_questions_29_"><text:span text:style-name="T1"> +</text:span></text:span>	</jinsi>	hausa_extrait_v51.xml
bloc sémantique des notices de même structure que la première notice "a"	<article id="([<+>)">(<kalma>[<+</kalma><furici>[<+</furici>)<ma_ana>[<+</ma_ana><misali>[<+</misali><makwatanci>[<+</makwatanci></article>	<article id="\$1">\$2<rukunin_ma_ana id="\$1_1">\$3</rukunin_ma_ana></article>	hausa_extrait_v52.xml

bloc sémantique des notices de même structure que la seconde notice "a"	<code><article id="([^\<]+)">(<kalma>[^\<]+</kalma><furici>[^\<]+</furici><nau_i>[^\<]+</nau_i>) (<ma_ana>[^\<]+</ma_ana><misali>[^\<]+</misali><makwatanci>[^\<]+</makwatanci>)</article></code>	<code><article id="\$1">\$2<rukunin_ma_ana id="\$1_1">\$3</rukunin_ma_ana></article></code>	hausa_extrait_v53.xml
bloc sémantique des notices de même structure que la notice "abacada"	<code><article id="([^\<]+)">(<kalma>[^\<]+</kalma><nau_i>[^\<]+</nau_i>) (<ma_ana>[^\<]+</ma_ana><misali>[^\<]+</misali>) (<jinsi>[^\<]+</jinsi>) (<makwatanci>[^\<]+</makwatanci>)</article></code>	<code><article id="\$1">\$2\$4<rukunin_ma_ana id="\$1_1">\$3\$5</rukunin_ma_ana></article></code>	hausa_extrait_v54.xml
bloc sémantique des notices de même structure que la première notice "abara"	<code><article id="([^\<]+)">(<kalma>[^\<]+</kalma><furici>[^\<]+</furici><nau_i>[^\<]+</nau_i>) (<ma_ana>[^\<]+</ma_ana><misali>[^\<]+</misali>) (<yare>[^\<]+</yare>) (<jinsi>m.</jinsi>) (<makwatanci>[^\<]+</makwatanci>)</article></code>	<code><article id="\$1">\$2\$5\$4<rukunin_ma_ana id="\$1_1">\$3\$6</rukunin_ma_ana></article></code>	hausa_extrait_v55.xml
bloc sémantique des notices de même structure que la seconde notice "abara"	<code><article id="([^\<]+)">(<kalma>[^\<]+</kalma><furici>[^\<]+</furici><nau_i>s.</nau_i>) (<ma_ana>[^\<]+</ma_ana>) (<jinsi>[^\<]+</jinsi>) (<makwatanci>[^\<]+</makwatanci>)</article></code>	<code><article id="\$1">\$2\$4<rukunin_ma_ana id="\$1_1">\$3\$5</rukunin_ma_ana></article></code>	hausa_extrait_v56.xml
bloc sémantique des notices de même structure que la notice "abarba"	<code><article id="([^\<]+)">(<kalma>[^\<]+</kalma><furici>[^\<]+</furici><nau_i>[^\<]+</nau_i>) (<ma_ana>[^\<]+</ma_ana><misali>[^\<]+</misali>) (<jinsi>m.</jinsi>) (<makwatanci>[^\<]+</makwatanci>)</article></code>	<code><article id="\$1">\$2\$4<rukunin_ma_ana id="\$1_1">\$3\$5</rukunin_ma_ana></article></code>	hausa_extrait_v57.xml
bloc sémantique des notices de même structure que la notice "ada"	<code><article id="([^\<]+)">(<kalma>[^\<]+</kalma><furici>[^\<]+</furici><nau_i>[^\<]+</nau_i>) (<ma_ana>[^\<]+</ma_ana><misali>[^\<]+</misali>) (<jinsi>[^\<]+</jinsi>) (<makwatanci>[^\<]+</makwatanci>)</article></code>	<code><article id="\$1">\$2\$4<rukunin_ma_ana id="\$1_1">\$3\$5</rukunin_ma_ana></article></code>	hausa_extrait_v58.xml

bloc sémantique des notices de même structure que la notice " <i>adabali</i> "	<article id="([<]+)"><kalma>[<] +</kalma><furici>[<]+</furici><nau_i>[<] +</nau_i>)<ma_ana>[<]+</ma_ana><misali>[<] +</misali>)<jinsi>[<]+</jinsi><jam_i>[<] +</jam_i>)<makwatanci>[<] +</makwatanci></article>	<article id="\$1">\$2\$4<rukunin_ma_an a id="\$1_1">\$3\$5</rukunin_ma _ana></article>	hausa_extrait_v59.xml
bloc sémantique des notices de même structure que la notice " <i>adana</i> "	<article id="([<]+)"><kalma>[<] +</kalma><furici>[<]+</furici><nau_i>[<] +</nau_i>)<ma_ana>[<]+</ma_ana><misali>[<] +</misali>)<siffolin_kalma>[<] +</siffolin_kalma>)<makwatanci>[<] +</makwatanci></article>	<article id="\$1">\$2\$4<rukunin_ma_an a id="\$1_1">\$3\$5</rukunin_ma _ana></article>	hausa_extrait_v60.xml
bloc sémantique des notices de même structure que la notice " <i>adananne</i> "	<article id="([<]+)"><kalma>[<] +</kalma><furici>[<]+</furici><nau_i>[<] +</nau_i>)<ma_ana>[<]+</ma_ana><misali>[<] +</misali>)<mace>[<]+</mace>)<jam_i>[<] +</jam_i>)<makwatanci>[<] +</makwatanci></article>	<article id="\$1">\$2\$5\$4<rukunin_ma_ ana id="\$1_1">\$3\$6</rukunin_ma _ana></article>	hausa_extrait_v61.xml
bloc sémantique des notices de même structure que la seconde notice " <i>ba</i> "	<article id="([<]+)"><kalma>[<] +</kalma><furici>[<]+</furici>)<ma_ana>[<] +</ma_ana><misali>[<]+</misali>)<yare>[<] +</yare>)<makwatanci>[<] +</makwatanci></article>	<article id="\$1">\$2\$4<rukunin_ma_an a id="\$1_1">\$3\$5</rukunin_ma _ana></article>	hausa_extrait_v62.xml

bloc sémantique des notices de même structure que la notice " <i>ba'abzine</i> "	<code><article id="([^\<]+)">(<kalma>[^\<]+</kalma><furici>[^\<]+</furici><nau_i>[^\<]+</nau_i>(<ma_ana>[^\<]+</ma_ana><misali>[^\<]+</misali>(<jinsi>n.</jinsi>(<mace>[^\<]+</mace>(<jam_i>[^\<]+</jam_i>(<makwatanci>[^\<]+</makwatanci>)</article></code>	<code><article id="\$1">\$2\$4\$6\$5<rukunin_ma_ana id="\$1_1">\$3\$7</rukunin_ma_ana></article></code>	hausa_extrait_v63.xml
--	---	---	-----------------------

Transformations du fichier tmp_v2.xml

conserver les classes	<code>^.*<nau_i>(.*?)</nau_i>.*\$</code>	\$1	tmp_v3.xml
ôter les notices restantes	<code>^<article.*\$</code>		tmp_v4.xml