

Ôðáàïòéè (CNRS) è Ëèíñêitâí óíèâåðñèòåòà. Áàçà íáñ÷èòûâàåò 75 òåéñòîâ íáùèì íáúâîíí áïëåå 3 500 000 òåéñòîòò. Èñòî÷íèéàìè BFM à íñííâîíí ýâéýþòý ââòòðèòåòíûâ ëðèòé-åñéèå èçääíéý, íäíâéé åñéèå ðàçâéâàþòñý ñîñâñòâåííûâ èçääíéý, iíèðàþùèåñý íá èéíåâéñòé-åñéè åúâåðåííûâ òðâíñéðèíöèé iðèéæíàéüíûò ðóéíèñâé. Å êà-åñòåå ïðèíâðà ïðæíi iðèéæíñòé è íòåðâèéòâííâ èçääíéå àííèííâðà ðíàïà XIII à. «Iéñèé Nâyòíâ Æðâàéý» («La Queste del saint Graal») iíä ðâáàéëöèåé. È. Iáðéâéëé-íèçüy [Queste 2011].

N ìàý 2012 ãïëà äïñòóí è BFM íñóùâñòâéýåòñý iíñðâäñòâîí iíðòàëà http://txm.bfm-corpus.org/bfm, iíñòðíâíñâ íá iéàòòòò å TXM.

Âñå òåéñòû BFM ðàçïå-åíû â ôïðìàòå XML íá íñííâå ðâáéíâäæöèé TEI, â ñîñòâåðñòâé ñí ñíâðòëéé-åõéâé, ðàçðâáâòòííé æéý íóæä iðíâéòà [Guillot et al. 2010] ñ ó-åòíi iäðñíâéòéåû èéíåâéñòé-åñéïíâ áïàéèçà. Iéàòòòò å TXM íá òïëüéí ñéðæòè æéý èíðïøâà BFM ñðâäñòåíí ãïñòóíâ iíëüçíâàòåéé, íí è iíçâéýåò ñîñóùâñòâéýöü ðýâ iíðâéëéå èâòòíàòè-åñéíé è iíëóâàòòíàòè-åñéíé ðàçïåòéé (â-åñòííñòè, iíðôíèíâè-åñéíé àííòòåëé, åúýâéåíéý iíðýíé ðâ-è).

Íâíé èç íåéâéåå ñëíæíûò çàëå- ïðè èñíñëüçíâåéé ñðâäñòâåòòòíò èâòòíàòè-åñéíé ëéíâéñòé-åñéïíâ áïàéèçà êíðïøâà íðèíâéèòåéüí è òåéñòàí, åééþ-åþùèí ãéøåâéóþ ðâäâæòòíðñéóþ ðàçïåòéé, ýâéýåòñý êíðâåéòíàý èäåíòéòéåòéý ñîñâ (óîéâíéçàöèý) è iíðâéëíæâééé, ê èíðöûòíí ÿòòí ãïàéèç åíëæåíí iðèíâéòüñý áâç iíðâåéò ñàííé ðâáàéëòíðñéíé ðàçïåòéé.

Nëâååðþùèé ïðèíâð ðàçïåòéé, ñîñâåðæâåéòéé ïðâäéëíæâííûé ðâáàéëòíðii ðôðâåâòíò ðâéñòà íá iåñòå éâéòíû, íà-èíâþùâéñý â èííòâ íäíâíñâ ñëíââ è çàéâí-èâåâþùâéñý íåñéèéüèé ñëíâàíè iíçæå, àáñíéþöíí êíðâåéòåí ñ òí-èé çðåééý ðâéíâíâåéòé TEI, íäíâéí áâñûíà ñëíâæåí ãéý ÿòâéíéçàöèé ñ ò-åòíí iñíâåííñòâé ýçüéâ XML (çàïðåò «iâðâéâòåèââéý» ýéâïâlòâà, ðéñé iíÿâéâéý iðíâåéíâ iâéæåó ÿââià è òåéñòíâùíè ÿçëàíè ïðè íâðâáíòéå):

en^{<supplied>}tra a
cheval en la^{</supplied>} sale une mout bele damoisele

Áùâ áïëåå ñåðüåçíûâ ïðíâéåíûâ áíçíééåþò ïðè ðàçïåòéâ ïðâäéëíæâéé, íñíâåíííâ ñ ñòðñòåò ñðóðòéåòðò, ååä íâðâéëòåèâåéé åâððè-åñéíé è ñéïòàéñè-åñéíé ñòðóéòòðû åñòðâ-åâòñý i-åííû ðàñòðò.

Ðàçóïåâòñý, iíæíí «iòòéëüòðíâåòöü» áñâ åéâïâéòûâéåòñý, íâðâéëòåèâåéàþùèåñý ñ íñíâíñòè ëéíâæñòé-åñéíé ñòðóéòòðâà (ñéââàíè è ïðâäéëíæâéíé), íäíâéé åïòòí ïðèååñòé è iíòåðâ ñóùâñòâåííé èíðïðíâòéé ãéý çàïðñîâ è åèçóâéèçàöèé (íâíðèíâð, í òí, iíââåâðâæàéñü èé ñëíâíòíðà ðâáàéëòíðñéíé ïðâåâéâ).

Níçääíéå åéâíðòíà òíéâéèçâöèé, êíòíðûé êíðâåéòíí íâðâåâòâåâæ áú ëþâíé òåéñòâ ñ åéøåíé ðâáâæòòíðñéíé ðàçïåòéíé â ñòàíâðòå XML, íðâñòâåâéþòñý iíðâéò-åñéè åâñçííæíû. Òâí íá iâíâå, iíæíí åâððüñý åíëéíå ñâíâæòâðòâðüäüíû ðâçóëüðâðâåå íðè ñóéâæéâ, ÷òò ðâçïåòéâ åñðâéåíí ãâéñàíòâ ïòââ-åâðò ðýâò iíñòðû ïðâæâéé. Íâíðèíâð, «óýâè, ðâñííëíæâííûâ áíóðò ñëíâ, åéææíû âúòù-åâðâí èâæíöéòéòðââíû» èéè «åñéè ðàçïå-åííûé ñââïâíò òåéñòâ ñà-èíâåòñý åíóðò è íâíâæéâ è çââåâòâåâæ åâðâéëëòíðñéíé ðâçäâéëèòü».

Òâéæå åíçííæíí ñïñòðâæðò ñïèñéè ÿââíâð TEI à çââéñèíñòè ïðè è òíçæöèé â èéíâæñòé-åñéíé èâðâåððéè òåéñòâ â ðâïâéâò ïðââéüíñâ iíðâéòâ. Íâíðèíâð, iíðâéëíæâéý ðâñííëâåâþòñý åíóðò è ÿéâïâlòâ òëà «âáçàö» <p> èéè <áéëé òåéñòâ> <ab>, à íá iâíâðò ðâ-òâðâòü

ýéâèâàëåíóíùè ñëïáô (íaiðèìåð, <abbr>, <num>, <pc>), à
íáñéîëüêî ýéâìáíòâ ïí-òè åñâåäà ðàñïëéàäàþòñ ÿáíòðè ñëïáà (<am>,
<c>, <ex>). Ðýä ýéâìáíòâ ñïäåðæàò ñååìáíòû òåññòà, éíòðûå íå ñëâåäåð
òíéâíçëöðíåäòû, ïñéïëüêò ìíò íä ïðéíäéàæàò è íåðåðééàò ñòðî-íééå (íaiðèìåð,
ðååàéòíðñéèå ïðéíà-áéý è ñíññéè å òåññòå êðèòè-åññéíâï ëçääíéý). Å ðàìéàõ
éííëðåoííï ïðíäéòà Íòéòåà Íòéòåà ñïññéè ïäñóò áúòù ñóùåñòåííï ðàñøèðåíû è óòí-íáíû, à
ðåçóëüòå ðåååò -åññéè ëéññé ÿéâìáíòâ, éíòðûå ïäðåðéðåùèåòüñÿ ñ
éëíâæñòè-åññéèò ñòðóéòðåíè, cíà-ëéòåëüíï ñíéðåùåòüñÿ.

Ðàçìåðòåà ðåññòåòâ BFM ïñíòâååòñòåò ðåòðèï ñòðî-íòééòðíåàííîé
ñïåòèòåëéàöèè ïðéíàíåíèý ðåéïíàíäåöèé TEI, òòðàæåííîé å ãéñóìåíòåòëè ODD
[Guillot et al. 2010]. Åäíàÿ ñïåëòéòéåëöèý åéëþ-ååò ïðååëëå èñíïëüçíåáíéý
åíóòðèñíïåíû ðýåáï (íaiðèìåð, èñíïðååéäåíû ðååàéðöðí ãóéå èééé çíåéïå
ïåðåííñà), à òåéæå ÿéâìáíòâ, éíòðûå ïäñóò íåðåðéðåùèåòüñÿ ñí ñòðóéòóðîé
ïðååëíæåíéé (íaiðèìåð, «Iøñòúå» ÿéâìáíòû <lb/> «Íàáÿ ñòðîéà»
èñíïëüçòþñý åíåñòí <l> «ñòëö», à ÿéâìáíò òëðèòðåíäíéý <q>
ðàññíàòðåðéååòñý èåéé åðååéåæåíéý). Ýóï îççáíéèé ðååéëçíåàòü å íæàòòðìå
TXM ýóðååòðèåíûé àéäíòðèò ðåéçàòëè åéý òåññóïå BFM.

Å ïñíøååíååå ååðåíå ùéé å ñòðåííï ïðåååååíí å ðýä ðåññòåòâ ïí ååðàëòåöèè
åéííñàí åéäíðòåà è åéñóìåíòà TEI XML, ïäñíòâåäåííû ìòðååéòàò. Ñðåæè
íéò ïñæíí íåçàòåò «Åèðòåäåéüíòå ãéåééòðåéå åóïàéñòåï»
<http://www.bvh.univ-tours.fr> è åçääíéå -åðííæééíå ðííàíá «Áóâàð è Íåéþøå»
Åþñòååà Öéíååðà (<http://dossiers-flaubert.ish-lyon.cnrs.fr>). Åääíòåòëý ñíñòïèò
å ïðéíàíåíéè ñïåëòéàëüíû ðéëüòò XSL íà åòíïå è íá åúöñå ïðåååååòðû
òíéâíéçåòëè. «Åñïáíéé» ðéëüòð óååéýåò ðýåé, éíòðûå íå ïðååñòååéýþò èíòååðåñà
åéý ýéñïëóàðåéëè ñ ïñíùüþ TXM, à òåéæå ðíðùåàò è ííðíàéèéçåò ðåéíòðû
ñíéíæíñå ñòðóéòðóðû ýéâìáíòâ XML. «Åûðíäíéé» ðéëüòð ïçåäéýåò èñíïðååéòü ðýä
íøéåé, éíòðûå ðååéòð-åññéè íååçíïæíí èçååææåòü å íðååññå íåðâè-íé
òíéâíéçåòëè (íaiðèìåð, ååéäíåíà ñëïáà ïðéíåññå ìòðååññå ìòðååññå ñòðàíéöû, éíñäà õåéüé
ðýä òýäïå íåæåò ðåññéååòüñÿ íåæååò ååñ íà-åéíí è éíñöíí).

Å ðåññòåò åéññéåäà ïú ïðéååååíí åíééðåòíûå ïðéíàðû ðååéëçíåàííû
ðåðååíéé è íðíæíñíñòðéðåíí åðåçéëüòåòû, éíòðûå ííæíí ïíéò-ëòü ïðé ýéñïëóàðåöèè
òåññóïå BFM è äðóæéò ïðíåéòåíå ñ ïñíùüþ íëàòòðìå TXM. Áíéåå ïäðåíñå ðàçëé-íúå
åéäíòðèòíù ðåéçàòëè, ðååéëçíåàííûå íà íëàòòðìå TXM, ííèñàíú å [Heiden 2010].

Niññíê ëëòåðåòóðû

Guillot, C., Heiden, S., Lavrentiev, A., Bertrand, L. Manuel
d'encodage XML-TEI des textes de la Base de Français Médiéval, Lyon: Équipe
BFM, 2010. —; Ååðåñ å Èíøååðåò http://bfm.ens-lyon.fr/IMG/pdf/Manuel_Encodage_TEI.pdf.

Heiden, S. The TXM
Platform: Building Open-Source Textual Analysis Software Compatible with the
TEI Encoding Scheme. 24th // Pacific Asia Conference on Language, Information
and Computation / Ed. Kiyoshi Ishikawa Ryo Otoguro. Institute for Digital
Enhancement of Cognitive Development, Waseda University, 2010. P. 389-398.

Heiden, S., Magué,
J.-P., Pincemin, B. TXM : Une plateforme logicielle open-source pour la

textométrie – conception et développement // Statistical Analysis of Textual Data - Proceedings of 10th International Conference JADT 2010 / Bolasco, S. et al. (Eds.). – Rome: Edizioni Universitarie di Lettere Economia Diritto, 2010. – P. 1021-1032. – Àäðåñ à Èíòåðíåò
http://www.ledonline.it/ledonline/JADT-2010/allegati/JADT-2010-1021-1032_025-Heiden.pdf.

Pincemin, B., Heiden,
S., Lay, M.-H., Leblanc J.-M. and Viprey, J.-M. Fonctionnalités textométriques:
Proposition de typologie selon un point de vue utilisateur. // Statistical
Analysis of Textual Data - Proceedings of 10th International Conference JADT
2010 / Bolasco, S. et al. (Eds.). – Rome: Edizioni Universitarie di Lettere
Economia Diritto, 2010. – P. 341-353. – Àäðåñ à Èíòåðíåò
http://www.ledonline.it/ledonline/JADT-2010/allegati/JADT-2010-0341-0354_023-Pincemin.pdf.

Queste del saint Graal. Édition
numérique interactive / Ed. Marchello-Nizia, Ch. – Lyon: Équipe de la BFM,
2011. – Àäðåñ à Èíòåðíåò <http://txm.bfm-corpus.org/txm>.