ÈÑÒÎÐÈ×ÅÑÊÈÅ ÊÎÐÏÓÑÀ È ÊÎÐÏÓÑÍÛÅ ÄÈÀÕÐÎÍÈ×ÅÑÊÈÅ ÈÑÑËÅÄÎÂÀÍÈß

© Âèêòîð Ïàâëîâè÷
Çàõàðîâ. Ðîññèÿ. Ñàíêò-Ïåòåðáóðã. Êàôåäðà ìàòåìàòè÷åñêîé
ëèíãâèñòèêè ôèëîëîãè÷åñêîãî ôàêóëüòåòà Ñàíêò-Ïåòåðáóðãñêîãî
óíèâåðñèòåòà

ßçûê, êàê èçâåñòíî, äèíàìè÷íàÿ ñèñòåìà;
â èñòîðè÷åñêèå ïåðèîäû ðàçíîé ïðîäîëæèòåëüíîñòè íà âñåõ åãî óðîâíÿõ (â ôîíåòèêå è ïèñüìå, ìîðôîëîãèè è ëåêñèêå, ñèíòàêñèñå è ñåìàíòèêå) ïðîèñõîäÿò èçìåíåíèÿ: ÷àñòîòà âñòðå÷àåìîñòè îäíèõ ýëåìåíòîâ, ÿâëåíèé óìåíüøàåòñÿ, à áûâàåò, ÷òî ÿâëåíèÿ, ýëåìåíòû ïîëíîñòüþ âûõîäÿò èç óïîòðåáëåíèÿ, äðóãèå æå âîçíèêàþò èëè ñòàíîâÿòñÿ áîëåå ÷àñòîòíûìè, ÷åì ïðåæäå. Èçìåíåíèÿ îáóñëîâëåíû äåéñòâèåì ôàêòîðîâ ðàçëè÷íîé ïðèðîäû &ndash; ïðåæäå âñåãî ïñèõîëîãè÷åñêèõ, ñîöèàëüíûõ è êóëüòóðíûõ. Äèàõðîíè÷åñêèå èññëåäîâàíèÿ ÿçûêà ïîçâîëÿþò âûÿâèòü ôàêòû è çàêîíîìåðíîñòè íå òîëüêî ëèíãâèñòè÷åñêîãî, íî è èñòîðèêî-êóëüòóðíîãî çíà÷åíèÿ.
È íå ñëó÷àéíî â ïîñëåäíåå âðåìÿ â ðàìêàõ êóëüòóðîëîãèè ïîÿâèëîñü íàïðàâëåíèå íàó÷íûõ èññëåäîâàíèé, íàçûâàåìîå «êóëüòóðîìåòðèÿ» (ñèíîíèì «êâàíòèòàòèâíàÿ êóëüòóðîëîãèÿ»).

Åùå íåäàâíî ïðîâåäåíèå òàêèõ
èññëåäîâàíèé òðåáîâàëî áîëüøèõ óñèëèé è çàòðàò âðåìåíè. Ñåãîäíÿ êîìïüþòåðíûå òåõíîëîãèè è êîðïóñíàÿ ëèíãâèñòèêà äàþò äëÿ íèõ ïðèíöèïèàëüíî íîâûå èíñòðóìåíòû. Äëÿ ýòîãî ñîçäàþòñÿ äèàõðîíè÷åñêèå (èñòîðè÷åñêèå) êîðïóñà. Ìîæíî ñêàçàòü, ÷òî ýòî êîðïóñà, êîòîðûå ïîçâîëÿþò èçó÷àòü ðàçâèòèå ÿçûêà íà ïðîòÿæåíèè êàêîãî-òî (äîñòàòî÷íî äëèííîãî) ïðîìåæóòêà âðåìåíè èëè èçó÷àòü ÿçûê â åãî ïðåæíèõ ñîñòîÿíèÿõ. Â ÷èñëå äèàõðîíè÷åñêèõ êîðïóñîâ ìîæíî íàçâàòü Ñàíêò-Ïåòåðáóðãñêèé êîðïóñ äèàõðîíè÷åñêèõ òåêñòîâ, Ðåãåíñáóðãñêèé äèàõðîíè÷åñêèé êîðïóñ ðóññêîãî ÿçûêà (äðåâíåðóññêèå òåêñòû), Êîëëåêöèþ äðåâíåéøèõ è ñðåäíåâåêîâûõ ñëàâÿíñêèõ è ðóññêèõ òåêñòîâ "Ìàíóñêðèïò", Ðóêîïèñíûå ïàìÿòíèêè Äðåâíåé Ðóñè (áåðåñòÿíûå ãðàìîòû, ëåòîïèñè, ðóêîïèñíûå êíèãè) è äð.

Ïîíÿòèå êîðïóñ ïîäðàçóìåâàåò òàêæå
èíñòðóìåíò äëÿ äèàõðîíè÷åñêèõ èññëåäîâàíèé. Ñ òî÷êè çðåíèÿ ôóíêöèîíàëüíîñòè áîëüøîé íàáîð ëèíãâèñòè÷åñêèõ ìîäóëåé ïðåäëàãàåò ñèñòåìà «Ìàíóñêðèïò». Â íàøåì êóðñå ìû ïîäðîáíî îñòàíîâèìñÿ íà äâóõ èíñòðóìåíòàõ, íàõîäÿùèõñÿ íåïîñðåäñòâåííî â ðàìêàõ êîðïóñíîé ëèíãâèñòèêè è âûïîëíÿþùèõ ñòàòèñòè÷åñêóþ îáðàáîòêó êîðïóñíûõ äàííûõ.

Ê òàêèì èíñòðóìåíòàì îòíîñÿòñÿ ñèñòåìà
Google Books Ngram Viewer è ñåðâèñ «Ãðàôèêè» Íàöèîíàëüíîãî êîðïóñà ðóññêîãî ÿçûêà (ÍÊÐß). Áàçà äàííûõ Google Books Ngram Viewer òàêæå ïðåäñòàâëåíà Ì. Äýâèñîì â ñîñòàâå åãî êîðïóñîâ íà ñàéòå Brigham Young University

Google books Ngram Viewer â íàñòîÿùåå
âðåìÿ ñîäåðæèò íåñêîëüêî êîðïóñîâ ðàçìå÷åííûõ òåêñòîâ êíèã íà 9 ÿçûêàõ. Íà êîíåö 2012 ã.
áàçà äàííûõ íàñ÷èòûâàë áîëåå 8 ìëí êíèã (òåêñòîâ), ÷òî ñîñòàâëÿåò îêîëî 6% âñåõ êîãäà-ëèáî îïóáëèêîâàííûõ ïå÷àòíûõ êíèã. Êîðïóñ êíèã íà ðóññêîì ÿçûêå ñîäåðæèò 591 310 òåêñòîâ îáùèì îáúåìîì áîëåå 67 ìëðä. ñëîâîóïîòðåáëåíèé. Ñàìûå ïîçäíèå ïóáëèêàöèè, äîñòóïíûå äëÿ ïîëüçîâàòåëåé â íàñòîÿùåå âðåìÿ, îòíîñÿòñÿ ê 2008 ãîäó.

Ñåðâèñ «Ãðàôèêè» ôóíêöèîíàëüíî ïîäîáåí
ñåðâèñó Google Books N Gramm Viewer.
Îáà îíè ïîêàçûâàþò õðîíîëîãè÷åñêîå ðàñïðåäåëåíèå çàäàííûõ è íàéäåííûõ ëåêñè÷åñêèõ åäèíèö â çàäàííîì âðåìåííîì ïðîìåæóòêå.

Ïðîáëåìà, èìåþùàÿ îñîáóþ âàæíîñòü â
äèàõðîíè÷åñêèõ èññëåäîâàíèÿõ ðóññêîãî ÿçûêà &ndash; ýòî ïðåäñòàâëåíèå ðóññêèõ òåêñòîâ
â ãðàôèêå è îðôîãðàôèè, äåéñòâîâàâøåé äî 1918 ã. âêëþ÷èòåëüíî. Â Google books Ngram
Viewer ñòàðàÿ îðôîãðàôèÿ ñîõðàíÿåòñÿ. Ñëåäóåò ñêàçàòü, ÷òî è â ÍÊÐß ñåé÷àñ
ïîÿâèëñÿ èñòîðè÷åñêèé ïîäêîðïóñ ñ îðèãèíàëüíîé îðôîãðàôèåé îáúåìîì áîëåå 7 ìëí
ñëîâîóïîòðåáëåíèé.

Ïðîãðàììà
ê êóðñó

«Èñòîðè÷åñêèå
êîðïóñà è êîðïóñíûå äèàõðîíè÷åñêèå èññëåäîâàíèÿ»

(10
÷àñîâ ëåêöèîííûõ çàíÿòèé è ïðàêòèêóìîâ, ñðîêè: 14-18 íîÿáðÿ 2015 ã.)

Ëåêöèÿ 1:
"Êîðïóñíàÿ ëèíãâèñòèêà â Ðîññèè"

1. Êîíöåïöèè êîðïóñíîé ëèíãâèñòèêè, ìåñòî äèñöèïëèíû è
ñîáñòâåííî êîðïóñîâ â ðÿäó ëèíãâècòè÷åñêèõ èíôîðìàöèîííûõ òåõíîëîãèé

2. Èñòîðè÷åñêèé ýêñêóðñ è ñîâðåìåííîå ñîñòîÿíèå ðîññèéñêîé
êîðïóñíîé ëèíãâèñòèêè.

3. Îáçîð êîðïóñîâ ðóññêîãî ÿçûêà ðàçëè÷íûõ òèïîâ.

Ëåêöèÿ 2:
Âîçìîæíîñòè êîðïóñíîé ëèíãâèñòèêè â ïðîâåäåíèè äèàõðîíè÷åñêèõ èññëåäîâàíèé

1. Îáçîð êîðïóñîâ è êîðïóñíûõ èíñòðóìåíòîâ, ðåàëèçóþùèõ ýòè
âîçìîæíîñòè

2. Âîïðîñû ðàçìåòêè äèàõðîíè÷åñêèõ êîðïóñîâ.

Ñåìèíàð:
îáñóæäåíèå òåì ëåêöèé 1 è 2

Ïðàêòè÷åñêèå çàíÿòèÿ

1) Ðàáîòà ñ ñåðâèñîì "Ãðàôèêè" Íàöèîíàëüíîãî êîðïóñà ðóññêîãî ÿçûêà

2) Ðàáîòà ñ ñåðâèñîì Google Books Ngram Viewer

Ëèòåðàòóðà

Çàõàðîâ Â.Ï., Ìàñëåâè÷ À.Ö.

Äèàõðîíè÷åñêèå èññëåäîâàíèÿ íà îñíîâå êîðïóñà ðóññêèõ òåêñòîâ Google Books
Ngram Viewer. // Ñòðóêòóðíàÿ è ïðèêëàäíàÿ ëèíãâèñòèêà. Âûïóñê 10. ÑÏá.: Èçä-âî
Ñ. Ïåòåðá. óí-òà, 2014. Ñ. 303-327.

Çàõàðîâ Â.Ï., Áîãäàíîâà Ñ.Þ. Êîðïóñíàÿ ëèíãâèñòèêà: Ó÷åáíèê äëÿ ñòóäåíòîâ íàïðàâëåíèÿ
«Ëèíãâèñòèêà». 2-å èçä., ïåðåðàá. è äîïîëí., &ndash; ÑÏá.: ÑÏáÃÓ. ÐÈÎ. Ôèëîëîãè÷åñêèé
ôàêóëüòåò, 2013. &ndash; 148 ñ.

Çàõàðîâ Â. Corpora of the Russian Language //
Text, Speech and Dialogue: Proceedings of the 16th International Conference,
TSD 2013, Plzen, Czech Republic, September 1&ndash;5, 2013. (Lecture Notes in
Artificial Intelligence, 8082) / Ivan Habernal, Václav Matou&scaron;ek (Eds.). &ndash; Springer-Verlag,
Berlin Heidelberg, 2013. &ndash; P. 1-13.

Webography

https://spbu.academia.edu/VictorZakharov.