كومپيۇتېر تىلشۇناسلىقى ۋە ئۇيغۇر تىلى تەتقىقاتى

كومپيۇتېر تىلشۇناسلىقى كومپيۇتېر ئىلمى بىلەن تىلشۇناسلىق تەتقىقاتىنىڭ بىرلەشكەنلىكىنىڭ مەھسۇلى بولۇپ ، ئەڭ تېز تەرەققىي قىلىۋاتقان پەنلەرنىڭ بىرى. بۇ پەننى ۋە ئۇنىڭ ئۇيغۇر تىلى تەتقىقاتىدىكى قوللىنىلىشىنى تونۇشتۇرۇش مەقسىتىدە تۆۋەندىكى ماقالە يوللاندى.
بۇ ماقالە «تىل ۋە تەرجىمە» ژۇرنىلىنىڭ 2006-يىلى 2-سانىغا بېسىلغان ، ماقلە يېزىلغىلى خېلى بولغانلىقتىن بەزى تەتقىقات ئەھۋاللىرى تولۇق ئەكس ئەتتۈرۈلمىگەن بولۇشى مۇمكىن. تۈزىتىش ئوقۇشۇڭلارنى ۋە باھا-تەكلىپ بېرىشىڭلارنى ئۈمىد قىلىمەن .

كومپيۇتېر تىلشۇناسلىقى ۋە ئۇيغۇر تىلى تەتقىقاتى
计算语言学与维吾尔语研究

سەمەت مەمتىمىن تىلچىن

قىسقىچە مەزمۇنى:
بۇ ماقالىدە كومپيۇتېر تىلشۇناسلىقىنىڭ ئېنىقلىمىسى ، تەتقىقات مەزمۇنلىرى ۋە ئەنئەنىۋى تىلشۇناسلىق بىلەن بولغان پەرقى سۆزلەنگەن ، ئاندىن مۇشۇ ئاساستا كومپيۇتېر تىلشۇناسلىقىنىڭ ئۇيغۇر تىلى تەتقىقاتىدا قوللىنىلىش ئەھۋالى ۋە ساقلانغان مەسىلىلەر خۇلاسىلانغان ھەم كومپيۇتېر تىلشۇناسلىقى نەزەرىيىسى ۋە ئۇسۇلىدىن پايدىلىنىپ ئۇيغۇر تىلى تەتقىقاتىدا ھەل قىلىشقا تېگىشلىك مەسىلىلەر ئوتتۇرىغا قويۇلغان.
ئاچقۇچلۇق سۆزلەر: كومپيۇتېر تىلشۇناسلىقى ؛ ئۇيغۇر تىلى ؛ تەتقىقات ئەھۋالى.

كومپيۇتېر تېخنىكىسىنىڭ راۋاجلىنىشى نۇرغۇن ساھەلەردە يېڭى ئۆزگىرىش ۋە تەرەققىياتلارنى بارلىققا كەلتۈردى ، كومپيۇتېر تېخنىكىسىنىڭ ئىلىم-پەن تەتقىقاتىدىكى رولىمۇ كۈنسېرى كۈچەيدى ، بۇنىڭ بىلەن نۇرغۇن پەنلەر كومپيۇتېر ئىلمى بىلەن گىرەلەشتى ۋە بەزى يېڭى ئارىلىق پەنلەر (边缘学科) شەكىللىنىشكە باشلىدى. جۈملىدىن ، بىر تەرەپتىن تىلشۇناسلىق تەتقىقاتىدا كومپيۇتېر تېخنىكىسى كەڭ قوللىنىلدى ، بۇ تىل تەتقىقاتىنى زور دەرىجىدە ئىلگىرى سۈردى ؛ يەنە بىر تەرەپتىن تىل تەتقىقاتىدىكى ئەڭ يېڭى نەزەرىيە ۋە بايقاشلار كومپيۇتېر ساھەسىدە ئىشلىتىلىپ كومپيۇتېر تېخنىكىسىنى راۋاجلاندۇردى ھەم ئۇنىڭ ئىشلىتىلىش دائىرىسىنى تېخىمۇ كېڭەيتتى. تەبىئىي تىلنى ئاپتوماتىك بىر تەرەپ قىلىش تېخنىكىسى (自然语言自动处理技术) بارلىققا كېلىپ راۋاجلاندى ، بۇ ئۆز نۆۋىتىدە يەنە تىل تەتقىقاتىغا نۇرغۇن تەلەپلەرنى قويدى ، بۇنىڭ بىلەن كومپيۇتېر ئىلمى بىلەن تىلشۇناسلىق ئىلمىنىڭ ھەقىقىي گىرەلىشىشى ئىشقا ئېشىپ يېڭى بىر ئارىلىق پەن كومپيۇتېر تىلشۇناسلىقى (计算语言学) بارلىققا كەلدى. كومپيۇتېر تىلشۇناسلىقى پەن بولۇپ شەكىللەنگەندىن بۇيان تەتقىقاتچىلار قىزغىن تەتقىق قىلىشقا كىرىشىپ كەتتى ، ھازىرمۇ كومپيۇتېر تىلشۇناسلىقى ناھايىتى قىزىق پەنلەرنىڭ بىرى بولۇۋاتماقتا. كومپيۇتېر تىلشۇناسلىقى يەنە يېڭى بىر تەتقىقات ئۇسۇلى سۈپىتىدە ھەر قايسى تىللار تەتقىقاتىدا قوللىنىلىپ نۇرغۇن نەتىجىلەر قولغا كەلتۈرۈلدى.
ئۇيغۇر تىلى تەتقىقاتچىلىرى ۋە كومپيۇتېر تەتقىقاتچىلىرىمۇ بۇ يېڭى پەنگە دىققەت قىلىشقا ۋە تەتقىق قىلىشقا كىرىشتى ، شۇنداقلا بەزى نەتىجىلەرگە ئېرىشتى. مۇشۇنداق ئەھۋالدا كومپيۇتېر تىلشۇناسلىقى توغرىسىدىكى چۈشەنچىلەرنى ئومۇملاشتۇرۇش ، كەڭ تەتقىقاتچىلارنىڭ بۇ يېڭى پەنگە بولغان دىققىتىنى قوزغاش ، باشقا مىللەتلەردىكىگە ئوخشاش تەتقىقاتنى كۈچەيتىپ تىل تەتقىقاتىدىمۇ يۈكسىلىش ھاسىل قىلىش زۆرۈر بولۇپ قالدى. بۇ ماقالىدە دەل مۇشۇ مەقسەت چىقىش قىلىنغان. يەنى ماقالىدە كومپيۇتېر تىلشۇناسلىقى توغرىسىدا چۈشەنچە بېرىش ئاساسىدا ، كومپيۇتېر تىلشۇناسلىقىنىڭ ئۇيغۇر تىلى تەتقىقاتىدا قوللىنىلىش ئەھۋالى ۋە ساقلانغان مەسىلىلەر خۇلاسىلاندى ھەم كومپيۇتېر تىلشۇناسلىقى نەزەرىيىسى ۋە ئۇسۇلىدىن پايدىلىنىپ ئۇيغۇر تىلى تەتقىقاتىدا ھەل قىلىشقا تېگىشلىك مەسىلىلەر ئوتتۇرىغا قويۇلدى.

1.كومپيۇتېر تىلشۇناسلىقى توغرىسىدا

(1ئېنىقلىمىسى

كومپيۇتېر تىلشۇناسلىقىنىڭ تەۋەلىكى توغرىسىدا تېخى بىرلىككە كەلگەن قاراش يوق ، تىلشۇناسلار ئۇنىڭ تەتقىقات ئوبيېكتىنىڭ تەبىئىي تىل ئىكەنلىكىنى ئاساس قىلىپ ئۇنى تىلشۇناسلىقىنىڭ تارمىقى دەپ قارىسا ، كومپيۇتېر تەتقىقاتچىلىرى قوللىنىلغان تەتقىقات ۋاسىتىسىنىڭ كومپيۇتېر تېخنىكىسى ئىكەنلىكىگە قاراپ ئۇنى كومپيۇتېر ئىلمىنىڭ تارمىقى دەپ قارايدۇ ، ھەتتا بەزى ماتېماتىكا ئالىملىرى ئۇنى ئەمەلىي ماتېماتىكا ئىلمىنىڭ تارمىقى دەپ قارايدۇ. لېكىن خۇلاسىلەپ ئاددىي قىلىپ ئېيتقاندا ، كومپيۇتېر تىلشۇناسلىقى (Computational Linguistics) فورمۇلالاشتۇرۇلغان

ھېسابلاش قېلىپى قورۇش ئاساسىدا تىلنى تەھلىل قىلىدىغان ، چۈشىنىدىغان ۋە بىر تەرەپ قىلىدىغان پەن. كومپيۇتېر تىلشۇناسلىقى بىر ئارىلىق پەن بولۇپ ، ئۇ تىلشۇناسلىق ، لوگىكا ، پسىخىك تىلشۇناسلىق ، كومپيۇتېر ئىلمى ، پەلسەپە ، سۈنئىي ئىقتىدار ئىلمى ، ماتېماتىكا ۋە ستاتىستىكا ئىلمىنىڭ تەتقىقات نەتىجىلىرىنى قوللىنىپ كومپيۇتېر ئارقىلىق تىلنى تەھلىل قىلىدۇ.

(2تەتقىقات مەزمۇنى

ئادەتتە تىل تەتقىقاتىدا تىل فونېتىكا ، لېكسىكا ، گىرامماتىكا ، سېمانتىكا ۋە پراگماتىكىدىن ئىبارەت بىر قانچە قاتلامغا بۆلىنىدۇ. كومپيۇتېر تىلشۇناسلىقمۇ مۇشۇ ئاساستا بىر قانچە تارماققا بۆلۈنگەن بولۇپ ئۇلارنىڭ تەتقىقات مەزمۇنى ئاساسلىقى تۆۋەندىكىدەك:

(1)كومپيۇتېردىن پايدىلىنىپ تىل تاۋۇشلىرى ئۇچۇرلىرىنى بىر تەرەپ قىلىش ، تىلنى ئاپتوماتىك سىنتېزلەش ۋە تونۇتۇشنى تەتقىق قىلىدۇ.

(2)تىلدىكى سۆزلەرنى كومپيۇتېردا بىر تەرەپ قىلىش ئۇسۇلى ، سۆزلۈك ئامبىرى ، ئاتالغۇلار ئامبىرى قاتارلىق ماشىنىلاشقان لۇغەتلەرنى قۇرۇش ئۇسۇللىرىنى تەتقىق قىلىدۇ.

(3)كومپيۇتېردىن پايدىلىنىپ تىلنىڭ گىرامماتىكىسىنى تەھلىل قىلىش ئۇسۇلىنى تەتقىق قىلىدۇ.

(4)تەبىئىي تىل ئىپادىلىگەن مەنىنى كومپيۇتېردا قانداق تەھلىل قىلىشنى تەتقىق قىلىدۇ.

(5)كۆپ مىقداردىكى تەبىئىي تىل ماتېرىياللىرىنى كومپيۇتېردا توپلاش ، ساقلاش ۋە ستاتىستىكا قىلىش ئۇسۇلىنى ، شۇنداقلا تىل ماتېرىياللىرى ئامبىرى تىلشۇناسلىقى ئۇسۇلىنىڭ تىلغا قارىتا مىقدار ئانالىزى ئېلىپ بېرىش ، لۇغەت تۈزۈش ، ئەسەرنىڭ ئۇسلۇبىنى تەھلىل قىلىش ، تەبىئىي تىلنى چۈشىنىش ، ماشىنا تەرجىمىسى قىلىش قاتارلىق ساھەلەردە قوللىنىلىشىنى تەتقىق قىلىدۇ .

كومپيۇتېر تىلشۇناسلىقىنىڭ تەتقىقات دائىرىسى مانا مۇشۇنداق ناھايىتى كەڭ بولۇپ ، تەتقىقات دائىرىسى تەبىئىي تىلنى كومپيۇتېردا بىر تەرەپ قىلىشقا ئائىت بارلىق ساھەلەرگە چېتىلىدۇ ؛ باشقىچە قىلىپ ئېيتقاندا كومپيۇتېر تىلشۇناسلىقى ئارقىلىق ئېرىشىلگەن نەتىجىلەر تىلشۇناسلىقتا قوللىنىلىشتىن سىرت ماشىنا تەرجىمىسى)ئاپتوماتىك ھالدا بىر تىلنى ئىككىنچى بىر تىلغا تەرجىمە قىلىش) ، تاۋۇشلارنى بىر تەرەپ قىلىش (ئاۋازنى تونۇتۇش ۋە ئاۋاز ھاسىل قىلىشنى ئۆز ئىچىگە ئالىدۇ (، ئاپتوماتىك كوررېكتورلۇق قىلىش (ئالاھىدە ياسالغان يۇمشاق دېتاللار ئارقىلىق ئەسەر ، كىتابلارنىڭ كوررېكتورلۇقىنى قىلىش) ، ئاپتوماتىك ئۇچۇر ئىزدەش (كومپيۇتېر ئارقىلىق زور مىقداردىكى ئۇچۇرلار ئارىسىدىن ئېھتىياجلىق ئۇچۇرلارنى توپلاش) ، سۆز-ئاتالغۇلار ئامبىرى قۇرۇش (ھەر خىل تىپتىكى بىر تىللىق ياكى كۆپ تىللىق ئېلېكترونلۇق لۇغەتلەرنى تۈزۈش) ، تەبىئىي تىلنى چۈشىنىش (كومپيۇتېر بىلەن ئادەمنىڭ سۆزلىشىشى ، تەبىئىي تىل بىلەن بىۋاسىتە ئۇچۇر ئالماشتۇرۇشى) ، ھەر خىل ئوقۇتۇش ياردەمچى دېتاللىرىنى ياساش قاتارلىقلاردا كەڭ قوللىنىلىدۇ.

(3كومپيۇتېر تىلشۇناسلىقى ۋە ئەنئەنىۋى تىلشۇناسلىقنىڭ پەرقى

كومپيۇتېر تىلشۇناسلىقى ۋە باشقا ئەنئەنىۋى تىلشۇناسلىقنىڭ تەتقىقات ئوبيېكتى ئوخشاشلا ئىنسان تىلىدىن ئىبارەت بولسىمۇ تەتقىقات ئۇسۇلى ۋە ۋاستىسىدە پەرقلەر بولغاچقا ئۇلار ئوتتۇرىسىدا مۇئەييەن ئوخشىماسلىقلار مەۋجۇت.

(1)كومپيۇتېر تىلشۇناسلىقى كومپيۇتېرنىڭ تىلدىكى ئومۇمىيلىققا ۋە يۈتۈنلۈككە ئىگە ئادەتتىكى مەسىلىلەرنى ھەل قىلىشنى تەتقىق قىلسا ، ئەنئەنىۋى تىلشۇناسلىقتا كۆپرەك تىلدىكى ئالاھىدە مەسىلىلەر ، ھادىسىلەر تەتقىق قىلىنىدۇ .

(2)كومپيۇتېر تىلشۇناسلىقى تىل تەتقىقاتىدا ئاۋۋال تەھلىل قىلىپ ئاندىن چۈشىنىدۇ ، چۈشىنىش تەھلىل قىلىشنىڭ نەتىجىسى. ئەنئەنىۋى تىلشۇناسلىقتا ئاۋۋال چۈشىنىلىپ ئاندىن تەھلىل قىلىنىدۇ ، چۈشىنىش تەھلىل قىلىشنىڭ ئالدىنقى شەرتى .

(3)كومپيۇتېر تىلشۇناسلىقىنىڭ تەتقىقات نەتىجىسى چوقۇم كومپيۇتېر تەبىئىي تىلنى بىر تەرەپ قىلىش داۋامىدا سىنىلىدۇ. كومپيۇتېر تىلشۇناسلىقى تەتقىقاتىنىڭ يەكۈنى ئەمەلىيەتتە قوللىنالايدىغان بولۇشى كېرەك ، يەنى نەزەرىيىنىڭ ئەمەلىيەتچانلىقىغا ئەھمىيەت بېرىلىدۇ. ئەنئەنىۋى تىلشۇناسلىقتا قائىدە-داۋلى سۆزلەش مۇھىم ، كۈچلۈك لوگىكىلىققا ئىگە بولۇشى تەلەپ قىلىنىدۇ.

(4)ئەنئەنىۋى تىلشۇناسلىقتا تەسۋىرلەش ئاساسىي ئورۇندا تۇرىدۇ ؛ كومپيۇتېر تىلشۇناسلىقىدا تىل نەزەرىيىسىنىڭ كۈچلۈك ئەمەلىي مەشغۇلاتچانلىققا ئىگە بولۇشى تەلەپ قىلىنىدۇ ، شۇڭا ھەر بىر جۈملىدىكى گىرامماتىكىلىق ئۇچۇر ۋە مەنە ئۇچۇرلىرىنىڭ يۇكسەك دەرىجىدە فورمۇلالاشتۇرۇلغان بولۇشى ، كومپيۇتېر چۈشىنەلەيدىغان دەرىجىدە قائىدىلەشتۈرۈلگەن بولۇشى تەلەپ قىلىنىدۇ .

.2كومپيۇتېر تىلشۇناسلىقىنىڭ ئۇيغۇر تىلى تەتقىقاتىدا قوللىنىلىشى

كومپيۇتېر تىلشۇناسلىقى بارلىققا كەلگەندىن تارتىپ ئارىلىق پەن بولۇش سۈپىتى بىلەن كومپيۇتېر تەتقىقاتى ۋە تىلشۇناسلىق تەتقىقاتىدا ئوخشاشلا مۇھىم ئورۇن تۇتۇپ كەلدى ؛ جۇملىدىن تىلشۇناسلارمۇ بۇ ساھەگە دىققەت قىلدى. ئىنگلىز ۋە خەنزۇ تىللىرىدا مەخسۇس كومپيۇتېر تىلشۇناسلىقى بىلەن شۇغۇللىنىدىغان تەتقىقاتچىلار ، مۇتەخەسسىسلەر قوشۇنى شەكىللەندى ۋە نۇرغۇن تەتقىقات نەتىجىلىرىگە ئېرىشتى. بۇ نەتىجىلەر بىر تەرەپتىن تىلشۇناسلىق تەتقىقاتىنى كۈچەيتتى ، يەنە بىر تەرەپتىن شۇ تىل-يېزىقنىڭ كومپيۇتېردا قوللىنىلىشىنى ۋە نۇرغۇن يېڭى تېخنىكا ، يۇمشاق دېتاللارنىڭ ياسىلىشىنى ئىلگىرى سۈردى. جۇملىدىن ئۇيغۇر تىلى ۋە كومپيۇتېر تەتقىقاتچىلىرىمۇ بۇ پەنگە دىققىتىنى قاراتتى ، شۇنداقلا بەزى نەتىجىلەرنى قولغا كەلتۈردى. لېكىن تەتقىقاتتا بەزى مەسىلىلەر ساقلىنىۋاتىدۇ ، نۇرغۇن مەسىلىلەر بىزنىڭ تەتقىق قىلىشىمىزنى كۈتىۋاتىدۇ. تۆۋەندە بۇ جەھەتتە توختىلىپ ئۆتىمىز.

(1ھازىرقى تەتقىقات ئەھۋالى

ئۇيغۇرلاردا كومپيۇتېر تىلشۇناسلىقى تەتقىقاتىنى تېخى ئەمدىلا باشلاندى دېيىشكە بولىدۇ، شۇنداقتىمۇ بەزى تىلغا ئېلىشقا تېگىشلىك نەتىجىلەر بارلىققا كەلدى. ئەمما ھازىرچە بۇ خىل تەتقىقات بىلەن شۇغۇللىنىدىغانلار ئاساسلىقى كومپيۇتېر تەتقىقاتچىلىرى ۋە بەزى كومپيۇتېر ھەۋەسكارلىرى بىلەن چەكلەنگەن. ئۇنىڭ ئۈستىگە ئۇلارنىڭ تەتقىقاتىدا كومپيۇتېر قىسمى ئاساسىي ئورۇندا تۇرىدۇ. تىلشۇناسلىق نۇقتىسىدىن كومپيۇتېر تىلشۇناسلىقى تەتقىقاتى بىلەن شۇغۇللىنىدىغانلار يوق دېيەرلىك. تۆۋەندە تىلغا ئالماقچى بولغانلىرىمىزنىمۇ ساپ كومپيۇتېر تىلشۇناسلىقىنىڭ تەتقىقات نەتىجىسى دەپ ئېيتالمايمىز، لېكىن بۇلار تىل بىلەن كومپيۇتېرنىڭ مۇناسىۋىتىنى چىقىش قىلىپ ئېلىپ بېرىلغان كومپيۇتېر تىلشۇناسلىقىغا مۇناسىۋەتلىك تەتقىقاتلارنىڭ نەتىجىلىرىدىن ئىبارەت.

(1)ئۇيغۇرچە يېزىقنى كومپيۇتېردا بىر تەرەپ قىلىش دېتاللىرى ياسالدى.

ئالدىنقى ئەسىرنىڭ 90-يىللىنىڭ بېشىدا DOS مۇھىتىدا ئۇيغۇرچە يېزىق كىرگۈزۈش سىستېمىسى «بۇغدا» مۇۋاپىقىيەتلىك ياسالغاندىن تارتىپ ئۇيغۇرچە يېزىقنى كومپيۇتېرغا كىرگۈزۈش، ھەر خىل يۇمشاق دېتاللاردا ئۇيغۇرچىنى تونۇتۇش، بىر قىسىم مىللىي يېزىقتىكى يۇمشاق دېتال ۋە ئۈسكۈنىلەرنى ئىشلەپچىقىرىش جەھەتلەردە خۇشاللىنارلىق نەتىجىلەر بارلىققا كەلدى.

ئەڭ دەسلەپ شىنجاڭ ئۇنىۋېرسىتېتى، ئۇيغۇر سوفت كومپيۇتېر شىركىتى قاتارلىق تەتقىقات ئورۇنلىرى كېيىنچە ئۇنىڭغا ئەگىشىپ يەنە باشقا تارماق ۋە شەخسلەرنىڭ تىرىشچانلىقى ئاساسىدا ھازىر ئۇيغۇرچە يېزىقنى كومپيۇتېرغا كىرگۈزۈش، تونۇتۇش ۋە ئاددىي بىر تەرەپ قىلىشلار ئاساسىي جەھەتتىن ئىشقا ئاشتى.

(2)ئاددىي كوررېكتورلۇق قۇراللىرى ياسالدى.

بىزنىڭ بىلىشىمىزگە ئاساسلانغاندا، ھازىرچە ئۇيغۇر سوفت شىركىتى ئىشلەپچىقارغان «ئەل كوررېكتور» يۇمشاق دېتالى (بازارغا سېلىنغان)، شىنجاڭ ئۇنىۋېرسىتېتىدىكى تەتقىقاتچىلار تەتقىق قىلىپ ياسىغان «كۆپ تىللىق مۇھىتتىكى ئۇيغۇرچە يېزىق كوررېكتورلۇق سىستېمىسى» (تېخى تارقىتىلمىغان) ۋە غەيرەت توختى ئەپەندى ياسىغان كوررېكتورلۇق ئىقتىدارى بار «ئۇيغۇر ئېدىت UyghurEdit» قورالى (ھەقسىز تارقىتىلغان) قاتارلىقلارنى تىلغا ئېلىش مۇمكىن. بۇلار بۇ جەھەتتىكى دەسلەپكى ئىزدىنىش بولغاچقا تېخى ئىقتىدارى تازا مۇكەممەل، توغرىلىق دەرىجىسى ئانچە يۇقىرى ئەمەس.

(3)ئېلېكترونلۇق لۇغەتلەر ياسالدى.

ئېلېكترونلۇق لۇغەت گەرچە كومپيۇتېر تىلشۇناسلىقىنىڭ بىۋاسىتە تەتقىقات نەتىجىسى بولمىسىمۇ، لېكىن ھەر خىل ئېلېكترونلۇق لۇغەتلەر (كومپيۇتېر لۇغىتى ۋە مەخسۇس ئېلېكترونلۇق ئۈسكۈنە ھالىتىدىكى لۇغەتنى ئۆز ئىچىگە ئالىدۇ)نى تەتقىق قىلىپ ياساشتا كومپيۇتېر تىلشۇناسلىقىنىڭ تەتقىقاتى كەم بولسا بولمايدۇ. مەسىلەن، ئۇيغۇر سوفت شىركىتى ئىشلەپچىقارغان «ئۇيغۇرچە- خەنزۇچە لۇغەت»، ئىرپان شىركىتىنىڭ «ئىرپان ئېلېكترونلۇق لۇغىتى» ۋە باشقىلار.

(4)ماشىنا تەرجىمە تەتقىقاتى ئېلىپ بېرىلىپ بەلگىلىك نەتىجە قولغا كەلتۈرۈلدى.

ھازىرغىچە تېخى كەڭ ئومۇملاشقان ئۇيغۇرچە تەرجىمە يۇمشاق دېتالى ياساپ چىقىلمىدى، بۇنىڭغا كېرەكلىك تەتقىقات، ئادەم كۈچى، مالىيە كۈچى يېتەرلىك ئەمەس. شۇنداقتىمۇ بىر قىسىم تەتقىقاتچىلار بۇ جەھەتتە ئىزدىنىپ بەزى نەتىجىلەرنى قولغا كەلتۈردى. شىنجاڭ ئۇنىۋېرسىتېتى قاتارلىق تەتقىقات ئورۇنلىرىدا بۇ جەھەتتە تەتقىقاتلار قىلىنىۋاتىدۇ.

(5)تاۋۇش، بوغۇم، سۆز چاستوتا تەتقىقاتى قىلىندى.

كومپيۇتېردىن پايدىلىنىپ كۆپ مىقداردىكى تىل ماتېرىيالىغا قارىتا تاۋۇش، بوغۇم ۋە سۆزلۈكلەرنىڭ چاستوتىسى جەھەتتىن ئوخشىمىغان نۇقتىدىن تەتقىقاتلار داۋاملىشىۋاتىدۇ؛ مەسىلەن، ئۇيغۇر سوفت شىركىتى ئەدەبىي ئەسەرلەرنى ئاساس قىلغان ھالدا 10 مىليون سۆزلۈك تىل ماتېرىيالىغا تاۋۇش، بوغۇم ۋە سۆزلۈك جەھەتتىن ستاتىستىكا ئېلىپ بېرىپ نەتىجىسىنى تور بېتىدە ئېلان قىلدى؛ بىلىشىمىزچە ئۇنىڭدىن سىرت شىنجاڭ پېداگوگىكا ئۇنىۋېرسىتېتى ۋە شىنجاڭ ئۇنىۋېرسىتېتى قاتارلىق ئورۇنلاردىكى تەتقىقاتچىلارمۇ بۇ جەھەتتە ئىزدىنىۋاتىدۇ. لېكىن بۇ ستاتىستىكىلاردا بەزى مەسىلىلەر ساقلانغان. يىغىپ ئېيتقاندا تېخى تولۇق ئېتىراپ قىلىنغان بىرەر نەتىجە چىققىنى يوق.

(6)سۆز، سۆز بىرىكمىلىرىنى پارچىلاش ۋە ئاپتوماتىك بەلگە قويۇش ئۇسۇلى توغرىسىدا تەتقىقاتلار ئېلىپ بېرىلدى.

سۆز ۋە سۆز بىرىكمىلىرىگە ئاپتوماتىك ھالدا گراماتىكىلىق بەلگە قويۇش ئۇيغۇر كومپيۇتېر تىلشۇناسلىقىدىكى مۇھىم ئاساس خاراكتېرلىق خىزمەت بولۇپ، باشقا تەتقىقاتلار مۇشۇ ئاساستا راۋاجلىنىدۇ. شىنجاڭ پېداگوگىكا ئۇنىۋېرسىتېتىدىن پروفېسسور يۈسۈپ ئەبەيدۇللا باشتىن- ئاخىرى بۇ جەھەتتە كۆپرەك ئىزدەنگەن تەتقىقاتچىلارنىڭ بىرى، ئۇنىڭدىن سىرت باشقا يولداشلارمۇ تەتقىق قىلىۋاتىدۇ. لېكىن كۆپىنچە تەتقىقات نەتىجىسى تېخى ئاشكارىلانمىغان ھالەتتە تۇرۇۋاتىدۇ، بۇ مەسىلىنى ھەل قىلالايدىغان يۇمشاق دېتاللارمۇ ئېلان قىلىنمىدى.

(7)يېزىق تونۇتۇش سىستېمىسى ياسالدى.

يېزىق تونۇتۇش سىستېمىسى دېگەندە قوليازما ياكى باسما ماتېرىياللاردىكى خەتلەرنى كومپيۇتېردا ئاپتوماتىك تونۇتۇش ئىقتىدارىغا ئىگە قۇراللار كۆزدە تۇتۇلىدۇ. ئىگىلىگەن ئەھۋالدىن قارىغاندا، شىنجاڭ ئۇنىۋېرسىتېتىدىكى تەتقىقاتچىلار دەسلەپكى قەدەمدە يېزىق تونۇتۇش سىستېمىسىنى ياساپ مۇۋاپىقىيەت قازانغان، بۇ سىستېما پات ئارىدا بازارغا سېلىنىدىكەن. بۇنىڭدىن باشقا يەنە قولدا يېزىپ كومپيۇتېرغا خەت كىرگۈزۈش قورالى ئالىلقاچان ياسىلىپ ئىشلىتىلىشكە باشلىدى، لېكىن بۇ قۇرال قوشۇمچە ئۈسكۈنىنىڭ ياردىمىدە ئىشلەيدۇ، تېخى كومپيۇتېر بىلەن بىر گەۋدىلىشىپ بولالمىدى.

(2تەتقىقاتتا ساقلانغان مەسىلىلەر

گەرچە ئۇيغۇر تىلى كومپيۇتېر تىلشۇناسلىقىغا دائىر يۇقىرىقىدەك نەتىجىلەر بارلىققا كەلگەن بولسىمۇ ، باشقا مىللەتلەردىكىگە ۋە دەۋرىمىزدىكى ئىلغار سەۋىيىگە سېلىشتۇرغىنىمىزدا نۇرغۇن مەسىلىلەر ساقلانغان:

(1)تەتقىقات نەتىجىلىرىنىڭ ئاشكارىلىنىش دەرىجىسى تۆۋەن.

بىلىشىمىزچە ، كومپيۇتېر تىلشۇناسلىقىغا ئائىت تەتقىقات بىلەن شۇغۇللىنىۋاتقانلار مەيلى كومپيۇتېر تەتقىقاتچىسى ياكى تىلشۇناسلىق تەتقىقاتچىسى بولسۇن خېلى سانىنى ئىگىلىسىمۇ ، نۇرغۇن تەتقىقات نەتىجىلىرى پەقەت شەخسلەرنىڭ قولىدىلا ساقلىنىپ باشقىلارنىڭ پايدىلىنىشى ئۈچۈن ئېلان قىلىنمايۋاتىدۇ. بەزى تەتقىقاتلارنىڭ قىلىنىۋاتقانلىقىنى ، ھەتتا ئاياغلاشقانلىقىنى ئاڭلىساقمۇ ، بىر نەچچە يىلغىچە بۇنىڭغا ئائىت ماقالە ياكى ئېلان قىلىنغان باشقا يۇمشاق دېتاللارنى كۆرەلمەيمىز. بەلكىم شۇ تەتقىقاتچى تەتقىقاتنى تېخىمۇ مۇكەممەللەشتۈرۈشنى ئويلىغان بولۇشى ، يېرىم يولدا توختاپ قالغان بولۇشى ، ئىقتىسادىي مەنپەئەت ياكى باشقا ئامىللارنى ئويلاپ نەتىجىسنى ئېلان قىلمىغان بولۇشى مۇمكىن. ئەلۋەتتە بۇنداق قىلىشمۇ توغرا ، لېكىن بىرەر تەتقىقاتنىڭ پۈتۈنلەي مۇكەممەل بولۇپ كېتىشى ناتايىن ، ئۇنىڭ ئۈستىگە بۇ جەھەتتە ئىقتىسادىي ئۇنۇم يارىتىشنىلا كۆزلەش ئانچە ئاقىلانىلىق ئەمەس. شۇڭا تەتقىقات بولغانىكەن ، نەتىجىسى ئېلان قىلىنسا ، باشقىلار ئۇنىڭدىن پايدىلانسا ھەم شۇ ئاساستا يەنىمۇ تەتقىقات چوڭقۇرلاشسا تېخىمۇ ياخشى بولاتتى.

(2)تەتقىقاتتا ھەمكارلىق كەمچىل ، ئەمەلىي ئىشلىتىش قىممىتى تۆۋەن بولماقتا.

ئۇيغۇر كومپيۇتېر تىلشۇناسلىقى تەتقىقاتىدا كەمچىل بولۇۋاتقىنى ھەمكارلىق بولۇپ ، تەتقىقاتلىرىمىز تارقاق ھەم تەكرار بولىۋاتىدۇ. ئۇنىڭ ئۈستىگە رايونىمىزدا بۇ جەھەتتە تەتقىقاتقا ئۇيۇشتۇرىدىغان نوپۇزلۇق ئورگانلار بولمىغاچقا ھەمكارلاشماق ناھايىتى تەس. كومپيۇتېر تىلشۇناسلىقى بىر ئارىلىق پەن بولغاچقا تەتقىقاتچىدىن كومپيۇتېر بىلىمى ھەم تىلشۇناسلىق بىلىمى بولۇشنى تەلەپ قىلىدۇ .لېكىن بۇ ئىككىسىنىڭ بىر ئادەمدە تەڭ بولۇشى ناتايىن ، بۇ خىل ئەھۋالدا تىل تەتقىقاتچىلىرى بىلەن كومپيۇتېر تەتقىقاتچىلىرى ھەمكارلاشقاندىلا بىر بىرىنى تولۇقلاپ زور نەتىجىلەرنى قولغا كەلتۈرۈلەيدۇ.

بىر قاتار سەۋەبلەر تۈپەيلى بەزى تەتقىقات نەتىجىلىرىنىڭ ئەمەلىي ئىشلىتىش قىممىتى بەكلا تۆۋەن بولىۋاتىدۇ. تەتقىقات ئۈچۈن ياكى تەكشۈرۈشتىن ئۆتتۈپلىش ئۈچۈنلا تەتقىقات قىلىدىغان ئەھۋال مەۋجۇت. بىر قىسىم چوڭ-چوڭ تەتقىقات تېمىلىرىنىڭ تەتقىقات نەتىجىلىرى تېما تەكشۈرۈپ ئۆتكۈزۈۋېلىش گۇرۇپپىسى تەرىپىدىن ئۆتكۈزۈۋېلىنغاندىن كېيىنمۇ ئەمەلىي ئىشلىتىلمەي مەخپىي ساقلىنىپ قېلىۋاتىدۇ. دۆلەت مەبلەغ ئاجراتقان بۇ تەتقىقاتلار نېمە ئۈچۈن ئومۇمنىڭ پايدىلىنىش ئۈچۈن ئېلان قىلىنمايدۇ؟ ئەلۋەتتە بۇ بىر ئويلىنىپ بېقىشقا تېگىشلىك مەسىلە.

(3)ساپ تىل تەتقىقاتىدا قوللىنىش يېتەرلىك بولمايۋاتىدۇ.

ھازىرغىچە كومپيۇتېر تىلشۇناسلىقى تەتقىقات دېيىلگەندە ئاساسلىقى كومپيۇتېر ساھەسىدىكىلەرنىڭ تەبئىي تىل بىر تەرەپ قىلىش (自然语言

处理)قا ئائىت تەتقىقاتى كۆپرەك نەزەردە تۇتۇلىدۇ. ئەمەلىيەتتە كومپيۇتېر تىلشۇناسلىقى ساپ تىل تەتقىقاتىدىمۇ مۇھىم ئەھمىيەتكە ئىگە ،

كومپيۇتېر تىلشۇناسلىقىنىڭ ئۇسۇلىدىن پايدىلىنىپ تىل ساھەسىدە نۇرغۇن تەتقىقاتلارنى قىلىش ۋە بەزى مەسىلىلەرنى ھەل قىلىش مۇمكىن. بۇنىڭ ئۈچۈن تىل تەتقىقاتچىلىرى بۇ ساھەگە نەزىرىنى ئاغدۇرۇشى ، يېڭى تەتقىقات ئۇسۇللىرىنى ئىگىلىشى ھەم بىر تۈركۈم كومپيۇتېرغا پىششىق بولغان تىل تەتقىقاتچىلىرى يېتىشىپ چىقىشى كېرەك.

(4)تەتقىقاتقا سېلىنغان مەبلەغ يېتەرلىك ئەمەس.

سەل قاراشقا بولمايدىغان بىر ئامىل شۇكى ، ھازىر ئۇيغۇر كومپيۇتېر تىلشۇناسلىقى تەتقىقاتىغا كېتەرلىك مەبلەغ كەمچىل. بۇمۇ تەتقىقاتنىڭ ئىلگىرىلىشىگە توسالغۇ بولىۋاتىدۇ. شۇنداقتىمۇ ئامال قىلىپ مۇناسىۋەتلىك تارماقلارنىڭ ، مۇمكىن بولسا شەخسلەرنىڭ قوللىشىغا ئېرىشىش كېرەك. مۇناسىۋەتلىك ئورۇنلار ۋە تەتقىقات تارماقلىرىمۇ كومپيۇتېر تىلشۇناسلىقى تەتقىقاتىغا يېتەرلىك ئەھمىيەت بېرىشى كېرەك.

(3بۇندىن كېيىن جىددىي ھەل قىلىشقا تېگىشلىك مەسىلىلەر

(1)ئۇيغۇر تىلىنىڭ كومپيۇتېرغا ماسلاشقان بەزى ئىملا قائىدىلىرىنى تۈزۈپ چىقىش ، توغرا ئىملانى ئومۇملاشتۇرۇش كېرەك.

يېقىندىكى بىر تەتقىقاتىمىزغا ئاساسلانغاندا ئۇيغۇر يېزىقىدىكى نەشر ماتېرىياللىرىدا ئېغىر دەرىجىدىكى ئىملا خاتالىقلىرى مەۋجۇت ئىكەن. مەسىلەن ، خېلى نوپۇزلۇق گېزىت -ژۇرنالدىن ئېلىنغان 200 مىڭ سۆزلۈك تېكستنى تەھلىل قىلغىنىمىزدا ئىملاسىدا مەسىلە بار سۆز 1.5 پىرسەنت ئەتراپىدا بولۇپ ، بۇ كىچىك ساندەك كۆرۈلسىمۇ 200مىڭ سۆزگە نىسبەتەن 3 مىڭ سۆزدە ئىملا خاتالىقىنىڭ كۆرۈلۈشى ھەقىقەتەن كىچىك ئىش ئەمەس. گەرچە بۇ ئىملا خاتالىقلىرى ئوقۇرمەنلەرگە سىزىلمەسلىكى مۇمكىن ، لېكىن كومپيۇتېر بىر ماشىنا بولغاچقا بىر سۆزدە بىرەر ھەرپ خاتا ياكى ئارتۇق-كەم بولسا ئۇنى توغرا تونىيالمايدۇ.

بۇ ئىملا خاتالىقلىرىدىن سىرت ، كومپيۇتېردىكى تېكىستلەرنىڭ ئىملاسىدا بەزى قالايمىقانچىلىقلار مەۋجۇت. يېزىقتىكى ئىملا ۋە تىنىش بەلگىلىرى توغرىسىدا ئېنىق قائىدىلەر تۈزۈلگەن ، گەپ شۇنى ئومۇملاشتۇرۇشتا. لېكىن كومپيۇتېردىكى ئىملاغا نىسبەتەن بەزى قائىدىلەرنى بېكىتىش زۆرۈر. مەسىلەن ، سۆزلەرنى ئايرىشتا قوللىنىلىدىغان بوشلۇق بەلگىسىنىڭ قوللىنىلىشى بەكلا قالايمىقان بولىۋاتىدۇ ، خەتلەر بېسىپ چىقىرىلغاندا بوشلۇق

گەرچە مۇھىم بولمىسىمۇ ، كومپيۇتېردا سۆزلەرنى پارچىلاش ، سۆزلەرنى تەھلىل قىلىش جەھەتلەردە ئۇنىڭ رولى بەك چوڭ. بىزنىڭ قارىشىمىزچە پەش (،) ، چېكىت (.) ، ئۈندەش (!) ، سوراق (؟) ، چېكىتلىك پەش (؛) ۋە قوش چېكىت (:) بەلگىسىنىڭ ئالدىغا ئەمەس ، كەينىگە بىر بوشلۇق بەلگىسى قويۇلىشى ؛ ئىككى تىنىش بەلگىسى بىللە كەلسە بىرلا بوشلۇق بەلگىسى قويۇلۇشى كېرەك ؛ ھەر خىل تىرناقلارنىڭ تېشىغا بوشلۇق بەلگىسى قويۇلىشى كېرەك. ئۇندىن سىرت سىزىقچە (-) بىلەن سىزىق (__) پەرقلەندۈرۈلىشى ؛ ئۇيغۇر تىلىدا بولمىغان "“‘ ، قاتارلىق تىنىش بەلگىلىرى ئىشلىتىلمەسلىكى ، مۇمكىن بولسا مۇناسىۋەتلىك تارماقلار كومپيۇتېردىكى ئىملا توغرىسىدا ئىنچىكرەك بەزى قائىدىلەرنى بېكىتىشى كېرەك. چۈنكى بۇ جەھەتتىكى قالايمىقانچىلىق تىلىمىزنى كومپيۇتېردا بىر تەرەپ قىلىش ۋە تەتقىق قىلىشقا توسالغۇ بولىدۇ.

(2)ئۆلچەملىك كوررېكتورلۇق قوراللىرىنى ياساپ چىقىش كېرەك.

توغرا ئىملانىڭ ئومۇملاشماسلىقىدىكى يەنە بىر سەۋەب ئۆلچەملىك كوررېكتورلۇق قوراللىرىنىڭ بولماسلىقىدا. بەزى ئىملا خاتالىقلىرىنى بايقاش خېلىلا قىيىن ، ئەگەر توغرىلىق دەرىجىسى يۇقىرى بولغان كوررېكتورلۇق يۇمشاق دېتاللىرى ياسالسا كوررېكتورلۇق خىزمىتى تېزلەشكەن ھەم سۈپىتىمۇ يۇقىرى بولغان بولاتتى .ئادەتتىكى كىشىلەرمۇ خەت كىرگۈزۈش داۋامىدا كوررېكتورلۇق قورالىنىڭ ياردىمىدە ئىملادىكى خاتالىقلىرىنى تۈزىتىۋالغان بولاتتى.

(3)ئۇچۇر بىر تەرەپ قىلىشنى مەقسەت قىلغان ئۇيغۇر تىلىنىڭ گىرامماتىكا ئۇچۇر ئامبىرى (لۇغىتى) قۇرۇش كېرەك.

خەنزۇ تىلىدا ئالىملاچان “现代汉语语法信息词典” تەتقىق قىلىنىپ تۈزۈپ چىقىلدى » .تىلنىڭ گىرامماتىكا ئۇچۇرلىرى لۇغىتى« تىل تەتقىقاتى ۋە تىل ئۇچۇرلىرىنى بىر تەرەپ قىلىشتىكى ئۇل خاراكتېرلىق قۇرۇلۇش بولۇپ ، بۇنداق لۇغەتتە تىلدىكى ھەر بىر سۆزگە قەدەر ئۇنىڭ گىرامماتىكىلىق ئالاھىدىلىكى ئىنتايىن تەپسىلىي بېرىلىدۇ. بۇنداق چوڭ لۇغەت نۇرغۇن تەتقىقاتچىنىڭ ھەمكارلىشىپ ئۇزۇن مۇددەت ئىزدىنىشى ئاساسىدا ۋۇجۇدقا چىقىدۇ. ئەگەر بۇ لۇغەت ياساپ چىقىلسا كومپيۇتېر تىلشۇناسلىقى تەتقىقاتى ۋە باشقا تەتقىقاتلار زور دەرىجىدە ئالغا باسىدۇ.

(4)ئۇيغۇر تىلىنىڭ سۆز چاستوتىسى تەتقىقاتىنى قىلىش كېرەك.

ھازىرغا قەدەر ئۇيغۇر تىلىدىكى سۆزلەر ئىشلىتىلىش چاستوتىسىغا ئاساسەن دەرىجىگە تۇرغۇزۇلمىدى. بۇ جەھەتتە ئۇيغۇر تىلى تەتقىقاتى باشقا تىللاردىن نەچچە يىل كېيىن قالدى. بۇنىڭدىكى ئاچقۇچلۇق مەسىلە تېخى كەڭ كۆلەمدىكى سۆز چاستوتىسىنى ستاتىستىكىلاش خىزمىتى ئىشلەنمىدى ياكى ئىشلەنسىمۇ نوپۇزلۇق ئورگانلارنىڭ ئېتىراپ قىلىشىغا ئېرىشەلمىدى.

سۆز چاستوتىسىنى ستاتىستىكىلاشتا ھەر خىل ژانىر ، ھەر خىل تېمىدىكى ۋەكىللىككە ئىگە بولغان تىل ماتېرىياللىرىدىن كۆپ مىقداردا توپلاپ كومپيۇتېر ئارقىلىق سۆزلەرنىڭ ئىشلىتىلىش ۋە تارقىلىش چاستوتىسىنى ستاتىستىكا قىلىشقا توغرا كېلىدۇ. بۇنىڭغا ئائىت تەتقىقاتلار ئالىملاچان باشلاندى ، يېقىن كەلگۈسىدە نەتىجە چىقىشى مۇمكىن. ئەگەر بۇ خىزمەت تاماملانسا ، تىل ئوقۇتۇشى ۋە تەتقىقاتىدا ھەل قىلغۇچ رول ئوينايدۇ

پايدىلانمىلار:

[1]ۋېڭ فۇلياڭ ، ۋاڭ يىيى قاتارلىقلار تۈزگەن » :كومپيۇتېر تىلشۇناسلىقى توغرىسىدا بايان» ، جۇڭگو ئىجتىمائىي پەنلەر ئاكادېمىيىسى نەشرىياتى ، 1998

[2]فېڭ جىۋېي: «كومپيۇتېر تىلشۇناسلىقىدىن ئاساس» ، شاڭۋۇ نەشرىياتى ، 2001

[3]ليۇ يىڭ: «كومپيۇتېر تىلشۇناسلىقى» ، چىڭخۇا ئۇنىۋېرسىتېتى نەشرىياتى ، 2002

[4]خۇاڭ چاڭنىڭ ، لى جۈەنزى قاتارلىقلار: «تىل ماتېرىيال ئامبىرى تىلشۇناسلىقى» ، شاڭۋۇ نەشرىياتى ، 2002