ڊاڪٽر مظهر علي ڏوتيو
اسسٽنٽ پروفيسر، ڪمپيوٽر سائنس ڊپارٽمينٽ
بينظير ڀٽو شهيد يونيورسٽي لياري، ڪراچي، سنڌ، پاڪستان

سنڌي ٻوليءَ جي ڪمپيوٽيشنل لسانيات وارن مسئلن جو حل ۽ تجزيو

AN ANALYSIS AND SOLUTION OF COMPUTATIONAL LINGUISTICS PROBLEMS OF SINDHI LANGUAGE

Abstract

Sindhi language is one of the oldest and indigenous languages of the world having complex grammar and rich morphology. It is read, written and spoken all over the world. There are good number of newspapers, magazines, social websites, blogs, general and specific websites available online in Sindhi language which provide big data in form of Sindhi text. The research is in progress on several languages of the world in the field of computational linguistics and natural languages process (NLP). However, the more research work has done on English language. Therefore, this language is resourced language of world computationally. To work on languages other than English, especially on right hand written languages, is really difficult job because there are less computerized resourced available online for these types of languages. Nevertheless, a good work has been done on Arabic and Urdu languages. Sindhi language has remained poor and less resourced language in the field of computational linguistics and natural languages process. Viewing the computational linguistics problems of Sindhi language, a scientific methodology is designed, which helps in developing different types of algorithms and NLP based tools for the solution of computational linguistics and NLP problems of Sindhi language. This research study presents the computational linguistics and natural languages process related problems of Sindhi language text and their solution in form of word tokenization, universal and Sindhi part of speech (POS) tagging, syntactic parsing and morphological analysis. **Key words:** Sindhi; NLP; Computational Linguistics; UPOS; SPOS; Tagging; Syntactic parsing

تعارف

ڪمپيوٽيشنل لسانيات بنيادي طورلسانيات جي شاخ آهي، جيڪا ڪمپيوٽر سسٽم جي ذريعي قدرتي ٻولين کي سمجهي انهن کي ترڪيب ڏئي ٿي ۽ انهن جو تجزيو ڪري ٿي ته جيئن انهن ٻولين کي مختلف ڪمپيوٽرائيزڊ ڪمن لاءِ استعمال ڪرڻ سان گڏ دنيا جي سمورين ٻولين کي هڪ عالمي گرامر تحت هڪ ڪري سگهجي. ان عمل سان دنيا حقيقي معنا ۾ عالمي ڳوٺ ۾ تبديل ٿي سگهندي. ڪمپيوٽيشنل لسانيات جي شعبي ۾ ڪم ڪندڙ ڪمپيوٽر سائنٽسٽس کي ڪمپيوٽيشنل لسانيات جو ماهر چئبو آهي. گڏريل ٽن ڏهاڪن کان ڪمپيوٽيشنل لسانيات، ڪمپيوٽيشنل انجنيئرنگ ۽ هٿراڌو ڏاهپ کي استعمال ڪندي بهترين قسم جا الگوريٽم ۽ اوزار ٺاهيا ويا آهن (1) جن دنيا جي ٻولين کي عالمي گرامر۽ قدرتي ٻولين جي پروسس جي گهرجن مطابق تمام گهڻي ترقي ڏياري آهي. هاڻي ڪمپيوٽر پنهنجو پاڻ لفظن۽ نشانين وغيره کي سڃاڻي ٿو، انهن کي ترڪيب ڏئي ٿو، انهن کي گرامر مطابق نشان ڏئي ٿو، لفظن جي ڀنڊار يا ڪارپس مان ڄاڻ حاصل ڪري ٿو، جنهن سان دنيا ۾ تيزيءَ سان ترقي ٿي رهي آهي. خاص ڪري انڊيسٽريز ترقي ڏياريندڙ تنظيمن ۽ ادارن کي تمام گهڻو فائدو پيو آهي. سماجي ويب سائيٽس يا ميڊيا کي ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس جي ڪري تمام گهڻو فائدو پيو آهي. هاڻي اهڙي قسم جون ويب سائيٽس ماڻهن جي نفسيات، سوچ، احساس، راين ۽ خيالن جو بهتر طريقي سان تجزيو ڪري سگهن ٿيون. اهڙيءَ طرح مختلف قسمن جون اڳڪٿيون ڪري ماڻهن کي سهولتون مهيا ڪن ٿيون. ڪمپيوٽيشنل لسانيات مصنوعي ذهانت ذريعي ٺاهيل روبوٽس ۽ ٻين اوزارن لاءِ پڻ اهم ڪم ڪري ٿي، ڇاڪاڻ ته روبوٽس کي قدرتي ٻولين سيکارڻ لاءِ، ٻولين ۾ موجود احساسن وغيره کي سمجهڻ لاءِ ڪمپيوٽيشنل لسانيات بنيادي ۽ اهم ڪردار ادا ڪري ٿي.

هن وقت ڪمپيوٽر سائنس ۾ مختلف شعبن سان گڏ ڪمپيوٽيشنل لسانيات۽ نيچرل لينگئيجز پروسس تي تمام گهڻو تحقيقي ۽ ترقياتي ڪم ٿي رهيو آهي، جنهن ۾ دنيا جي مختلف ٻولين کي پاڻ ۾ ڳنڍڻ سان گڏ مختلف ٻولين جي بناوٽ کي به سمجهيو وڃي ٿو. تنهنڪري ٻولين جي ترجمي ڪرڻ ، احساسن کي ڳولهڻ ۽ انهن جي لفظن يا متن جي تجزيي ڪرڻ، لفظن کي سمجهڻ ۽ ڄاڻ جي حصول ۾ آساني ٿئي ٿي. ان سلسي ۾ انگريزي ٻوليءَ تي تمام گهڻوڪم ٿيو آهي(2) جڏهن ته ساڄي پاسي کان لکجندڙ ٻولين ۾ عربي ۽ اردوءَ تي گهڻو ڪم ٿيو آهي.جنهنڪري اهي ٻوليون موجوده دور ۾ ترقي ڪندڙ ٻوليون سڏيون وڃن ٿيون. سنڌي ٻولي دنيا جي پراڻين ٻولين مان هڪ ٻولي آهي، جنهنجي بناوٽ، صرفيه ۽ گرامر دنيا جي ڪيترين ئي ٻولين کان شاندار ۽ شاهوڪار آهن. (3، 4، 5)، پر ان ٻوليءَ تي ڪمپيوٽر لسانيات ۽ نيچرل لينگئيجز پروسس جي

حوالي سان تمام گهڻ ڪم ٿيو آهي، جنهنڪري هيءَ ٻولي ڪمپيوٽر ايپليڪيشنس جي ٻولين جي فهرست ۾ گهڻ نظر اچي رهي آهي. ان ڪان سواءِ ڳولها جي نظام، ڇاڻ جي حصولءِ ترجمي واري نظام ۾ به پٺتي رهي آهي. پر هن وقت سنڌي ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس تي بهتر طريقي سان ڪم ٿي رهيو آهن، جنهن سلسلي ۾ ڪيترائي ڪمپيوٽيشنل لسانياتءِ نيچرل لينگئيجز پروسس جا اوزار آن لائين ڪيا ويا آهن، جيڪي سنڌي ٻوليءَ جي ترقيءَ ۾ اهم ڪردار ادا ڪري رهيا آهن. هي اوزار سنڌي ٻوليءَ ۾ لکجندڙ جملن کي سمجهن ٿا، انهن کي جدا ۽ آزاد لفظن ۽ نشانين ۾ ظاهر ڪن ٿا، لفظن کي عالمي گرامر ۽ سنڌي گرامر مطابق نشان لڳائين ٿا، لفظن کي ٽوڙي، انهن جون اڳاڙيون ۽ پڇاڙيون هٽائي انهن جا بنيادي لفظ ڪڍن ٿا جيڪي ڳولها جي نظام ۽ ڇاڻ جي حصول لاءِ اهم آهن، لفظن ۽ جملن مان ليما ڪڍن ٿا. ليما لفظ کي مڪمل ڪري پيش ڪري ٿو ۽ سنڌي متن يا جملن مان احساس ڳولهڻ ۾ مدد ڪري ٿو ته جيئن انهن جو بهتر تجزيو ڪري معلوم ڪري سگهجي ته اهي جملا مثبت آهن يا منفي يا نيوٽرل.ان ڪان سواءِ ليما ترجمي ڪرڻ واري اوزار يا نظام ٺاهڻ لاءِ پڻ اهم ۽ ضروري آهن. هن تحقيق جي محقق سنڌي ٻوليءَ لاءِ لفظن جو ڇارپڻ ٺاهيو آهي، جيڪاحقيقت ۾ هڪ وڏي ڪاميابي آهي. اهڙي قسم جي تحقيق ۽ ڊولپمينٽ سان سنڌي ٻوليءَ دنيا جي تحقيق جو موضوع بڻجي پئي. ان تحقيق ۽ ڊولپمينٽ سان دنيا ۾ سنڌي ٻولي جو هڪ مقام ٺهيو آهي، جنهنڪري سنڌي ٻولي عالمي سطح تي ٺهندڙ ٻولين جي گروه ۾ شامل ٿي چڪي آهي.

نوٽ: مقالي ۾ ڪم ايندڙ مخصوص انگريزي لفظن کي سنڌي صورتخطيءَ ۾ لکيو ويو آهي، انهن جي انگريزي هِجي (Spelling) حوالن ڪان پوءِ ڏني وئي آهي.

لفظن کي گرامر مطابق نشان لڳائڻ

هيءَ تحقيق جملن ۾ موجود لفظن کي يونيورسل يا عالمي ڳالهائڻ جي لفظن مطابق نشان لڳائي ٿي ته جيئن لفظن کي صحيح طريقي سان سڃاڻي سگهجي. دنيا جي مختلف ٽري بينڪس (6) جو مطالعو ۽ تجزيو دنيا جي مختلف ٻولين لاءِ عالمي يا يونيورسل نشانن جي اهميت کي ظاهر ڪن ٿا (7). هي عالمي ڳالهائڻ جا لفظ يا نشان دنيا جي مختلف ٻولين کي يونيورسل ڊپينڊينسي تحت ترڪيب ڏين ٿا، جنهن سان دنيا جون مختلف ٻوليون هڪ ٽري بينڪ تحت ڳنڍجن ٿيون يا هڪ قسم جو گرامر استعمال ڪن ٿيون. عالمي ڳالهائڻ جي لفظن جي اهميت کي محسوس ڪندي هيءَ تحقيق به يونيورسل ڳالهائڻ جي لفظن کي استعمال ڪري ٿي ته جيئن سنڌي ٻوليءَ کي عالمي گرامر مطابق ترڪيب ۽ تصريف ڏئي سگهجي ۽ اها ڪمپيوٽر کي آساني سان سمجهه ۾

اچي سگهي. ان ڪان سواءِ ڪمپيوٽر ٽيڪنالاجيءَ جي مختلف ڪمن لاءِ استعمال ٿي سگهي. مثال طور ترجمي لاءِ مختلف قسم جي ٻُولها لاءِ، ڄاڻ جي ٻُولها لاءِ، احساسن جي ٻُولها لاءِ ۽ وغيره وغيره . ٽيبل 1 سنڌي ڳالهائڻ جي لفظن ۽ عالمي ڳالهائڻ جي لفظن کي ظاهر ڪري ٿي . ان ٽيبل ذريعي خبر پوي ٿي ته سنڌي ۽ عالمي ڳالهائڻ جي لفظن ۾ ڪيترو فرق آهي، اهو فرق سنڌي گرامر جي اهميت کي اجاگر ڪري ٿو. سنڌي ۽ عالمي ڳالهائڻ جي لفظن ۾ ٿورو فرق آهي، جيڪو عالمي ڳالهائڻ جي لفظ پارٽ جي ڪري ظاهر ٿئي ٿو. لفظ پارٽ بنيادي طور پار ٽيسيپل آهي، جيڪو عالمي گرامر ۾ ناڪاري لفظن ۽ حرف اضافت لاءِ استعمل ٿئي ٿو. جڏهن ته سنڌيءَ ۾ ناڪاري لفظن لاءِ ظرف استعمال ٿئي.اهڙيءَ طرح، حرف اضافت وارن لفظن کي ترڪيب اضافي ذريعي جملن ۾ ظاهر ڪجي ٿو. مرزا قليچ بيگ پنهنجي ڪتاب سنڌي ويا ڪرڻ ڇاپي چوٿين جي صفحي 89 ۾ لکي ٿو ته: ”اضافت يا ملڪيت ۽ واسطيداري به گهڻن قسمن جي ٿئي ٿي ۽ انهيءَ ۾ هميشه حرف اضافت ڪم اچي ٿي“ (5). ترڪيب اضافت جي ڪري پنهنجي ملڪيت يا واسطيداريءَ جي خبر پوي ٿي. مثال طور، منهنجو ڪتاب، اسلم جو قلم وغيره وغيره. سنڌيءَ ۾ ناڪاري ۽ اضافي حالتن لاءِ عليحده ڳالهائڻ جا لفظ استعمال ڪيا وڃن ٿا پر يونيورسل گرامر ۾ هڪ ئي نشان استعمال ڪيو وڃي ٿو. ان هڪ نشان ڪان سواءِ سنڌي ڳالهائڻ جي لفظن ۽ يونيورسل گرامر جي لفظن ۾ ڪو خاص فرق ڪونه آهي. ان ننڍي فرق جي ڪري سنڌي ٻوليءَ جي حيثيت اهم ٿي وڃي ٿي. ڇاڪاڻ ته يونيورسل گرامر جا ٻيا ڳالهائڻ جا لفظ سنڌي ٻوليءَ جي ڳالهائڻ جي لفظن سان ملن ٿا. يونيورسل گرامر ۾ هڪ لفظ ايڪس استعمال ڪيو ويو آهي، جيڪو اهڙن لفظن کي نشان لڳائي ٿو جيڪي نامعلوم آهن يا انهن لفظن جي ويا ڪرڻ مطابق ڪا حيثيت نه آهي. جيئن ته يونيورسل گرامر هڪ عالمي گرامر آهي، جيڪو دنيا جي سمورين ٻولين لاءِ ٺاهيو ويو آهي تنهنڪري ايڪس هڪ نشان رکيو ويو آهي ته، جيئن ڪنهن اڻ سڃاتل لفظ کي به نشان ڏئي سگهجي. ان سان ڪمپيوٽر يا ڪنهن به قسم جي ڪمپيوٽرائزڊ مشين يا روبوٽس کي دنيا جي ڪنهن به ٻوليءَ کي سمجهڻ ۾ آساني ٿئي ٿي.

ٽيبل 1: يونيورسل ۽ سنڌي ڳالهائڻ جي لفظن جو ٽيگ سيٽس

سنڌي ڳالهائڻ جي لفظن جو ٽيگ سيٽ	ڳالهائڻ جي لفظن جو مڪمل نالو	يونيورسل ڳالهائڻ جي لفظن جو ٽيگ سيٽ
صفت	Adjective	ADJ
حرف جر	Adposition	ADP
ظرف	Adverb	ADV
فعل معاون	Auxiliary Verb	AUX
حرف جملو	Conjunction	CONJ

DET	Determiner	ضمير اشارو
INTJ	Interjection	حرف ندا
NOUN	Noun	اسم
NUM	Number	صفت عددي
PART	Participle	ظرف (ناڪاري لفظ) ۽ حرف اضافت
PRON	Pronoun	ضمير
PROPN	Proper Noun	اسم خاص
PUNCT	Punctuation	بيهڪ جون نشانيون
SCONJ	Subordinating conjunction	حرف جملو شرطيه
SYM	Symbol	نشانيون
VERB	Verb	فعل
X	Other	ٻيا يا نامعلوم

مسئلي جو بيان

ڪمپيوٽر ٽيڪنالاجي دنيا کي بدلائي رهي آهي. هن وقت دنيا جي ظاهري يا جسماني بناوت ورچولائيزيشن طرف وڃي رهي آهي، جنهن ذريعي اڳتي هلي شايد دنيا جي موجوده بناوت بہ بدلجي سگهجي ٿي. ان ترقيءَ ۾ دنيا جي ٻولين کي ڪمپيوٽرائزڊ ڪرڻ وارو قدم بہ هڪ اهم اڳڀرائي آهي. ان سلسلي ۾ يونيورسل گرامر ٺاهيو ويو آهي، جيڪو پوريءَ دنيا جي ٻولين کي هڪ گرامر تحت ترڪيب ڏئي ٿو تہ، جيئن پوريءَ دنيا جا ماڻهو هڪ ٻئي جي ٻوليءَ کي آسانيءَ سان سمجهي سگهن. جيئن تہ ڪمپيوٽر هڪ مشين آهي تنهنڪري ڪمپيوٽر کي قدرتي يا انساني ٻولين سمجهائڻ خاطر ڪمپيوٽرسائنس مان نيچرل لينگئيجز پروسس ۽ لسانيات مان وري ڪمپيوٽيشنل لسانيات جا شعبا پيدا ڪيا ويا آهن. انگريزي ، عربي ۽ دنيا جي ڪجھ ٻين ٻولين ان نظام مان فائدو وٺندي پنهنجي ٻولين کي ڪمپيوٽر سان روشناس ڪرائڻ لاءِ جوڳا قدم کنيا آهن. سنڌي ٻولي ان سلسلي ۾ پٺ تي پيل ٻولي هئي، تنهنڪري دنيا جي مختلف ٻولين لاءِ ٺاهيل ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس جي الگوريٿم کي ٿورو تبديل ڪري انهن کي سنڌي ٻوليءَ لاءِ استعمال ڪرڻ تمام گهڻو اهم هئو. ان مسئلي کي نظر ۾ رکي ڪري تحقيق جو سلسلو شروع ڪيو تہ جيئن سنڌي ٻوليءَ جا ڪمپيوٽيشنل لسانياتي مسئلا حل ڪري سگهجن ۽ هن ٻوليءَ کي بين الاقوامي ترقي يافتہ ٻولين جي سٽ ۾ شامل ڪري سگهجي. هي تحقيقي مقالو ان قسم جي تحقيق جي هڪ ڪڙي آهي. هن تحقيق سان سنڌي ٻوليءَ جا ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس جي حوالي سان پيش ايندڙ مسئلا حل ڪري سگهبا. سنڌي ٻولي صرفي جي حوالي سان شاهوڪار ٻولي آهي، تنهنڪري ڪمپيوٽيشنل لسانيات مطابق

سنڌي ٻوليءَ جي متن کي تصريف ۽ ترڪيب ڏيڻ، جملي کي جدا جدا لفظن ۽ بيهڪ جي نشانين ۾ ظاهر ڪرڻ ڏکيو ڪم آهي ڇاڪاڻ تہ سنڌي ٻولي پنهنجي متن ۾ زيرن، زبرن ۽ پيش وار لفظ، مرتب، مرڪب ۽ ٻٽا لفظ وڏيڪ استعمال ڪري ٿي. زير، زبر ۽ پيش جملن ۾ لفظن جي معنا مٽائي ڇڏين ٿا. اهڙيءَ طرح ساڳي اچار وارا لفظ پڻ سنڌي ڪمپيوٽيشنل لسانياتي عمل لاءِ مسئلا پيدا ڪن ٿا. ڪمپيوٽر لاءِ اهو ڏکيو ٿيندو تہ هو ساڳين اچارن وارن لفظن، زير، زبر ۽ پيش وارن لفظن، مرڪب ۽ ٻٽن لفظن کي آسانيءَ سان سمجهي. تنهنڪري هيءَ تحقيق سنڌي ٻوليءَ جي لسانياتي مسئلن کي ڪمپيوٽيشنل لسانيات ذريعي حل ڪري ٿي. هن تحقيق ذريعي سنڌي متن جي لفظن کي جدا جدا ڪيو ويندو، پوءِ لفظن کي گرامر مطابق نشان ڏنا ويندا. متن کي تصريف ۽ ترڪيب ڏني ويندي ۽ ان جو تجزيو ڪيو ويندو تہ جيئن ڪمپيوٽر سنڌي ٻوليءَ جي متن ۽ لفظن کي عالمي گرامر مطابق سمجهي سگهي.

اڳ ۾ ٿيل ڪم جو مطالعو

ٻولي رابطي جو بهترين ذريعو آهي، جنهن ذريعي اسان پنهنجا خيال، خواب، سوچون، خوشيون، مسئلا وغيره هڪ ٻئي سان ونڊيون ٿا. اهڙيءَ طرح رابطو لکت ۾ بہ ٿئي ٿو تہ ڳالهائڻ ذريعي بہ تہ اشارن ذريعي بہ. ڪمپيوٽرٽيڪنالاجي بہ مختلف قسمن جون ٻوليون ۽ پروگرام استعمال ڪري انساني يا قدرتي ٻولين جا مسئلا حل ڪري ٿي. ان سلسلي ۾ ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس تي تحقيق جاري آهي تہ جيئن دنيا جي مختلف ٻولين جا لسانياتي ۽ ڪمپيوٽيشنل مسئلا حل ڪري سگهجن. هن وقت سڀ کان وڌيڪ ڪمپيوٽرائزڊ اثاثا انگريزي ۽ عربي ٻولين لاءِ موجود آهن، جيڪي سنڌي ٻوليءَ لاءِ گهٽا ڪارائتا نہ آهن ڇاڪاڻ تہ سنڌي ٻوليءَ جي بناوٽ انهن ٻولين کان مختلف آهي .هن تحقيق جو مقصد سنڌي ٻوليءَ تي ڪمپيوٽيشنل لسانيات جي حوالي سان ٿيندڙ ڪم جو جائزو وٺڻ ۽ ان کي وڌيڪ تحقيق ڪرڻ لاءِ پيش ڪرڻ آهي تہ، جيئن سنڌي ٻوليءَ کي بين الاقوامي سطح تي تحقيق جي موضوع بڻائڻ سان گڏ يونيورسل گرامر مطابق ترڪيب ڏيڻ ۽ ان جي تجزيي ڪرڻ لاءِ پيش ڪري سگهجي.

سنڌي ٻوليءَ تي ڪمپيوٽيشنل لسانيات جي حوالي سان تمام گهٽ ڪم ٿيو آهي. ان سلسلي ۾موتلاٽي رويش پنهنجي تحقيقي مقالي ۾ لکي ٿو تہ: ”سنڌي ٻولي ڪمپيوٽر ٽيڪنالاجيءَ جي حوالي سان هڪ غريب ٻولي آهي، جنهن تي ڪمپيوٽيشنل ٽيڪنالاجيءَ جي حوالي سان گهٽ ڪم ٿيو آهي“ (8)، جاويد مهر ۽ ان جي ساٿين تحقيق ڪندڙن پنهنجيءَ تحقيق ۾ ظاهر ڪيو آهي تہ: سنڌي جملن ۾ ڳنڍيل يا ٻٽن لفظن کي ٽوڙڻ هڪ ڏکيو عمل آهي ڇاڪاڻ تہ انهن کي ٽوڙڻ سان انهن جي معنا ۽ مقصد ضايع ٿي

وجن ٿا" (9). تنهن هوندي بہ جاويد مهر ۽ گڏ تحقيق ڪندڙ ساٿين سنڌي جملن کي ٽوڙڻ لاءِ هڪ ماڊل پيش ڪيو آهي، جيڪو مختلف ٿهن ذريعي سنڌي جملن کي ٽوڙي جدا ۽ آزاد لفظن ۾ پيش ڪري ٿو. پر ضرورت هئي تي سنڌي ٻوليءَ تي وڏي پيماني تي ڪم ڪيو وڃي تہ جيئن سنڌ ٻوليءَ جا ڪمپيوٽيشنل لسانياتي مسئلا حل ڪري سگهجن. ان سلسلي ۾ مون ۽ عاصم امداد وگڻ پنهنجيءَ تحقيق ذريعي ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس جي گهرجن مطابق خودڪار ڪمپيوٽرائزڊ نظام ٺاهيو، جيڪو سنڌي متن کي جدا جدا لفظن ۾ پيش ڪرڻ سان گڏ انهن لفظن کي يونيورسل ڳالهائڻ جي لفظن مطابق نشان لڳائي ٿو. ان سان گڏ سنڌي متن کي فقرن ۾ پڻ ورهائي ٿو (10) جيڪو هن وقت تائين ٿيندڙ تحقيقي ڪم ۾ مٿانهين حيثيت رکي ٿو. مون ۽ عاصم امداد وگڻ پنهنجيءَ تحقيق ذريعي سنڌي ٻوليءَ جو لسانيات مطابق ڊيٽا سيٽ ٺاهيو، جنهن ۾ سنڌي لفظن جا گرامر مطابق نشان، صرفيہ صورتن مطابق حيثيت، فعل مطابق حيثيت، لفظن مان نڪرندڙ احساس وغيره شامل آهن (11). ان ڊيٽا سيٽ جي ڪري سنڌي ڪارپس تي قومي ۽ بين الاقوامي سطح تي تحقيق ڪرڻ جا دروازا کلن ٿا. هيءَ تحقيق هڪ خاص ماڊل ذريعي سنڌي ٻوليءَ جا لسانياتي مسئلا ڪمپيوٽيشنل لسانيات جي اوزارن ذريعي حل ڪري ٿي. اهي اوزار خاص الگوريٿم ذريعي ٺاهيا ويا آهن، تنهنڪري هي تحقيق مستقبل جي تحقيق لاءِ پڻ دروازا کولي ٿي تہ جيئن هنن اوزارن ۽ طريقن کي استعمال ڪندي وڌيڪ بهتر نتيجا حاصل ڪري سگهجن.

سنڌي ليماٽائزيشن تي تحقيقي ڪم نہ هجڻ جي برابر ٿيو آهي. ان سلسلي ۾ مون (مظهر علي ڏوتيي) ۽ عاصم امداد وگڻ پهريون ڀيرو هن اهم موضوع تي ڪم ڪري سنڌي ٻوليءَ کي مختلف قسمن جي ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس جي مخلتلف عملن لاءِ پيش ڪيو تہ جيئن لفظن جي ڳولها، ترجمي ڪرڻ جي نظام، ڄاڻ جي حصول وغيره لاءِ انهن لفظن کي يعني ليما کي استعمال ڪري سگهجي. هن تحقيق ۾سنڌي ٻوليءَ لاءِ اسٽيمنگ ۽ ليماٽائزيشن پروسس تي ڪم ڪيو ويو آهي، جيڪو ٻولين تي ٿيندڙ ٽين بين الاقوامي ڪانفرنس ۾، جيڪا 2017ع ۾ ڪراچي ۾ ٿي هئي، پيش ڪيو ويو هو (12). ان تحقيق ۾ سنڌي ليما ۽ اسٽيمنگ تي بحث ڪيو ويو آهي. موجوده دور ۾ نيچرل لينگئيجز پروسس جي شعبي ۾ سڀ کان وڌيڪ ڪم احساسن کي ڳولهڻ، انهن کي ماپڻ ۽ انهن جي تجزيي تي ٿي رهيو آهي. ان سلسلي ۾ مشين کي سکيا ڏئي ماهر ڪيو وڃي ٿو تہ جيئن اها ماڻهن جا احساس، رايا ۽ سوچون سمجهي انهن جو تجزيو ڪري سگهي. سنڌي ٻوليءَ ۾ ان موضوع تي ڪڏهن بہ ڪم نہ ٿيو هو. پر هن مقالي جي محقق ان ڪم کي اهم سمجهندي پهريون دفعو سنڌي ٻوليءَ ۾موجود احساسن ۽ راين تي ڪم ڪيو تہ جيئن ڪمپيوٽر آسانيءَ سان سنڌي ٻوليءَ ۾ موجود

احساسن ۽ راين کي سمجهي سگهي ۽ انهن جو تجزيو ڪري سگهي (13). دنيا ۾ سياسي، سماجي ۽ واپاري ادارا احساسن کي ماپڻ وارن اوزارن ۽ طريقن کي اهميت ڏين ٿا ڇاڪاڻ ته ان ذريعي انهن کي خبر پوي ٿي ته ماڻهو انهن جي باري ۾ ڇا ٿا سوچين؟

ڪم جو طريقيڪار

ڪمپيوٽر سائنس ۾ منطق ۽ ميٿميٽيڪل ماڊل کي تمام گهڻي اهميت حاصل آهي. ڇاڪاڻ ته ڪمپيوٽر ڳالهيون نه پر منطق۽ ميٿميٽيڪل ماڊلس کي سمجهندو آهي.اهو ئي سبب آهي جو ڪمپيوٽر جي پنهنجي ٻولي صرف ٻن انگن تي ٻڌل آهي جيڪي آهن 0 ۽ 1. سنڌي متن کي ترڪيب ڏيڻ ، جملن ۾ لفظن ۽ نشانين کي الڳ ڪرڻ لاءِ ڪجھ مخصوص قسم جا پرگرام ٺاهيا ويا آهن. اهي پروگرام مختلف قسمن جي الگوريٿم ذريعي ٺاهيا ويا آهن، جيڪي ڪمپيوٽر کي هدايتون ڏين ٿا ته جيئن هُو سنڌي ٻوليءَ جي متن کي سمجهي ڪري ان جو حل پيش ڪري. هر ڪم لاءِ الڳ الگوريٿم ٺاهيو ويو آهي ته جيئن ڪمپيوٽر کي سمجهڻ ۾ آساني ٿئي. انهن الگوريٿمس جي ڪري سنڌي ٻوليءَ لاءِ ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس تحت مختلف قسم جا اوزار ٺاهيا ويا آهن، جيڪي هن وقت به انٽرنيٽ تي سنڌي اين ايل پي جي نالي سان موجود آهي. اهڙي قسم جا اوزار هن وقت تائين دنيا جي ٿورين ٻولين لاءِ انٽرنيٽ تي موجود آهن. تنهنڪري انهن آن لائين اوزارن جي ڪري دنيا ۾ سنڌي ٻوليءَ جي هڪ منفرد حيثيت ٺهي چڪي آهي. انهن اوزارن کي دنيا جي مختلف ملڪن ۾ استعمال ڪري سنڌي ڪارپس تي تحقيق ڪئي وڃي ٿي. هن تحقيقي مقالي جي نتيجي واري حصي ۾ پيش ٿيندڙ نتيجا به هنن اوزارن ذريعي حاصل ڪيا ويا آهن.

ان ڪان سواءِ مشين لرننگ ۽ ڊيپ لرننگ‌ذريعي ڪمپيوٽر کي ماهر ڪيو ويو آهي ته هو سنڌي متن کي سمجهي سگهي جيڪو پڻ هڪ وڏو ڪم آهي. هن وقت دنيا ۾ مشين لرننگ ۽ ڊيپ لرننگ ذريعي ٻولين کي سمجهيو وڃي ٿو، جيڪو هڪ ڪامياب طريقو آهي. ڪمپيوٽر مشين قدرتي ٻولين جي لفظن کي سڌي طريقي سان نه ٿي سمجهي بلڪ ڪجھ مخصوص ڪوڊ آهن، جن جي ذريعي ڪمپيوٽر مختلف ٻولين جي لفظن کي سمجهي ٿو. تنهنڪري سنڌي ڪارپس يا لفظن جي ڀنڊار کي مختلف ويڪٽرس۽ ميٽرائسزجي شڪل ۾ تبديل ڪيو ويو آهي ته جيئن اڪيلن ۽ ڳنڍيل لفظن کي ويڪٽرس ۾ تبديل ڪري سگهجي ۽ ڪمپيوٽر کي سمجهائي سگهجي. هيءُ هڪ مڪمل سائينٽيفڪ طريقو آهي، جنهنڪري ڪمپيوٽر کي سمجهڻ ۾ آساني ٿئي ٿي. جيتوڻيڪ ان ڪم ڪرڻ ۾ تمام گهڻيون ڏکيائيون پيش اچن ٿيون ڇاڪاڻ ته مشين لرننگ ۽ ڊيپ لرننگ لاءِ جيڪي عمل ضروري آهن، اهي سنڌي ٻوليءَ لاءِ موجود نه آهن. جڏهن ته اهي

عمل انگريزي، فرينچ، عربي ۽ ڪجھ ٻين ٻولين لاءِ موجود آهن. انهن مسئلن کي ڏسندي سنڌي ٻوليءَ لاءِ پنهنجي تحقيق ذريعي عليحده اهڙي قسم جا عمل ۽ طريقيڪار ٺاهيا ويا آهن تہ جيئن سنڌي ٻوليءَ کي مشين لرننگ ۽ ڊيپ لرننگ تحت پروسس ڪري نتيجا حاصل ڪري سگھجن. تنهنڪري اها پڻ هڪ وڏي سوڀ آهي جيڪا اڳتي سنڌي ٻوليءَ تي ڪم ڪندڙن لاءِ آساني پيدا ڪندي. تحقيق ذريعي مشين لرننگ ۽ ڊيپ لرننگ لاءِ سنڌي ٻوليءَ لاءِ مخصوص ماڊلس ۽ پروگرام ٺاهيا ويا آهن، جيڪي سنڌي ڪارپس يا لفظن جي ڀنڊار کي مختلف قسمن جي عملن ۾ مدد ڪن ٿا.

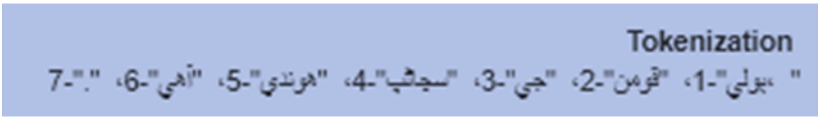
نتيجا ۽ بحث

الگوريٿمس ۽ مشين لرننگ جي بنياد تي سنڌي ٻوليءَ لاءِ ٺاهيل اوزار سنڌي متن کي بهتر طريقي سان پروسس ڪن ٿا ۽ سٺا نتيجا ڏين ٿا. اهڙيءَ طرح سنڌي ٻوليءَ جا ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس سان واسطو رکندڙ مسئلا حل ٿين ٿا. نتيجن جو تفصيل هيٺ ڏجي ٿو.

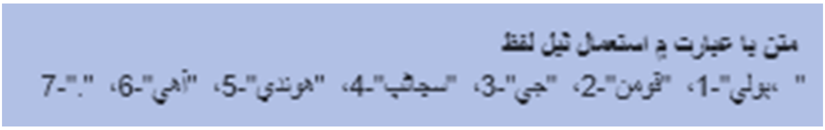
سنڌي ٽيڪسٽ پارسر

ٽيڪسٽ پارسرهڪ ڪمپيوٽرائيزڊ پروگرام آهي، جيڪو ڪنهن بہ ٻوليءَ جي متن کي ٽڪرن ۾ ورهائي ان کي گرامر مطابق هڪ مڪمل لفظ طور ظاهر ڪري ٿو، جنهن سان لفظ يا ڪنهن بہ بيهڪ جي نشانيءَ جي سڃاڻپ آسانيءَ سان ٿي سگھي ٿي. ماڻهو پنهنجيءَ ٻوليءَ جي لفظن کي آسانيءَ سان سمجھي سگھي ٿو پر ڪمپيوٽر کي ان لفظ کي سمجھڻ ڏکيو هوندو آهي. سنڌي متن جي پارسنگ ڪرڻ لاءِ ڪمپيوٽر کي هدايتون ڏئي ماهر بڻايو ويو آهي. انهن هدايتن موجب، سنڌي ٽيڪسٽ پارسر سڀ کان پهريان سنڌي متن کي ساڄي پاسي ڪري ان جو اسٽائيل بدلائي ٿو، ڇاڪاڻ تہ اسان وٽ سنڌي ساڄي پاسي کان لکي وڃي ٿي. ان کان پوءِ متن کي عليحده لفظن ۽ نشانين ۾ ظاهر ڪري ٿو. بعد ۾ انهن کي عالمي ۽ سنڌي گرامر مطابق نشان لڳائي ٿو. شڪل1 ۾ اهڙي عمل جو نتيجو ظاهر ڪجي ٿو. بعد ۾ انهن لفظن ۽ نشانين کي سڃاڻي انهن کي عالمي ۽ سنڌي گرامر مطابق نشان لڳائي ٿو. مثال طور اسم، ضمير يا نائون ، پرونائون وغيره. شڪل 2 ۽ 3 ۾ خودڪار نظام ذريعي سنڌي متن کي عالمي ۽ سنڌي گرامر مطابق نشان لڳايا ويا آهن. شڪل 4 ۽ 5 ۾ سنڌي ڳالهائڻ جي لفظ ظرف ۽ يونيورسل ڳالهائڻ جي لفظ پارٽ جي فرق کي ظاهر ڪيو ويو آهي. جڏهن تہ شڪل 6 ۾ هڪ ٻيو سنڌي جملو پروسس ڪيو ويو آهي، جنهن ۾ مختلف قسمن جا يونيورسل ڳالهائڻ جا لفظ استعمال ڪيا ويا آهن. شڪل 7 ۾لفظن کي فقرن ذريعي ظاهر ڪيو ويو آهي، جنهن

۾ سنڌي متن ڪي سنٽيڪٽڪ پارسنگ واري عمل سان جدا جدا لفظن ۽ نشانين ۾ ورهائي انهن ڪي فقرن ۽ نشانن ذريعي ظاهر ڪيو ويو آهي.شڪل 8 ۾ سنڌي ٽيڪسٽ پارسر لفظن جون صرفيي واريون صورتون ٻڌائي ٿو تہ جملي ۾ ڪيترا لفظ صرفيي جي ڪهڙين صورتن ۾ ڪيتري تعداد ۾ استعمال ٿيل آهن.

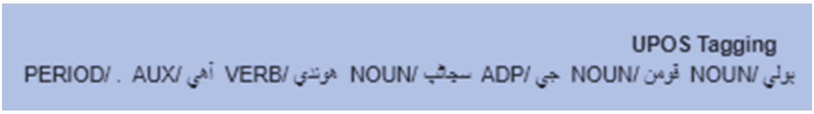


شڪل 1: جملو ٽوڙي جدا جدا لفظن ۾ ظاهر ڪرڻ وارو عمل

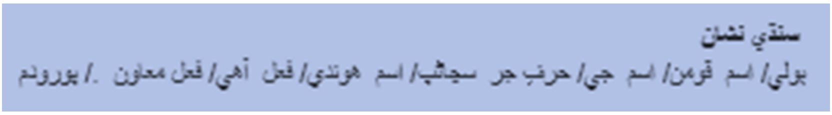


شڪل 2: جملو ٽوڙي جدا جدا لفظن ۾ ظاهر ڪرڻ وارو عمل

شڪل 1 ۽ 2 ۾ سنڌي متن ڪي ٽوڙي ڪري انهن ڪي جدا جدا لفظن ۽ نشانين ۾ ظاهر ڪرڻ سان گڏ ترتيبوار نمبر پڻ لڳايا ويا آهن تہ جيئن لفظن جي جڳهہ معلوم ٿي سگهي. ان عمل سان لفظ جي بناوت جي خبر پوي ٿي، جيڪا ڳاٽيٽي ۾ پڻ ڪم اچي ٿي. ڪمپيوٽر ڪنهن بہ متن ۾ موجود لفظن ۽ نشانين جي تعداد ڪي آسانيءَ سان ٻڌائي سگهي ٿو. اهڙيءَ طرح ڪمپيوٽر ڪي لفظن ڪي گرامر مطابق نشان لڳائڻ ۾ پڻ آساني ٿئي ٿي. ڳولها جي نظام ۾ پڻ هيءُ طريقو مددگار ٿئي ٿو، جنهن ذريعي گوگل يا ٻين ڳولها جي نظامن ۾ ڳولها ڪرڻ آسان ٿئي ٿي. شڪل 1 ۽ 2 سنڌي جملي ڪي صحيح طريقي سان ٽوڙڻ واري نتيجي ڪي ظاهر ڪري الگوريٿم ڪي صحيح ثابت ڪن ٿيون.



شڪل 3: جملي ۾ موجود لفظن ۽ نشانين ڪي عالمي ڳالهائڻ جي لفظن مطابق **نشان ڏيڻ وارو** عمل

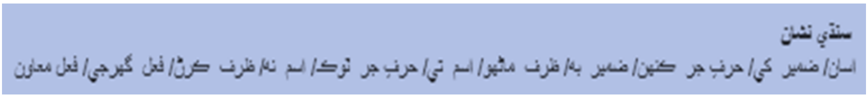


شڪل 4: جملي ۾ موجود لفظن ۽ نشانين ڪي سنڌي ڳالهائڻ جي لفظن مطابق نشان ڏيڻ وارو عمل

شڪل 3 ۾ خودڪار نظام ذريعي، جيڪو هن تحقيق ذريعي ٺاهيو ويو آهي، سنڌي لفظن ڪي عالمي گرامر جي ڳالهائڻ جي لفظن مطابق نشان لڳايا ويا آهن. شڪل 4 ۾ ساڳي خودڪار نظام ذريعي سنڌي لفظن ڪي سنڌي گرامر جي ڳالهائڻ جي لفظن

مطابق نشان لڳايا ويا آهن. لفظن ڪي خودڪار نظام ذريعي صحيح نشان لڳائڻ حقيقت ۾ الگوريٿم جي بنياد تي ناهيل اوزارن جي ڪم ڪي صحيح ثابت ڪن ٿا. سنڌي لفظن ڪي نشان لڳائڻ سان لفظن جي سڃاڻپ آسان ٿئي ٿي، جنهنڪري ڪمپيوٽر سنڌي لفظن ڪي سڃاڻي انهن ڪي مختلف قسمن جا جملا ٺاهڻ ۽ ترجمي ڪرڻ لاءِ استعمال ڪري سگهي ٿو. ان ڪانسواءِ ڪمپيوٽر انهن لفظن ڪي احساسن ڪي ڳولهڻ، ڇاڻ ڪي حاصل ڪرڻ ۽ مختلف قسمن جي ڳولها جي نظام لاءِ پڻ استعمال ڪري سگهي ٿو.

شڪل 5 ۽ 6 سنڌي گرامر ۽ عالمي گرامر ۾ فرق ڪي ظاهر ڪن ٿيون. انهن شڪلين ذريعي عالمي ڳالهائڻ جي لفظ پارٽ ۽ سنڌي ڳالهائڻ جي لفظ ظرف ۾ فرق ڏيکاريو ويو آهي. سنڌي لفظ نہ لاءِ عالمي گرامر جو لفظ پارٽ استعمال ڪري ٿو ، جڏهن تہ سنڌي گرامر ڳالهائڻ جو لفظ ظرف استعمال ڪري ٿو. ڪمپيوٽر انهن لفظن ڪي ٻنهي گرامرن مطابق سمجهي نشان لڳائي ٿو ۽ صحي نتيجا ڏئي ٿو جنهن سان پيش ڪيل الگوريٿم جي اهميت ظاهر ٿئي ٿي.



شڪل 5: جملي ۾ موجود لفظن ۽ نشانين ڪي سنڌي ڳالهائڻ جي لفظن مطابق نشان ڏيڻ وارو عمل

شڪل 6: جملي ۾ موجود لفظن ۽ نشانين ڪي عالمي ڳالهائڻ جي لفظن مطابق نشان ڏيڻ وارو عمل

شڪل 7 ۾ سنڌي متن ڪي مختلف قسمن جي عالمي گرامر جي لفظن جا نشان لڳايا ويا آهن، جيڪي سنڌي ٻوليءَ لاءِ ناهيل الگوريٿم جي بنياد تي ناهيل اوزار جي اهميت ڪي ظاهر ڪن ٿا تہ هيءُ اوزار هر قسم جي لفظن ۽ نشانين ڪي عالمي گرامر مطابق نشان ڏئي سگهي ٿو. اهو عمل ثابت ڪري ٿو تہ: سنڌي ٻولي بين الاقوامي گرامر مطابق آسانيءَ سان ترڪيب ڏئي سگهجي ٿي جيڪو ڪمپيوٽر ڪي سنڌي سمجهڻ لاءِ ۽ سڪيا ڏيڻ لاءِ پڻ اهم آهي. پيش ٿيل سنڌي متن ۾ ڪيترائي مرتب، مرڪب ۽ ٻٽا لفظ موجود آهن، جن ڪي سنڌي ڪمپيوٽيشنل لسانيات جي خودڪار اوزار وسيلي بهتر طريقي سان نشان لڳايا ويا آهن. پيش ٿيل متن ۾ سڀ ڪان وڌيڪ مفرد، مرتب لفظ ۽ مرڪب لفظ استعمال ٿيا آهن. جڏهن تہ ٻٽا يا ڍهرايل لفظ گهٽ استعمال ٿيا آهن. مرتب ۽ مرڪب لفظن جو وڌيڪ استعمال سنڌي ٻوليءَ ڪي شاهوڪار ٻولي بڻائي ٿو. متن ۾ بيهڪ جون نشانيون پڻ استعمال ڪيون ويون آهن جن ڪي سنڌي ڪمپيوٽيشنل لسانيات واري خودڪار نظام ذريعي نشان لڳايا ويا آهن.

اجڪلهه ADV/ غريب ADJ/ ماڻهو NOUN/ ننڍڙڪو NOUN/ ٿي VERB/ پيو VERB/ آهي AUX/ ۔ PUNC/
ڪيس PRON/ ڪوئي PRON/ بچڻ VERB/ وارو ADP/ ئي ADP/ ڪونهي ADV/ ۔ PUNC/
بلڪل ADV/ ڏک NOUN/ پيو VERB/ ڏسي ADJ/ . PERIOD/ دولتمند ADJ/ ماڻهو NOUN/ ڪيس PRON/
ڏڪوداب NOUN/ ڪري VERB/ پنهنجا PRON/ ڪم NOUN/ ونس VERB/ ٿا VERB/ ۔ PUNC/
جيڪڏهن CONJ/ گوڙگڙبڙ NOUN/ ڪري VERB/ ته ADP/ ڪيس PRON/ ڏڪادهمان NOUN/ ۽ CONJ/
ڌمڪيون NOUN/ ملنس PRON/ ٿيون VERB/ . PERIOD/ / اسانجي PRON/ معاشري NOUN/ ۾ ADP/
غريب ADJ/ ويچارا ADJ/ ڏاڍا ADJ/ پريشان ADJ/ ۽ CONJ/ ڏکايل ADJ/ رهن VERB/ ٿا VERB/ . PERIOD/

شڪل 7: جملي ۾ موجود لفظن ۽ نشانين ڪي عالمي ڳالهائڻ جي لفظن مطابق نشان ڏيڻ وارو عمل

ڪمپيوٽيشنل لسانيات ۾ سنٽيڪٽڪ پارسنگ جي تمام گهڻي اهميت آهي . سنٽيڪٽڪ پارسنگ ذريعي جملن کي ٺاهيو وڃي ٿو. ان ذريعي ترجمي جو بهتر نظام ٺهي ٿو، سوالن جوابن جو ڪمپيوٽرائزڊ خودڪار نظام پڻ ٺهي ٿو، ان جي ڪري ڪمپيوٽر سسٽم قدرتي ٻولين کي بهتر طريقي سان سمجهي ٿو. اسان پنهنجيءَ تحقيق ذريعي سنڌي ٻوليءَ لاءِ سنٽيڪٽڪ پارسنگ جو الگوريٿم ۽ اوزار ٺاهيو آهي، جيڪو سنڌي ٻوليءَ جي متن کي سنٽيڪٽيڪلي پارس ڪري بهتر نتيجا ڏئي ٿو. ان سان سنڌيءَ کي ٻين ٻولين ۾ ترجمي ڪرڻ جي نظام ٺاهڻ سان گڏ وڳڏ ٻيا ڪيترائي ڪمپيوٽرائزڊ پروگرام ۽ اوزار ٺاهي سگهجن ٿا، جن سان سنڌي ٻولي ڪمپيوٽر جو بهتر اوزار بڻجي سگهي ٿي. سنٽيڪٽڪ پارسنگ جا نتيجا مختلف طريقن سان پيش ٿين ٿا جن مان هڪ طريقو شڪل 8 ۽ 9 ۾ پيش ڪيو وڃي ٿو. هن نتيجي ۾ لفظن کي فقرن ۽ گرامر جي نشانن ذريعي ظاهر ڪيو وڃي ٿو،جنهن سان لفظ جي گراميٽيڪل شڪل سامهون اچي ٿي. شڪل8 عالمي گرامر مطابق سنڌي متن جي سنٽيڪٽڪ پارسنگ ظاهري ڪري ٿي ۽ شڪل9 سنڌي گرامر مطابق سنڌي متن جي سنٽيڪٽڪ پارسنگ کي ظاهر ڪري ٿي.

تصريف ۽ ترڪيب Parsing

S)
(((ٻولي) NOUN) NP)
(((قومن) NOUN) NP)
(((جي) ADP) PP)
(((سڃاڻپ) NOUN) NP)
(((هوندي) VERB) VP)
(((آهي) AUX) VP)
(((.) PERIOD) PUNC)

شڪل 8: جملي۾ موجود لفظن ۽ نشانين کي عالمي گرامر کي استعمال ڪندي سنٽيڪٽڪ پارسنگ مطابق تصريف ۽ ترڪيب ڏيڻ وارو عمل

```
تصريف ۽ ترڪيب ;

S)
(((بولي ) اسم ) NP)
(((قومن ) اسم ) NP)
(((جي ) حرف جر ) PP)
(((سجاڳ ) اسم ) NP)
(((هوندي ) فعل ) VP)
(((آهي ) فعل معاون ) VP)
(((پوروم ( . ) PUNC)
```

شڪل 9: جملي ۾ موجود لفظن ۽ نشانين کي سنڌي گرامر کي استعمال ڪندي سنٽيڪٽڪ پارسنگ مطابق تصريف ۽ ترڪيب ڏيڻ وارو عمل

شڪل 10 ۽ 11 ذريعي سنڌي جملي **بولي قومن جي سجاڳ هوندي آهي** جو صرفيه جي صورتن ذريعي اسٽيسٽيڪل يا انگن اکرن ذريعي تجزيو پيش ڪن ٿيون. هيءُ اوزار جملن کي سمجهي انهن مان لفظن کي صرفي صورتن مطابق جدا جدا ڪري ٿو ۽ آخر ۾ هر صورت مطابق نتيجو ڏيکاري ٿو. پيش ڪيل جملو ٿورن لفظن تي ٻڌل آهي تنهنڪري صرفيي جي سمورين صورتن موجب نتيجا نہ پيا ظاهر ڪيا وڃن. هن جملي ۾ چار مفرد ۽ ٻہ مرتب صورتن مطابق لفظ موجود آهن، جن کي سنڌي ڪمپيوٽيشنل لسانيات جو اوزار سمجهي ظاهر ڪري ٿو. هن اوزار ذريعي خبر پئي ٿي تہ سنڌي ٻوليءَ ۾ ڪيترين صرفيي جون ڪيترين صورتون موجود آهن. ان تجزيي سان ٻوليءَ جي شاهوڪاري ۽ اهميت جي خبر پوي ٿي.

Analysis of Sindhi Morphological Tokens

Morphological Words	Total number in text	Percentage
Simple Words	4	57.14
Complex Words	2	28.57
Compound Words	0	0.00
Reduplicated Words	0	0.00

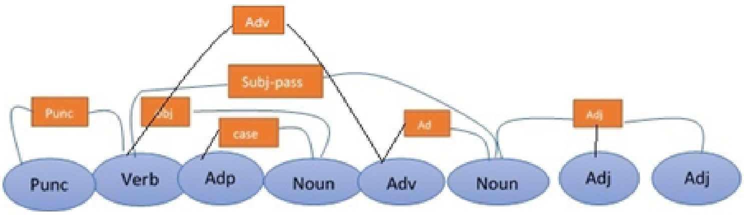
شڪل 10: جملي ۾ موجود لفظن کي انگريزي صرفيي صورتن مطابق سڃاڻڻ وارو عمل

سنڌ لفظن جو علم صرف مطابق تجزيو

علم صرف جا قسم	لفظ جي ڪل تعداد	سيڪڙو
مفرد لفظ	4	57.14
مرتب لفظ	2	28.57
مرڪب لفظ	0	0.00
ٻهرايل يا ٻٽا لفظ	0	0.00

شڪل 11: جملي ۾ موجود لفظن کي سنڌي صرفيي صورتن مطابق سڃاڻڻ وارو عمل

عالمي گرامر جو مقصد دنيا جي سمورين ٻولين کي هڪ گرامر تحت ڳنڍڻ ۽ انهن کي نشان ڏيڻ آهي. سنڌي ٻوليءَ کي اڳ ۾ ئي عالمي گرامر جي ڳالهائڻ جي لفظن سان جوڙيو ويو آهي، تنهنڪري يونيورسل ڊپينڊينسي جي پروسس ۾ سنڌي ٻوليءَ کي آڻڻ ڪو گهٽو ڏکيو ڪم نه آهي. هن تحقيق ذريعي سنڌي ٻوليءَ تي عالمي گرامر جي گهرجن مطابق ڪم ڪيو ويو آهي، جنهن سان سنڌي ٻولي يونيورسل ڊپينڊينسي جي پروسس جو حصو بڻجي ويندي. سنڌي جملي **ننڍو ڪارو ڪتو اڇتو ٻليءَ تي ٽپو ڪيو.** کي يونيورسل ڊپينڊنسي پروسس لاءِ استعمال ڪيو ويو آهي. هن جملي کي ٻن حصن يا فقرن ۾ ورهائجي ٿو. هڪ: ننڍو ڪارو ڪتو ۽ ٻيو: **اوڇتو ٻليءَ** تي ٽپو ڪيو. هن جملي ۾ فاعل ڪتو ۽ مفعول ٻلي آهي، جڏهن ته فعل ٽپو ڪيو جملي ۾ اهم ڪردار ادا ڪري ٿو. فعل ٽپو ڪيو ڪتي ۽ ٻليءَ ٻنهي سان ڳنڍيل آهي، تنهنڪري هو ڪتي ۽ ٻليءَ جي وچ ۾ هڪ رشتو پيدا ڪري ٿو. ٻليءَ کي مفعول طور به فعل ٽپو ڪيو ظاهر ڪري ٿو ڇاڪاڻ ته ٻليءَ تي ٽپو ڪرڻ جو عمل ٿيو آهي، جيڪو فاعل ڪتي طرفان ٿيو آهي. هن جملي ۾ لفظ ننڍو ۽ ڪارو اسم ڪتي جي صفت بيان ڪن ٿا. حرف جر طور استعمال ٿيندڙ لفظ تي ڪتي جو ٻليءَ تي ٽپو ڪرڻ کي ظاهر ڪري ٿو ۽ ظرف **اوڇتو** جملي ۾ وقت ۽ حالت کي ظاهر ڪري ٿو، جنهن ۾ ڪتو **ٻليءَ** تي ٽپو ڪيو. آخر ۾ جملي کي پوري دم سان بند ڪيو ويو آهي. يونيورسل ڊپينڊينسي تحت جملي کي عالمي گرامر تحت نشان ڏنا ويا آهن. ان ڪان سواءِ جملي کي انگريزيءَ ۾ ترجمو ڪري ان کي ساڳي عالمي گرامر مطابق نشان ڏنا ويا آهن ته جيئن پڪ ڪري سگهجي ته سنڌي ٻولي يونيورسل ڊپينڊنسي تحت دنيا جي ٻين ٻولين سان ڳنڍجي سگهجي ٿي. شڪل 12 سنڌي جملي کي يونيورسل ڊپينڊنسي تحت گرامر جا نشان ڏئي پيش ڪري ٿي، جنهن ۾ اهو ڏيکاريو ويو آهي ته ڪهڙو لفظ يا فقرو ڪهڙي لفظ يا فقري تي منحصر ڪري ٿو. شڪل13 سنڌي جملي کي انگريزيءَ ۾ ترجمو ڪري ان کي يونيورسل ڊپينڊنسيءَ تحت گرامر جا نشان ڏئي پيش ڪري ٿي، جنهن مان ثابت ٿئي ٿو ته سنڌي ٻولي ترجمي ڪرڻ کان پوءِ به صحيح بيهڪ ظاهر ڪري ٿي، جيڪا سنڌي ٻوليءَ جي اهميت کي ظاهر ڪري ٿي.



ننڍو ڪارو ڪتو اوڇتو ٻليءَ تي ٽپو ڪيو.

شڪل 12: سنڌي جملي کي يونيورسل ڊپينڊنسيءَ تحت عالمي گرامر مطابق نشان ڏيڻ جو عمل

شڪل12 مان ظاهر ٿئي ٿو تہ، فاعل ڪتو، جيڪو اسم آهي، صفت سان بہ ڳنڍيل آهي تہ فعل سان بہ ڳنڍيل آهيء ظرف سان بہ ڳنڍيل آهي. اهڙيءَ طرح جملي ۾ صفت، اسم، ظرف ۽ فعل جي هڪ ٻئي تي ڀاڙڻ کي ظاهر ڪيو ويو آهي. مفعول ٻلي. جيڪا اسم پڻ آهي، جو ڳانڍاپو حرف جر تي سان آهي، جيڪو ڪتي جي ڀونڪڻ کي ظاهر ڪري ٿو ۽ فعل ڀونڪيو سان ڳانڍاپو آهي،جڏهن تہ ظرف 'اوچتو' جو ڳانڍاپو فعل ڀونڪيو سان پڻ آهي ڇاڪاڻ تہ ظرف ڪتي جي اوچتي ڀونڪڻ کي ظاهر ڪري ٿو. فعل ڀونڪيو جو ڳانڍاپو اسم ڪتو، ظرف اوچتو، اسم ٻلي ۽ پوري دم سان آهي. اهي ڳانڍاپا هيٺينءَ ريت ظاهر ڪري سگهجن ٿا.

ننڍو ڪارو ڪتو

ڪتو ڀونڪيو

ڪتو ٻليءَ تي ڀونڪيو

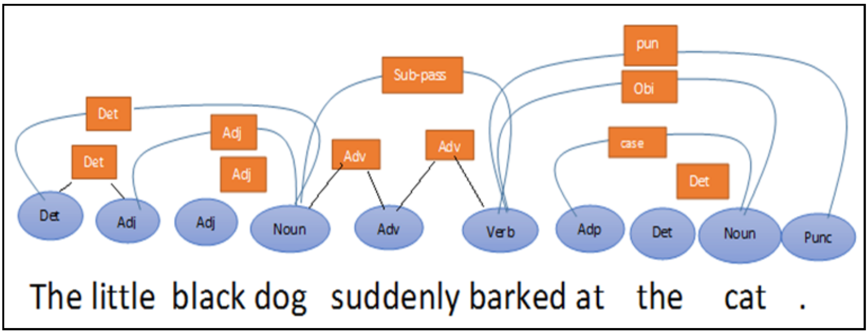
اوچتو ڀونڪيو

ڪتو اوچتو ٻليءَ تي ڀونڪيو

ٻليءَ تي ڀونڪيو

ڀونڪيو.

انگريزي جملي جا ڳانڍاپا بہ اهڙيءَ طرح آهن، تنهنڪري نتيجو ظاهر ڪري ٿو تہ: سنڌي ٻولي يونيورسل ڊيپينڊنسي لاءِ هڪ مڪمل ٻولي آهي.



شڪل 13: انگريزي جملي کي (جيڪو سنڌي مان ترجمو ٿيل آهي)

يونيورسل ڊيپينڊنسي تحت عالمي گرامر مطابق نشان ڏيڻ جو عمل

نچوڙ

هن تحقيق جو مقصد سنڌي ٻوليءَ جي ڪمپيوٽيشنل لسانيات ۽ نيچرل لينگئيجز پروسس جي حوالي سان مسئلا حل ڪرڻ هيو تہ جيئن سنڌي ٻولي يونيورسل ڊيپينڊنسيءَ جي سرشتي ۾ شامل ٿي سگهي. تحقيق ذريعي حاصل ٿيندڙ نتيجا سنڌي ٻوليءَ لاءِ

ڪمپيوٽيشنل لسانيات ۾ نيچرل لينگئيجز پروسس مطابق ناهيل اوزارن ۾ الگوريٽمس جي نتيجن کي بهتر ثابت ڪن ٿا.الگوريٽمس ۾ اوزار سنڌي ٻوليءَ جي ڪمپيوٽيشنل لسانيات ۾ نيچرل لينگئيجز پروسس جي مسئلن کي حل ڪرڻ لاءِ اهم ڪردار ادا ڪن ٿا. اهڙيءَ طرح مستقبل ۾ انهن الگوريٽمس ۾ اوزارن تي وڌيڪ ڪم ڪري انهن کي وڌيڪ بهتر بڻائي سگهجي ٿو ۽ سنڌي ٻوليءَ تي مختلف رخن کان بهتر تحقيق ۾ ڪم ڪري سگهجي ٿو. هن تحقيق ذريعي سنڌي ٻوليءَ جي مختلف جملن تي ڪمپيوٽيشنل لسانيات مطابق مختلف قسم جا عمل ڪيا ۽ بهتر نتيجا حاصل ڪيا، جنهن سان سنڌي ٻوليءَ تي مستقبل ۾ تحقيق ڪندڙن کي آساني ٿيندي.البت يونيورسل ڊپينڊينسي،لفظن جي معنا سمجهڻ، سنٽيڪٽڪ پارسنگ، چاڻ جي ڳولها، ترجمي جي نظام ۾ آوازن کي بهتر طريقي سان سمجهڻ لاءِ وڌيڪ تحقيق جي ضرورت آهي ڇاڪاڻ ته هيءَ تحقيق سنڌي ٻوليءَ لاءِ ڪمپيوٽيشنل لسانيات جو بنادي ڍانچو پيش ڪري ٿي. هيءَ تحقيق مستقبل ۾ ٿيندڙ تحقيق لاءِ دروازا کولي ٿي ته: جيئن سنڌي ٻوليءَ تي ڪمپيوٽيشنل لسانيات ۾ نيچرل لينگئيجز پروسس جي حوالي سان بهتر تحقيق ٿي سگهي.

حوالا:

1. Clark, A., Fox, C., &Lappin, S. (Eds.). (2013). *The handbook of computational linguistics and natural language processing*. John Wiley & Sons.
2. Tsarfaty, R., Seddah, D., Kübler, S., &Nivre, J. (2013). Parsing morphologically rich languages: Introduction to the special issue. *Computational linguistics*, *39*(1), 15-22.
3. Mahar, J. A., Memon, G. Q., &Danwar, S. H. (2011). ALGORITHMS FOR SINDHI WORD SEGMENTATION USING LEXICON-DRIVEN APPROACH. International journal of academic research, 3(3).
4. Rahman, M. U. (2009). Sindhi morphology and noun inflections. In *Proceedings of the Conference on Language & Technology* (pp. 74-81).
5. Mirza Kaleech Bag. Sindhi Vyakaran. Edition 4th. Sindhi Adabi Board Jamshoro, Sindh Pakistan. 2015
6. Berzak, Y., Kenney, J., Spadine, C., Wang, J. X., Lam, L., Mori, K. S., ... & Katz, B. (2016). Universal dependencies for learner English. *arXiv preprint arXiv:1605.04278*.
7. Nivre, J. (2015, April). Towards a universal grammar for natural language processing. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 3-16). Springer, Cham.
8. Motlani, R. (2016). Developing language technology tools and resources for a resource-poor language: Sindhi. In *Proceedings of the NAACL Student Research Workshop* (pp. 51-58).
9. Mahar, J. A., Shaikh, H., & Memon, G. Q. (2012). A Model for Sindhi Text Segmentation into Word Tokens. *Sindh University Research Journal-SURJ (Science Series)*, *44*(1).

10. Dootio, M.A., Wagan, A.I. Syntactic parsing and supervised analysis of Sindhi text. Journal of King Saud University – Computer and Information Sciences (2017), https://doi.org/10.1016/j. jksuci .2017.10.004
11. Dootio, M. A., & Wagan, A. I. (2018). Unicode-8 based linguistics data set of annotated Sindhi text. *Data in Brief*, Volume 19, 2018, pp. 198-213
12. Dootio, M. A., & Wagan, A. I. (2017). Automatic Stemming and Lemmatization Process for Sindhi Text. Proceeding of ICLAP, 3rd International conference of Linguistic Association Karachi, Sindh Pakistan
13. Mazhar Ali and Asim Imdad Wagan, "Sentiment Summerization and Analysis of Sindhi Text" International Journal of Advanced Computer Science and Applications (IJACSA), 8(10), 2017. http://dx.doi. org /10.14569/IJACSA.2017.081038

سنڌي صورتخطيءَ ۾ لکيل مخصوص انگريزي لفظن جي هِجي (Spelling).

1. Computational Linguistics
2. Computational Linguist
3. Artificial Intelligence
4. Algorithm
5. Universal Grammar
6. Process
7. Corpus
8. Morphology
9. Natural Languages Process
10. Search engine
11. Information Retrieval
12. http://www.sindhinlp.com/
13. Lemma
14. Word Net
15. Tree Bank
16. PART
17. Participle
18. X
19. Layers
20. Lemmatization
21. Lemma
22. Sentiments
23. Machine Learning
24. Deep Learning
25. Text
26. Vectors
27. Matrices
28. Text Parsing
29. Universal Part of Speech (UPOS)
30. PART
31. Simple word
32. Complex Word
33. Compound Words
34. Reduplicated words
35. Syntactic Parsing
36. Syntactically Parse
37. Statistical
38. Universal Dependency