ڈاکٹر حافظ صفوان محمد چوہان

صدر شعبہ کمپیوٹر و تربیت،پی ٹی سی ایل ٹریننگ کالج، فیصل آباد

ڈاکٹر ظہیر احمد

سی سی ایس آر، یونیورسٹی آف سرے،برطانیہ

انجینئر سارہ سلیم

اسسٹنٹ منیجر،سوئچنگ نیٹ ورک، فیصل آباد

مشینی ترجمے کی تکنیکیں، مسائل اور اُن کا حل:
اردو کے تناظر میں ایک مطالعہ

Dr Hafiz Safwan Muhammad Chohan

Head of Computer Tranning Department, PTCL Tranning College, Faisalabad

Dr Zaheer Ahmed

CCSR, University of Surrey, United Kingdom

Engineer Sara Saleem

Assistant Manager, Switching Network, Faisalabad

Techniques of Computer Aided Translation, Problems and Their Solutions:
A study focused on Urdu Literature

This article introduces 4 major types of computer-aided translation services with particular reference to Urdu. Machine Translation using Rule-based methods, Example-based methods, Statistical model and Hybrid approaches are in wide current. Available online translation services are mentioned categorically.

Environment of this article is made reasonably digestable for the people of Urdu literature so as to retain their attention in these pages of stark technical discussion.

......مشینی ترجمہ کار زیادہ سے زیادہ یہی کر سکتا ہے کہ مترجِم کے کام کا آسان تر حصہ خود کار طریقے سے کر دے۔ مشکل حصہ تو وہی ہوتا ہے جس میں ماخذ متن کے معانی میں ابہام دور کرنے کے لیے تحقیق پر بہت سا وقت صَرف کرنا پڑتا ہے، اور جسے مطلوبہ زبان

میں ڈھالتے وقت گرامر اور لفظیات کے بارے میں بہت سے فوری اور ہنگامی فیصلے کرنا پڑتے ہیں۔...... (Claude Piron, long-time translator for UN and the WHO)

اصولی اور تکنیکی طور پر اِس مقالے کا دائرۂ کار کسی بھی زبان سے اردو اور اردو سے کسی بھی زبان میں کسی بھی قسم کا متن ترجمہ کرنے پر پھیلا ہوا ہے۔ تاہم مثالوں کی حد تک یہ صرف اردو-انگریزی اور انگریزی-اردو سے بحث کرتا ہے۔

یہ مقالہ فروری ۲۰۱۰ء کے آخری ہفتے میں مکمل ہوا۔ اِس میں شامل معلوماتِ عامہ کو صرف اِنہی تاریخوں تک درست سمجھا جائے۔

کلیدی الفاظ

مشین ریڈایبل اردو، کارپس، ماخذ جملہ (Source Sentence)، انسانی زبان (Natural Language)، ماخذ زبان (Source Language)، مطلوبہ زبان (Target Language)، مترجم (Human Translator)، شعبے کی زبان (Term-base)، لفظیات (Lexis)، متن (Discourse)، نقلِ حرفی (Transliteration)، کمپیوٹیشنل گرامر (Computational Grammar)، فیصلہ کرنے کی صلاحیت (Decision-making)، قواعدی ترجمہ (Rule-based Translation)، مثالی ترجمہ (Knowledge-based Translation)، شماریاتی ترجمہ (Statistical Translation)، دوغلا ترجمہ (Hybrid Translation)، مصنوعی ذہانت (Artificial Intelligence)، معانی میں ابہام (Ambiguity)۔

☆ اِس مقالے میں لفظ زبان زیادہ تر انسانی زبان کے مفہوم میں استعمال کیا گیا ہے۔

☆ ترجمہ یا زبان کا ترجمہ سے بھی عموماً مراد زبانوں کے کسی جوڑے (مثلاً اردو، انگریزی) کا باہم ترجمہ ہے۔

مخففات

ٹرانسلیشن میموری:	TM: Translation Memory
مشین ٹرانسلیشن، مشینی ترجمہ:	MT: Machine Translation
ٹریڈوس:	TRADOS: Translator for DOS
مصنوعی ذہانت:	AI: Artificial Intelligence
شماریاتی مشینی ترجمہ:	SMT: Statistical Machine Translation
ناموجود الفاظ:	OOV: Out of Vocabulary Words
رک:	رجوع کیجیے
مرکزِ تحقیقاتِ اردو پاکستان:	CRULP: Centre of Research in Urdu Language Processing
مقتدرۂ قومی زبان پاکستان:	NLA: National Language Authority, Islamabad, Pakistan

اصطلاحات

شعبے کی زبان: Term-base

عالمی ترجمہ کار: Universal Translator

کمپیوٹیشنل لسانیات: Computational Linguistics

اِطلاعیات: Informatics

بالائے لغت معنی: Ultra-dictionary meaning

پلیٹ فارم سے ناوابستہ: Platform-Independent

روزمرہ زبان/ بول چال: Contemporary [use of] Language

انسانی/فطری زبان: Natural Language

کارپس/ مثال گھر/ قاموس الامثال: Corpus

لغتیاتی تجزیہ: Lexical Analysis

ڈیٹا کا ذخیرہ/ مال خانہ: Repository

☆تجارتی نشانات: اِس مقالے میں Microsoft،Google،وغیرہ الفاظ استعمال کیے گئے ہیں؛ یہ الفاظ مختلف اداروں کے تجارتی نشانات (ٹریڈ مارک) ہیں۔

0: تعارف

مشینی ترجمے سے مراد ایک انسانی زبان (ماخذ زبان) کے متن کا دوسری انسانی زبان (مطلوبہ زبان) میں سافٹ ویئر کی مدد سے ترجمہ کرنا ہے۔ مشینی ترجمہ شعبۂ کمپیوٹیشنل لسانیات کی ایک شاخ ہے۔

مشینی ترجمے کا سادہ ترین انداز یہ ہے کہ ایک زبان کے الفاظ کو دوسری زبان کے الفاظ سے صرف بدل دیا جائے (Substitution)۔ کارپس کی تکنیک کے استعمال سے نسبتًا پیچیدہ ترجمے جن میں زبان کی ساخت، جملوں کی بنت اور پہچان، اور محاوروں اور دیگر اجزائے کلام وغیرہ کو ملحوظ رکھا گیا ہو، بہتر انداز میں کیے جا سکتے ہیں۔

[۱]

1: ترجمے کا عمل کیا ہے؟

ترجمے کا عمل دو سطحوں پر مشتمل ہوتا ہے:

۱۔ ماخذ متن کا مفہوم سمجھنا (Decoding the meaning of the source text)، اور

۲۔ اِس مفہوم کو مطلوبہ زبان میں ڈھالنا (Re-encoding this meaning in the target language)۔

اِس بظاہر بہت سادہ سے عمل کے پیچھے ایک بے حد پیچیدہ کارروائی چھپی ہوئی ہے جس میں نرے الفاظ ہی نہیں بلکہ وجدان بھی کارفرما ہوتا ہے۔ ماخذ متن کو پورے طور سے سمجھنے کے لیے مترجم کو متن کی خصوصیات اور ساخت کا تجزیہ اور تشریح کرنا ہوتی ہے۔ یہ وہ عمل ہے جس کے لیے ماخذ زبان کی نہ صرف گرامر، نحو، لفظیات اور محاورات وغیرہ کا گہرا علم ہونا ضروری ہے بلکہ نواحی مطالعے کے ساتھ ساتھ زبان کے بولنے والوں کی

ثقافت سے آگاہی بھی اشد ضروری ہے۔ مترجم کو اِس متن کے مطلوبہ زبان میں ڈھالنے کے لیے مطلوبہ زبان کا بھی اتنا ہی گہرا علم رکھنا ضروری ہے۔

اب مشینی ترجمے کے لیے آزمائش شروع ہوتی ہے: کمپیوٹر کو کس طرح پروگرام کیا جائے کہ وہ متن کو ویسے ''سمجھے'' جیسا کہ انسان سمجھتا ہے، اور پھر مطلوبہ زبان کے لیے ایسا متن ''بنائے'' جو بالکل ''ویسا'' محسوس ہو جیسے کہ اسے کسی آدمی نے لکھا ہے۔

چنانچہ مشین کو استعمال کرتے ہوئے ترجمہ کرنے کے کئی طریقِ کار استعمال کیے جاتے ہیں۔

1.1: ترجمہ اور ترجمانی

ترجمے (Translation) اور ترجمانی (Interpretation) کے درمیان حدِ فاصل نہیں کھینچی جاسکتی۔ اصولاً ترجمے کو ترجمانی ہی کہنا چاہیے کیونکہ مافی الضمیر کی ادائیگی ترجمانی ہی سے ہوسکتی ہے۔ ترجمے کی دو بنیادی اقسام ہیں: بلاواسطہ (Direct) اور بالواسطہ (Indirect)۔ بلاواسطہ ترجمے کو بعض اوقات لفظ بلفظ (Word-for-word) اور لفظی (Literal) ترجمہ بھی کہا جاتا ہے، اور اس کا تقابل محاوراتی (Idiomatic) ترجمے سے کیا جاتا ہے۔

[۲]

2: کمپیوٹر کی معاونت سے کیے گئے ترجمے — طریقِ کار

مشینی ترجمہ کار لسانیاتی قواعد کی بنیاد پر بنایا گیا کوئی طریقہ استعمال کرسکتا ہے جس میں الفاظ لسانی حساب سے ترجمہ کیے گئے ہوں۔ کہنے کی حد تک یہ مناسب ترین طریقہ ہے جس سے مطلوبہ زبان کے الفاظ ماخذ زبان کے الفاظ کو تبدیل کردیتے ہیں۔ اِس نکتے پر اکثر بحث رہتی ہے کہ مشینی ترجمے کی کامیابی کے لیے ضروری ہے کہ ماخذ زبان کو سمجھنے کا مسئلہ پہلے حل کیا جائے، یعنی مشین کم سے کم ایک زبان کو پوری طرح سے ''سمجھنے'' کے قابل بنادی جائے۔

مشینی ترجمے کے بیشتر قواعدی طریقے (Rule-based methods) داخل کردہ متن کو زبان کی قواعدی حیثیت میں ڈھالتے (Parse) اور عام طور سے ایک عارضی رموزی صورت (Intermediary symbolic representation) میں منتقل کردیتے ہیں جس سے مطلوبہ زبان کا متن تشکیل پاتا ہے۔ متن کی اس عارضی رموزی صورت کی نوعیت کے مطابق بین اللسانی مشینی ترجمہ (Interlingual MT) یا تبدیلی کی بنیاد پر ترجمہ (Transfer-based MT) میں سے کوئی طریقہ اختیار کیا جاتا ہے۔ ان طریقوں سے ترجمہ کرنے کے لیے بڑے پیمانے پر گرامر کے قوانین کے ساتھ ساتھ بڑے بڑے ذخیرہ ہائے الفاظ اور لغات (Lexicons) کی بھی ضرورت ہوتی ہے جن میں تصریفی، نحوی اور معنوی معلومات ہوں۔ اگر بہت سارا تربیتی مواد (Training Data) مشینوں میں جمع کردیا جائے تو مشینی ترجمے کے پروگرام ماخذ متن کا لگ بھگ ویسا ہی ترجمہ کردیتے ہیں جیسا ایک زبان کا باصلاحیت بولنے والا دوسری زبان میں کرے گا۔ مشکل یہی ہے کہ مناسبِ حال ڈیٹا کافی مقدار میں مہیا نہیں ہوتا، مثلاً شماریاتی (Statistical) طریقوں سے کیے جانے والے ترجمے کے لیے ناگزیر بہت سے ذولسانی کارپس ڈیٹا کی ضرورت اُس وقت نہیں ہوتی اگر ہمیں گرامر کے قواعد کی مدد سے ترجمہ کرنا ہو۔ اس کے بالمقابل گرامر والے طریقوں سے کام لینے کے لیے کائیاں ماہرینِ لسانیات کی ضرورت ہوتی ہے جو گرامر کو مشین کی ضرورتوں کے مطابق ڈیزائن کرسکیں۔ باہم ملتی جلتی زبانوں میں ترجمہ کرنے کے لیے سطحی تبدیلی (Shallow-transfer) کی تکنیک بھی استعمال کی جاسکتی ہے۔

کمپیوٹر کی معاونت سے کیے جانے والے ترجموں کی چار اصولی شاخیں یا طریقے ہیں:

۱۔ قواعدی یا مبنی بر قواعد (Rule-based) مشینی ترجمے۔

۲۔ مثالی یا مبنی بر اَمثلہ (Example-based) ترجمے، جو ٹرانسلیشن میموری (TM) کی بنیاد پر کیے جاتے ہیں۔

۳۔ شماریات (Statistics) کی بنیاد پر کیے گئے ترجمے (SMT)۔

۴۔ دوغلا مشینی ترجمہ (Hybrid MT)۔

ترجمے کی پہلی قسم میں کمپیوٹر ایک سمجھدار (Intelligent) مشین کے طور پر کام کرتا ہے؛ دوسری قسم میں یہ ایک بے عقل (Duffer) مشین کی مانند کام کرتا ہے؛ تیسری قسم میں یہ ایک انتہائی سمجھدار مشین کے طور پر ناپ تول کر امکانات کی بنیاد پر کام کرتا ہے؛ چوتھی قسم میں یہ قواعدی اور شماریاتی طریقہائے تراجم کی خصوصیتوں کو ملا جلا کر استعمال کرتا ہے۔ اِس بات کا ذکر بے جا نہ ہوگا کہ ایک زبان سے دوسری زبان میں کمپیوٹر کی معاونت سے کیے گئے ہر طرح کے ترجمے (Computer-Aided/Assisted Translation) کو کمپیوٹر اور کمپیوٹیشنل لسانیات کی عملداری میں ایک ہی عمومی اصطلاح یعنی ''مشینی ترجمہ'' سے یاد کیا جاتا ہے۔

ذیل میں اِن سب طریقوں کا تعارف اور آپس میں کا فرق پیش کیا جاتا ہے۔

2.1: قواعدی یا مبنی بر قواعد (Rule-based) مشینی ترجمے

قواعدی ترجمہ کار پروگرام زبان کے قواعد، قوانین اور زبانوں کے ثقافتی ماحول کو سمجھ کر اُن پر عمل کرتے ہوئے ترجمہ کرتا ہے۔ اِس تکنیک میں کمپیوٹر کو فقرات کی بجائے الفاظ، اُن کے معنی اور اُن کی دیگر خصوصیات (Attributes) ''سمجھائی'' جاتی ہیں، اور ساتھ ہی زبان کی گرامر کے سارے قواعد اور نحو۔ ترجمے کی اِس تکنیک میں کمپیوٹر اپنی مکسوبہ ذہانت کو بروئے کار لاتا ہے جسے تکنیکی زبان میں مصنوعی ذہانت (Artificial Intelligence) کہتے ہیں۔ لفظوں، زبان اور قواعدِ زبان کے بارے میں ایسی معلومات کارپس سے بسہولت دستیاب ہو سکتی اور اخذ کی جا سکتی ہیں۔ چنانچہ قواعدی ترجمہ کار بنیادی طور پر کارپس پر انحصار کرتا ہے۔

(آگے چلنے سے پہلے یہ سمجھ لیا جائے کہ کمپیوٹر کی لفظیات میں ''جملہ'' کسے کہتے ہیں۔ کمپیوٹر کی زبان میں ''لفظ'' (Word) کی تعریف یہ ہے کہ یہ حروفِ تہجی کا ایسا سلسلہ (String) ہے جو دو خالی جگہوں کے درمیان ہوتا ہے یا جس کے ایک جانب کوئی علامتِ وقف لگی ہوتی ہے۔[۲] چنانچہ اس ڈیٹابیس میں رکھے گئے مواد — یعنی جملوں — کی تکنیکی تعریف یہ ہے کہ یہ حروفِ تہجی کے ایسے سلسلے ہوتے ہیں جن کے آخر میں رموزِ اوقاف میں سے کوئی ایسی علامت لگی ہو جن پر قواعد دانوں کے نزدیک جملہ جائز طور پر ختم ہوتا ہے۔)

مثال لیجیے کہ یہ ایک انگریزی جملہ، جسے اصطلاحاً Source Sentence (ماخذ جملہ) کہتے ہیں، اردو ترجمے کے لیے داخل کیا گیا:

He is coming.

اِس جملے کے داخل ہونے پر سافٹ ویئر یہ دیکھے اور طے کرے گا کہ یہ جملہ کس زمانے (Tense) سے تعلق رکھتا ہے۔ ''is'' سے معلوم ہوگا کہ یہ فعل حال ہے۔ ''-ing'' سے پتہ ملے گا کہ یہ فعل حال جاری ہے۔ پھر یہ سافٹ ویئر دیکھے گا کہ مطلوبہ زبان (اِس مقالے کی حد تک اردو) میں جملے کی جنس اور نحوی ترتیب کیا ہے (یعنی اردو کی صورت میں SOV اور انگریزی کی صورت میں SVO) اور پھر اِن قواعد کے مطابق ترجمہ کرکے اور ترجمہ شدہ مواد کو اردو کی نحوی ترکیب میں ترتیب دے کر مطلوبہ جملہ سامنے لائے گا: وہ آرہا ہے۔ (اگر سافٹ ویئر

جملے کو اردو کی نحوی ترکیب میں نہ ڈھال سکے تو لامحالہ یہ ترجمہ ملے گا: وہ بس آ رہا۔)

فرض کیجیے کہ صارف کو مطلوبہ جملہ درست ترجمہ شدہ نہیں ملا اور اس میں قواعد کی یا کوئی علمی غلطی پائی جاتی ہے، تو اب اگر وہ ترجمے میں درستی کرتا ہے تو یہ تبدیلی/ترمیم صرف صارف کی سطح تک رہے گی، اور وہ بھی صرف اسی موقع کے لیے۔ یہی صارف اگر آئندہ یہی جملہ ڈالتا ہے تو بھی یہی غلطی دوہرائی جائے گی تا آنکہ قواعد میں تبدیلی/ترمیم کی جائے۔ (یہ تبدیلی سافٹ ویئر کی پروگرامنگ سطح پر ہوتی ہے جس کا اختیار سافٹ ویئر بنانے والے اِدارے ہی کو ہوتا ہے نہ کہ صارف کو۔) اِس صورتِ حال کو مثالوں سے سمجھیے۔ ایک ماخذ جملہ داخل کیا گیا:

I took breakfast.

قواعدی ترجمہ کار سافٹ ویئر سے ترجمہ حاصل ہوا: میں نے ناشتہ حاصل کیا یا میں نے ناشتہ لیا۔ جب کہ درست ترجمہ ہے: میں نے ناشتہ کیا۔ لیکن یہ ترجمہ قواعدی ترجمہ کار سافٹ ویئر سے سیدھے سبھاؤ حاصل نہیں ہوتا۔

اِسی طرح ایک ماخذ جملہ داخل کیا گیا:

Springs spring in spring.

ترجمہ حاصل ہوا: بہار میں بہاریں بہاریں۔ جب کہ درست ترجمہ ہے: موسم بہار میں چشمے پھوٹتے ہیں۔ چنانچہ قواعدی ترجمہ کار سافٹ ویئر از خود کامل طور پر یہ فیصلہ کرنے سے تا حال قاصر ہے کہ Spring کا مطلب بہار کہاں آئے گا، پھوٹنا کہاں اور چشمہ یا اچھلنا کہاں۔ سافٹ ویئر کو دی جانے والی مصنوعی ذہانت کی کیفیت اور کمیت میں اضافہ ایسے مسائل کے حل سامنے لاتا رہتا ہے اور نتیجۃً ترجمے کی سہولت بہتر ہوتی جاتی ہے۔

2.1.1: قواعدی ترجمہ کار کا دائرۂ کار: اسماء و افعال کی بجائے نحوی ترجمہ

صَرف (Etymology) کا لفظی مطلب ہے ہیر پھیر، اور نحو (Syntax) کا لفظی معنیٰ سمت ہے۔ صَرف میں الفاظ بنتے ہیں، نحو میں کلام بنتا ہے۔ جملے میں الفاظ کی جوڑ جاڑ یعنی Assembling نحو کہلاتی ہے۔ جملے کی نحوی ترکیب کی جائے تو تین طرح کے الفاظ ملتے ہیں:

۱: اسماء (واحد یا/اور جمع صورتوں میں): اسم تنہا اپنا مفہوم دیتا ہے لیکن اِس میں زمانہ نہیں ہوتا۔

۲: افعال (تصریفی صورتیں): فعل تنہا مفہوم بھی دیتا ہے اور اِس میں زمانہ بھی پایا جاتا ہے۔

۳: حروف (نے، کو، پر، وغیرہ): حرف تنہا اپنا مفہوم ادا نہیں کرتا جب تک کہ وہ اسماء و افعال کے ساتھ مربوط نہ ہو۔

کمپیوٹر کے ذریعے کیے جانے والے تراجم میں ایک طویل آزمائشی دور گزارنے کے بعد تحقیقات کا حتمی رخ یہ بنا ہے کہ اسماء اور افعال کی بدلتی تصریفی صورتوں کے ترجمے کی بجائے سافٹ ویئر کو جملے کی نحوی ترکیب سمجھا کر اُسے دوسری زبان میں ڈھالنے پر لگایا جائے۔ یہ کام، ظاہر ہے کہ، تبھی ہو سکے گا جب سافٹ ویئر ماخذ زبان کے ساتھ ساتھ مطلوبہ زبان کی نحو کو بھی بخوبی سمجھتا ہو۔ یہ سب کام مصنوعی ذہانت کا ہے۔ چنانچہ تمام زبانوں کی نحو کا الگ الگ اور ان میں کی مشترکات کا مطالعہ کیا جا رہا ہے اور ایک زبان کے جملے کا دوسری زبان میں بامعنی اور مربوط ترجمہ کیے جانے کی کوششیں ہو رہی ہیں۔ نتائج حوصلہ افزا ہیں۔ اِنہی تحقیقات کے نتائج اگر مجموعی طور پر پھیلا لیے جائیں تو اصولاً کسی بھی ماخذ زبان کے پورے جملے کو کسی بھی مطلوبہ زبان میں صحیح طور سے ترجمہ کرنے میں مددگار رہوں گے۔

ہمارے پاس نحو کی بنیاد پر ترجمہ کرنے کی ایک تاریخی اور عمرانی شہادت بھی ہے، گو سائنس اسے پہلی نگاہ میں پس وپیش کے ساتھ قبول کرے۔ اللہ پاک نے قرآنِ پاک میں فرمایا ہے: وَعَلَّمَ آدَمَ الاَسْمَآءَ كُلَّها [۳] یعنی بنی نوعِ انسان کو اللہ پاک نے چیزوں کے نام سکھا دیے۔ ناموں کی حد تک یہ تحقیق (سائنسی بھی اور عمرانی بھی) سامنے آئی ہے کہ بہت سی چیزوں کے نام پوری دنیا کی بیشتر زبانوں میں ذرا ذرا سے فرق کے ساتھ ملتے جلتے ہیں۔ [۴] ہاں! الفاظ کو جملوں میں پرو کر مافی الضمیر کا اظہار کرنا...... یہ اللہ نے انسانوں پر چھوڑ دیا۔ اگرچہ نئے اسماء کا اضافہ بھی ہوتا رہتا ہے اور ایسی چیزیں جن کا وجود باقی نہیں رہا اُن کے نام بھی محو ہوتے جاتے ہیں۔ نحو، البتہ، انسان نے اپنی عقل سے خود بنائی ہے۔ جیسے جیسے انسانیت مہذب اور مدنی ہوتی گئی، بین اللسانی روابط بڑھے اور سبھی زبانوں کی تراش خراش ہوتی گئی۔ یاد رہے کہ تراش خراش سے مراد نحو میں آنے والی تبدیلیاں اور چستی ہے، اسماء تقریباً وہی کے وہی رہے ہیں البتہ اِن میں اضافہ روز افزوں ہے۔ لفظِ نحو کی تکرار سے استاد ابراہیم ذوق کا شعر کیا خوب یاد آیا: ؎

کبھی ہمت تھی مری قاعدۂ صَرف میں صَرف
تھی کبھی نحو میں ہر نحو مجھے محویت

2.1.2: قواعدی ترجمہ کار کا دائرۂ کار: کچھ اور باتیں

جملوں میں الفاظ اور افعال محض قواعدی (Grammatically) اعتبار سے نہیں بلکہ تناظر (Context) کے اعتبار سے بھی درست ہونا ضروری ہوتے ہیں۔ ممکن ہے کہ ایک جملہ قواعد کے اعتبار سے درست ہو لیکن عام سمجھ بوجھ، زمینی حقائق یا ثقافت کے اعتبار سے اُس میں کہی گئی بات ممکن الوقوع نہ ہو۔ الفبائی ترتیب میں لکھے چند مثالی جملے دیکھیے:

امام صاحب نے عید میلاد النبیؐ کی نماز پڑھائی۔
انڈے سے درخت نکلا۔
بھینس پڑھ رہی ہے۔
پانی نہا رہا ہے۔
پہاڑ اُڑ رہا ہے۔
تاج محل کا رنگ سرخ ہے۔
حج محرم کی دس تاریخ کو ہوتا ہے۔
درخت گا رہا ہے۔
درختوں پر چوزے لگ رہے ہیں۔
زمین سے لڑکا اگا۔
سورج سے ٹھنڈا اندھیرا نکلتا ہے۔
شیر انڈے دیتا ہے اور شیرنی انڈے سیتی ہے۔
گندم کا درخت کتنا اونچا ہوتا ہے؟

لڑکی کے سر پر سہرا سجا۔
لڑکی والے بارات لے کر آئے۔
لڑکے نے گھونگٹ نکالا۔
وائٹ ہاؤس سیاہ رنگ کی عمارت ہے۔

وغیرہ وغیرہ۔ یہ ایسے جملے ہیں جو قواعد کے اعتبار سے درست لیکن مختلف دیگر وجوہات سے نادرست ہیں۔ یہ عام، موٹی سمجھ کی بات ہے کہ ہر فاعل ہر فعل نہیں کرسکتا اور نہ ہی ہر مفعول پر ہر فعل ہوسکتا ہے۔ مشین کو ایسی باریکیاں، البتہ، مصنوعی ذہانت کے ذریعے سے سکھائی جاتی ہیں۔

تمام انسانی زبانوں کا اپنا اپنا ثقافتی پہناوا بھی ہوتا ہے۔ ایک زبان کا ایک ہی لفظ مختلف علاقوں اور ماحولوں میں از سرِ مختلف مفہوم رکھ سکتا ہے۔ مثلاً لفظِ مہتر اور مہترانی کا معنیٰ پاکستان کے شمالی علاقہ جات میں بالکل مختلف ہے۔ علاقائی اور انتظامی حدود میں بدلتے معنی دینے والے اور بھی کئی الفاظ ہیں مثلاً راؤنڈ، منشی، جمعدار وغیرہ۔ اسی طرح بے حد وسیع ثقافتی معنی رکھنے والے الفاظ کی بھی لمبی لمبی فہرستیں ہیں مثلاً بابا، مائی، اماں، بزرگ، بھائی، بہن وغیرہ۔ ترجمہ کرتے میں مطلوبہ جملے میں الفاظ وتراکیب کا ثقافتی طور پر قابلِ قبول ہونا بھی ایک اہم ضرورت ہوتا ہے۔ مشین کو ایسے مخصوص الفاظ وتراکیب کا دوطرفہ استعمال خاص طور سے سکھایا جاتا ہے۔

ہر زبان میں ایک لفظ (یعنی ایک صورتِ املاء) بیک وقت ایک سے زیادہ قواعدی تعریفوں کا حامل ہوسکتا ہے۔ اردو میں یہ صورتِ حال سب سے زیادہ نظر آتی ہے۔ مثلاً مصدر چمٹنا سے صیغۂ ماضی مذکر بنتا ہے: چمٹا۔ لیکن چمٹہ / چمٹا ایک آلہ بھی ہے جو گھریلو/ طباخی/ فنِ جراحت وغیرہ میں استعمال ہوتا ہے۔ اس آلے کی تصغیر چمٹی بھی ایک عام لفظ ہے۔ یہی لفظِ چمٹی مصدر چمٹنا سے صیغۂ ماضی مونث بھی ہے۔ پھر چمٹا ہماری لوک موسیقی کا ایک ساز بھی ہے، جس سے متعلقہ تراکیب چمٹا نواز، چمٹا بجانا، وغیرہ ہیں۔ چنانچہ ایسے الفاظ کے لیے مشین کو دی جانے والی ذہانت بہت پیچیدہ ہوتی ہے۔

اردو میں قواعد کے اعتبار سے مشین کو دی جانے والی ذہانت، اور الفاظ کے مفاہیم کو درست طور سے سمجھنے اور ترجمے کو تناظر کے اعتبار سے کرنے کے قابل بنانے کے لیے کیے گئے ایک مطالعے کے نتائج (صفوان محمد: ۲۰۰۹ء) اگر ملاحظہ کر لیے جائیں تو بہتر ہے۔ رک: اردو کارپس: تکنیکی تعارف، اہمیت، ضرورت اور دائرہ و لائحۂ عمل، مشمولہ جرنل آف ریسرچ، بہاء الدین زکریا یونیورسٹی ملتان، شمارہ-۱۴، ۲۰۰۸؛ عنوانات 4.4 و 4.5۔

2.1.3: قواعدی ترجمہ کار کی بنیادی ضرورت: کمپیوٹیشنل گرامر

چنانچہ قواعدی ترجمہ کار سافٹ ویئر میں زبان کی گرامر کے سارے قوانین، لفظوں سے متعلق تمام ضروری معلومات، الفاظ کی تصریفی شکلوں کے تمام تر امکانات، اور تمام ممکن زمانے (Tense) وغیرہ شامل کیے جاتے ہیں تاکہ مشین کسی بھی ماخذ/ مطلوبہ متن کی گرامر اور الفاظ کی حتمی صورتِ املاء کو جانچ کر فیصلے کرسکے۔ صارف کے لیے یہ سہولت ہوتی ہے کہ وہ حسبِ ضرورت گرامر کے اصولوں کے نفاذ میں نرمی اور سختی کرسکتا ہے۔ نیز موقع کی مناسبت سے کسی جملے یا پورے متن کے لیے گرامر کو غیر موثر (Inactive) بھی کیا جاسکتا ہے۔ وغیرہ۔

کمپیوٹیشنل گرامر سے متعلق یہ مشینی معلومات اس انداز میں رکھی جاتی ہیں کہ یہ دوسری زبانوں کی کمپیوٹیشنل گرامر کے ساتھ مل کر کام کرسکیں۔ ظاہری بات ہے کہ ترجمہ جس زبان سے ہوگا—اور جس زبان میں ہوگا— دونوں کی قواعد کا مشین پر یکساں طور پر کام کرتے ہونا

ایک بنیادی ضرورت ہے۔

اردو کے لیے کمپیوٹیشنل گرامر ابھی تک نہیں بن پائی ہے۔ کچھ اساسی کوششیں البتہ ہوئی ہیں۔ مثال کے طور پر اردو کے اجزائے کلام (Parts of Speech) پر Andrew Hardie کا کیا ہوا کام۔ ملاحظہ کیجیے:

www.ling.lancs.ac.uk/profiles/Andrew-Hardie/

2.1.4: قواعدی ترجمہ کار کی اقسام

قواعدی ترجمہ کار سافٹ ویئروں کے مثالی نمونوں (Paradigm) میں ترجمے کے مندرجۂ ذیل طریقے ہوتے ہیں:

۱: تبدیلی کی بنیاد پر مشینی ترجمہ (Transfer-based MT): اِن طریقوں میں ماخذ متن کے لفظوں، فقروں اور محاوروں کو مطلوبہ زبان میں مناسبت سے ترجمہ کرکے ایک عارضی رموزی صورت (Intermediary symbolic representation) میں منتقل کیا جاتا ہے۔ عارضی رموزی صورت میں موجود اِس مواد کا دونوں زبانوں سے تعلق (Dependence) ہوتا ہے، جس سے دونوں زبانوں کے ثقافتی رویوں کے تناظر میں مطلوبہ زبان میں ترجمہ کرنے کی کوشش کی جاتی ہے۔ یہ ترجمہ صرف نحوی بھی ہوسکتا ہے جسے Superficial یا Syntactic Transfer کہتے ہیں، اور یہ ایسی زبانوں کے لیے بہت کارآمد ہے جو ایک ہی خاندان سے ہوں یا جن کی ترکیبِ نحوی ملتی جلتی ہو۔ یہ ترجمہ معنوی بھی ہوسکتا ہے جسے Semantic یا Deep Transfer کہتے ہیں۔ اِس صورت میں یہ دور کا تعلق رکھنے والی زبانوں کے باہم ترجمے کے لیے بھی کارآمد ہوتا ہے۔

۲: بین اللسانی مشینی ترجمہ (Interlingual MT): اِن طریقوں میں ماخذ متن کو ایسی عارضی رموزی صورت میں تبدیل کیا جاتا ہے جو ہر زبان سے مکمل طور پر ناوابستہ (Independent) ہو۔ متن کی اِس عارضی رموزی صورت کو موقع کی مناسبت سے کسی بھی مطلوبہ زبان میں ترجمہ کرلیا جاتا ہے۔ اِن طریقوں کا فائدہ یہ ہے کسی بھی دو زبانوں کا آپس میں ساتھن بنا ضروری نہیں ہوتا، کہ لازماً فلاں زبان سے فلاں زبان ہی میں ترجمہ کیا جائے۔ ظاہراً یہ تکنیک بہت اچھی ہے، لیکن ایسا بین اللسانی ترجمہ (Interlingua) جو سبھی زبانوں سے بالکل ناوابستہ ہو اُس کا ایسی صورت میں سبھی زبانوں کے لیے کارآمد رہنا ایک طرح سے دیوانے کا خواب ہی ہے جب کہ مشینی ترجمے کے لیے زبانوں کی تعداد بھی بڑھتی جارہی ہے۔

۳: لغت کی بنیاد پر مشینی ترجمہ (Dictionary-based MT): اِس طریقے میں مشینی ترجمہ کار اپنے پاس موجود کسی خودکار لغت کی بنیاد پر ایک متن کا دوسری زبان میں نرا ترجمہ کرکے لفظوں کی قطار بنائے چلا جاتا ہے، ایسا ترجمہ جس میں الفاظ کا آپس میں کوئی تعلق نہ ہو۔ لغات الفاظ کی تصریفی صورتوں کے حامل ہو بھی سکتے ہیں اور نہیں بھی۔ ترجمے کی یہ تکنیک جہاں شاید سب سے زیادہ غیر پیچیدہ ہے وہیں بڑی تعداد میں رکھے گئے چھوٹے چھوٹے فقرات (جملے نہیں) کے ترجمے کے نہایت شاندار نتائج دیتی ہے جیسے مثلاً مصنوعات کی فہرستیں، تعارفی کتابچے، تشہیری مواد، وغیرہ۔ ترجمے کا یہ طریقہ دستی ترجمے کے لیے بہت معاون ہوسکتا ہے اگر مترجم دونوں زبانوں کا ماہر

ہواورنحواورگرامرکی درستی کرسکتا ہو۔

2.2: مثالی یا مبنی بر اَمثلہ ترجمے

مثالی ترجمہ کار سافٹ ویئر ذولسانی کارپس اوراپنے پاس رکھے ذخیرۂ علم (Knowledge-base) کے بیک وقت استعمال سے ترجمہ کرتا ہے، اورٹرانسلیشن میموری (TM) میں ماخذ متن سے مکمل یا جزوی مماثلت رکھنے والے پہلے سے موجود تراجم کے پارچوں کو ہاتھ کے ہاتھ سامنے لے آتا ہے۔ چنانچہ یہ Knowledge-based یا Example-based ترجمہ کہلاتا ہے جو مشینی سمجھ کا مثالوں یا مماثلتوں کی بنیاد پر استدلال (Case-based Reasoning Approach) سے کام لینے کا انداز ہے۔ مشینی ترجمے کا یہ طریقہ قواعدی ترجمے کی ضد ہے جو زبان کے ''سمجھائے'' گئے قواعد اور ثقافتی لگاوٹوں وغیرہ کو ''سمجھ'' کر اُن پر عمل کرتے ہوئے ترجمہ کرتا ہے، چنانچہ اسے کمپیوٹر سائنس کے سلینگ میں ایک بےعقل (Duffer) یا کھوتا کام (Donkey-work) کرنے والا ترجمہ کار کہتے ہیں یا اردو کے روزمرّے میں سفید اَن پڑھ۔

اِس ترجمہ کار کی تکنیکی صورت گری یوں ہوتی ہے کہ اِس کے پاس کمپیوٹر (یا کمپیوٹروں کے ایک سلسلے) میں ایک ڈیٹا بیس (یعنی ٹرانسلیشن میموری) رکھا ہوتا ہے جس میں پورے پورے جملوں/فقروں کا ترجمہ موجود ہوتا ہے۔ اسی موجود ''علم'' سے یہ ترجمہ کار آمدہ جملوں کا ترجمہ کرتا ہے۔ یہ ترجمہ کار اِس نظریے پر کام کرتا ہے کہ جس طرح انسان بھی ایک (لمبے) جملے کا ترجمہ پورے جملے کے سارے لسانی پہلوؤں کو سوچ کر یکلخت نہیں کرتے بلکہ پہلے جملے کو مناسب ٹکڑوں میں بانٹ کر ایک ایک ٹکڑے کا مناسبِ حال ترجمہ اپنے ذہن میں کر لیتے ہیں اور پھر اِن سب ترجمہ شدہ ٹکڑوں کو مطلوبہ زبان کی قواعد کے مطابق درست انداز میں ترتیب دے کر مطلوبہ زبان کا پورا جملہ بنا لیتے ہیں۔ بنیادی طور پر یہ ترجمہ کار سافٹ ویئر ذولسانی کارپسوں کی تربیت یافتہ (Trained) صورتیں ہوتے ہیں جن میں ہر جملہ اپنے ترجمے کے ساتھ ایک ایک جوڑے کی شکل میں رکھا ہوتا ہے۔ چنانچہ یہ کارپس اور Knowledge-base جن دو زبانوں کا ہو، مشینی ترجمے کا یہ طریقہ ان دونوں زبانوں میں ایک دوسری کا متن ترجمہ کرنے کی یکساں صلاحیت رکھتا ہے۔

2.2.1: مثالی ترجمہ کار کا طریقِ کار

جب ترجمے کے لیے کوئی متن اِس سافٹ ویئر میں داخل کیا جاتا ہے تو یہ سب سے پہلے اپنے پاس موجود جملوں سے اِس کا جملہ بہ جملہ موازنہ کرتا ہے اور دیکھتا ہے کہ آیا فل سٹاپ یا کسی اور ختمی علامت والا کوئی جملہ ایسا ہے جو اِس ماخذ جملے سے سو فیصد ملتا ہو۔ اگر ایسا کوئی جملہ مل جائے تو فبہا، اور اگر ایسا نہیں ہے تو ایک تکنیک استعمال کی جاتی ہے جسے تقابل (Concordance) کہتے ہیں۔ اِس میں تقابلی سطح (Concordance Level) طے کی جاتی ہے، کہ ماخذ جملہ مطلوبہ جملے سے کتنا میل کھاتا ہے۔ یہ سطح عام طور سے ۹۰٪ سے لے کر ۵۰٪ تک رکھی جاتی ہے (لیکن یہ کوئی ضروری نہیں)۔ یہ تقابل اِس امر کو روشنی میں لاتا ہے کہ ماخذ جملے کا کتنے فیصدی حصہ ٹرانسلیشن میموری سے فراہم کیا جا سکتا ہے۔ مثلاً داخل کیا گیا ایک ماخذ جملہ اگر یہ ہے:

Muhammad Ahmad is sleeping.

اور اِس وقت آپ نے تقابلی سطح ۸۰٪ رکھی ہوئی ہے۔ اب اگر ٹرانسلیشن میموری کے اندر کوئی جملہ ایسا رکھا ہے جس میں ''is sleeping'' آتا ہے اور اِس کا ترجمہ ''سو رہا ہے'' کیا ہوا ہے، اور فرض کیجیے کہ ٹرانسلیشن میموری کے اندر ''Muhammad Ahmad'' کا ترجمہ موجود نہیں

ہے، تو اب یہ سافٹ ویئر ماخذ جملے کے دو پارچے (Chunks) بنالے گا جن میں سے ایک پارچے کا ترجمہ موجود ہے اور دوسرے کا ناموجود۔ تو یوں ماخذ جملے کا پچاس فیصد ترجمہ کیا ہوا جملہ صارف کو ہمدست ہو جائے گا۔ اب اگر صارف حاصل شدہ جملے میں اِس ناموجود حصے کا ترجمہ خود سے کر کے جملے کو محفوظ (Save) کر لے تو یہ ترجمہ شدہ پورا جملہ ٹرانسلیشن میموری میں جا داخل ہوگا۔ یہ داخل کیا گیا ''علم''، یعنی نالج ہے جو اِس درج کردہ مثال میں ''Muhammad Ahmad'' ہے۔ چنانچہ آئندہ جب بھی کوئی ماخذ جملہ جس میں ''Muhammad Ahmad'' کا پارچہ موجود ہوا تو اُسے یہ ترجمہ مل جائے گا۔

یوں یہ ٹرانسلیشن میموری حجم میں بڑھتی چلی جاتی ہے اور زیادہ کارآمد ہوتی جاتی ہے۔

2.2.2: مثالی ترجمے کے مسائل اور اُن کا حل: شعبے کی زبان (Term-base)

لیکن اگر ٹرانسلیشن میموری کا حجم بہت زیادہ ہو جائے تو کمپیوٹر میں چھٹائی (Sorting) کے مسائل گھمبیر ہوتے جاتے ہیں۔ اِس کا لازمی نتیجہ مطلوبہ معلومات کے دیر سے ملنے یا بسا اوقات سافٹ ویئر کے جواب دے جانے (Crash) کی صورت میں نکلتا ہے۔ لہٰذا کوشش کر کے ٹرانسلیشن میموری کو مختلف حصوں میں بانٹ کر رکھا جاتا ہے تاکہ مجموعی حجم بھلے زیادہ ہو جائے، لیکن ایک وقت میں ایک حصہ ہی پراسیس کرنا پڑے۔

مثالی ترجمہ کار کا دوسرا مسئلہ الفاظ کے مرادی معنوں کا ہے۔ عام طور سے یہ ہوتا ہے کہ ایک لفظ ایک شعبے میں کچھ معنی رکھتا ہے اور دوسرے میں اور۔ مثلاً نسخہ کتاب کا بھی ہوتا ہے اور طبیب کا بھی۔ لفظِ لوٹا کو صرف اِملائی اعتبار سے دیکھیں تو یہ ایک برتن بھی ہے اور مصدر لوٹنا کا صیغۂ ماضی بھی۔ اِسی طرح انگریزی میں can کا معنیٰ سکنا بھی ہے اور ڈبہ بھی؛ table کا معنیٰ جدول بھی ہے اور میز بھی؛ اور trunk کا معنیٰ درخت کا تنا، پچھاڑی، صندوق اور نہ جانے کیا کیا ہے۔ مثالی ترجمہ کار میں درست یا حسبِ ضرورت معنی منتخب کرنے کا فیصلہ کرنے کی صلاحیت (Decision-making) نہیں ہوتی۔ مثلاً کسی وقت میں ایک ماخذ جملہ یہ داخل کیا گیا تھا:

I saw a mouse.

اور اِس کا ترجمہ ٹرانسلیشن میموری میں رکھا ہے: میں نے ایک چوہا دیکھا۔ اب اگر کسی نے یہ ماخذ جملہ داخل کیا:

I click the mouse button.

تو اِس کا ترجمہ مطلوبہ جملے کی صورت میں یوں آئے گا: میں چوہے کا بٹن دباتا ہوں۔ ظاہر ہے کہ یہ مطلوبہ جملہ نہیں ہے۔ مطلوبہ جملے میں mouse کا معنیٰ چوہا نہیں تھا۔ چنانچہ اِس قسم کے مسائل کے حل کے لیے ٹرانسلیشن میموری کے ''علم'' کو مختلف شعبوں میں تقسیم کر لیا جاتا ہے۔ اِس تقسیم کو تکنیکی زبان میں شعبے کی زبان (Term-base) کہتے ہیں۔ یہ ٹرم بیس، جو Knowledge-base ہی کی ذیلی شاخیں ہوتے ہیں، ریاضی، طبیعیات، کمپیوٹر سائنس، ادب، صحافت، موسمیاتی رپورٹیں، وغیرہ وغیرہ شعبوں کے ہوتے ہیں، اور اِن میں سے ہر ایک میں ترامیم اور اضافے بھی کیے جاتے رہتے ہیں۔ جس وقت جس موضوع کے متن کا ترجمہ مقصود ہو بس اُسی کا ٹرم بیس ٹرانسلیشن میموری میں لوڈ کر لیا جائے، ٹرانسلیشن میموری کا حجم بھی چھیج جائے گا اور مطلوبہ متن کے جملے بھی ضرورت سے قریب قریب ہوں گے کیونکہ اِن میں محدود تر الفاظ میں سے اپنے لیے مناسب لفظ منتخب کیے گئے ہوں گے۔ اِس سے یہ بات بھی معلوم ہوگئی کہ حکومتی اور قانونی دستاویزات کا، یعنی ایسے متون جو اصولی اور قواعدی زبان (Formal/Formulaic Language) میں لکھے گئے ہوں، ترجمہ زیادہ بہتر انداز میں اور جلد ہو سکتا

ہے بنسبت عام مکالموں یا غیر معیاری یا آزاد متون کے۔

چنانچہ ٹرم بیس بنانے کے دو فائدے ہیں: ایک تو یہ کہ کسی مخصوص شعبۂ علم اور اُس کی ذیلی شاخوں کی لفظیات (Lexis) بن گئی، جو کہیں بھی استعمال ہو سکتی ہے، کہیں بھی بھیجی جا سکتی ہے (Portability) اور جب چاہے بہتر (Improve) کی جا سکتی ہے۔ اور دوسرا فائدہ جو ضمناً حاصل ہو گیا، یہ ہے کہ ٹرانسلیشن میموری کا مجموعی حجم کم ہو گیا۔ اِس سے مشین کے پراسیسنگ ٹائم میں کمی ہوتی ہے اور نتیجۂ صارف کو مطلوبہ معلومات بھی جلد مل جاتی ہیں۔

2.2.3: شعبے کی زبان استعمال نہ کرنے کے نقصانات

یہ درست ہے کہ دورِ حاضر میں کمپیوٹر کی رفتار اِتنی زیادہ اور میموری اِتنی ارزاں ہو گئی ہے کہ پراسیسنگ ٹائم کوئی بڑا مسئلہ نہیں رہا، تاہم اگر ٹرم بیس نہ بنائے جائیں یا اِن کے بغیر کام کیا جائے تو کئی نقصانات ہوں گے۔ ذیل میں اردو ادب کے حوالے سے چند ایک مثالیں پیش کی جاتی ہیں تا کہ بات واضح ہو سکے۔

اگر غالب کے اشعار کا ترجمہ کرنا مقصود ہو اور غالب کی شاعری کا ٹرم بیس نہ بنایا گیا ہو تو مندرجۂ ذیل اشعار کے تراجم یوں ملیں گے:

تھی خبر گرم کہ غالب کے اُڑیں گے پرزے

دیکھنے ہم بھی گئے تھے پہ تماشا نہ ہوا

The news was hot that spare-parts of Ghalib will be flown

We also went to see but there was no exhibition

[مکرر]

دلِ ناداں تجھے ہوا کیا ہے

آخر اس درد کی دوا کیا ہے

O'stupid heart! what has happened to you

At last what is the medicine of this pain

یا اگر مندرجۂ ذیل اردو ماخذ جملہ ڈالا جائے: وہ گیا، ایسا گیا کہ بس گیا ہی گیا۔ تو اِس کا انگریزی ترجمہ یہ ملتا ہے:

He went, such went that went so went.

تفنن برطرف، اَدبی متن کا ترجمہ ویسے بھی خالہ جی کا گھر نہیں کیونکہ اَدبی متون (نظم/نثر) میں الفاظ اپنے لفظی/لغوی معنٰی کی بجائے استعاراتی معنوں ہی میں استعمال ہوتے ہیں۔ کمپیوٹر کی معاونت سے کیے گئے شاعری کے تراجم سے مزاح البتہ ضرور پیدا کیا جا سکتا ہے۔

2.2.4: مثالی ترجمے کی بنیاد پر کام کرنے والے مقبول سافٹ ویئر: ایک تعارف

ٹرانسلیشن میموری کی بنیاد پر کام کرنے والا مقبول ترین سافٹ ویئر TRADOS ہے۔ یہ لفظ Translator for DOS کا مخفف ہے۔ یہ پروگرام مائیکروسافٹ کی عام صارف فائلوں یعنی doc.، ppt.، xls. اور rtf. وغیرہ فارمیٹ اور اِس کے ساتھ ساتھ XML فارمیٹ میں لکھے گئے متن کو بسہولت ترجمہ کر سکتا ہے۔ اِس کے پاس بے شمار ٹرانسلیشن میموری ہے اور یہ پوری فائل کو بحیثیتِ ماخذ متن داخل

کرتے ہی فائل کا بڑا حصہ ترجمہ کر دیتا ہے، جب کہ ناموجود جملوں اور الفاظ کو الگ کر کے دکھا دیتا ہے اور صارف سے مطالبہ کرتا ہے کہ وہ اِن لفظوں اور جملوں کا ترجمہ کر کے فائل کو محفوظ کر لے۔ اس فائل کو محفوظ کرنے کے ساتھ ہی یہ مواد ٹرانسلیشن میموری میں داخل ہو جاتا ہے اور آئندہ ماخذ جملوں کے ترجمے کے لیے استعمال ہوتا ہے۔

یہ سافٹ ویئر، بلکہ اس کی طرح کے سارے ہی سافٹ ویئر، دائیں سے بائیں چلنے والی تقریباً تمام ہی زبانوں کے ماخذ متون کو مختلف زبانوں میں ترجمہ کر سکتے ہیں۔ اردو کے حساب سے دیکھیں تو اِن دونوں سافٹ ویئروں کی ایک خاص خوبی یہ ہے کہ یہ انگریزی-اردو اور اردو-انگریزی ترجمہ کر سکتے ہیں۔ سافٹ ویئر چلاتے وقت بتانا پڑتا ہے کہ ماخذ زبان کون سی ہے اور مطلوبہ زبان کون سی۔

تکنیکی زبان میں اس نوع کے ترجمہ کار سافٹ ویئروں کو ذاتی ترجمہ کار (Personal Translator) بھی کہتے ہیں۔

2.3: شماریات (Statistics) کی بنیاد پر کیا گیا ترجمہ (SMT)

شماریاتی مشینی ترجمہ ایسی تکنیک ہے جس میں ایسے شماریاتی ماڈل کی بنیاد ترجمہ کیا جاتا ہے جس کے پیرامیٹرز دولسانی کارپسوں کے تجزیے سے اخذ کیے گئے ہوں۔ ترجمے کی یہ تکنیک قواعدی اور مثالی ترجمہ کاری کے پہلو بہ پہلو کام کرتی ہے۔

شماریاتی مشینی ترجمے کا نظریہ ۱۹۴۹ء میں Warren Weaver نے امریکی ماہر برقیات شینن (Claude Elwood Shannon) کے انفارمیشن تھیوری کے نظریے کو لاگو کرتے ہوئے پیش کیا تھا۔ ۱۹۹۱ء میں یہ خیال IBM کے سائنسدانوں نے دوبارہ سے پیش کیا اور اب اس تکنیک نے مشینی ترجمے کی تحقیقات میں نئی سرگرمی پیدا کر دی ہے۔ مشینی ترجمے کے میدان میں سب سے زیادہ تحقیقات آج کل اِسی طریقے پر ہو رہی ہیں۔

شماریاتی مشینی ترجمے کی تکنیک کا کم و بیش یہ ہے کہ اوّلاً یہ گمان کر لیا جاتا ہے کہ مشین کے پاس موجود ڈیٹا ذخائر (Repositories) کا کوئی نہ کوئی پارچہ (String) ماخذ متن کے کسی پارچے کا ترجمہ ہو سکتا ہے۔ اب سافٹ ویئر اس متن کو مختلف انداز میں جانچتا ہے اور اُس پارچے کو سامنے لے آتا ہے جس کا ماخذ پارچے کا ترجمہ ہونے کا امکان (Probability) سب سے زیادہ ہو۔ یہ ترجمہ کار کام کو دو سطحوں میں تقسیم کر کے کرتا ہے: پہلا حصہ ترجمے کا ماڈل (Translation Model) ہے اور دوسرا لسانی ماڈل (Language Model)۔ ترجمے کا ماڈل ماخذ متن کے ہر ہر لفظ کے لیے ممکن ترجمے کا پارچہ مطلوبہ زبان میں ڈھونڈ لیتا ہے، جب کہ لسانی ماڈل اِن ڈھونڈے گئے پارچوں کو مطلوبہ زبان کے قواعد کے مطابق ترتیب دے دیتا ہے۔ بہترین ترجمہ وہ ہوگا جس میں یہ دونوں امکانات زیادہ سے زیادہ ہوں۔ یہ تکنیک اِس لیے پسندیدہ ہے کہ یہ مسئلے کو دو حصوں میں تقسیم کر کے کام کرتی ہے۔ اِس تکنیک کو اگر بے محابا نافذ کیا جائے تو ماخذ متن کے سبھی پارچے تلاش اور امکان کے اِس عمل سے مکمل طور پر گزر جاتے ہیں۔ تلاش کے اِس عمل کی کارکردگی میں بہتری لانا مشینی ترجمے کو Decode کرنے والے سافٹ ویئر کا کام ہے جو بیرونی پارچوں، خود ساختہ معاونین اور دوسرے طریقوں سے تلاش کے عمل کو نہ صرف بے ضرورت پھیلنے سے روکتا ہے بلکہ اِس کے معیار کو بھی اطمینان بخش رکھتا ہے۔ معیار اور وقت کے درمیان یہ سمجھوتہ آواز پر تحقیقات (آواز سے متن/متن سے آواز) میں بھی ملتا ہے۔

اب چونکہ ترجمے کے سسٹم سبھی زبانوں کے سبھی پارچوں اور اُن کے ترجموں کو اپنے پاس جمع نہیں رکھ سکتے، اِس لیے ہر دستاویز جملہ بہ جملہ ترجمہ ہوتی ہے۔ لیکن یہ بھی کافی نہیں ہے۔ لسانی ماڈلوں کو بھی عام طور سے نپے تلے اندازوں کے ساتھ مقید رکھا جاتا ہے اور

اسی طرح سے ترجمے کے ماڈلوں کو بھی، لیکن مختلف زبانوں میں جملوں کی لمبائی کے فرق اور الفاظ کی ترتیب کے اختلاف کی وجہ سے کچھ اضافی پیچیدگیاں ہیں۔

مشینی ترجمے کی یہ تکنیک ذولسانی کارپسوں پر انحصار کرتی ہے جیسے مثلاً Canadian Hansard Corpus، کنیڈین پارلیمان کا انگریزی-فرانسیسی ریکارڈ اور یوروپین پارلیمان کا ریکارڈ EUROPAL۔ جہاں ایسے کارپس مہیا ہیں وہاں ملتے جلتے متون کے ترجمے کے نہایت شاندار نتائج حاصل ہوتے ہیں۔ لیکن ایسے کارپس ابھی تک بہت ہی کم تعداد میں ہیں۔ شماریاتی مشینی ترجمے کا پہلا سافٹ ویئر IBM کا تیار کردہ CANDIDE تھا۔ گوگل والوں نے بہت سالوں تک Systran استعمال کیا لیکن بالآخر ستمبر ۲۰۰۷ء میں شماریاتی مشینی ترجمے پر منتقل ہو گئے۔ ابھی کچھ عرصہ پہلے اُنھوں نے تقریباً ۲۰۰ بلین الفاظ پر مشتمل اقوامِ متحدہ کی دستاویزات اپنے سسٹم میں بطورِ تربیتی مواد (Training Data) ڈال کر اپنی ترجمے کی سہولیات کو بہتر کیا ہے۔

2.3.1: شماریاتی ترجمے کی اقسام

شماریاتی مشینی ترجمے کے ماڈل ابتدا میں لفظ کی بنیاد پر تھے۔ پھر جب اِن ماڈلوں سے فقروں کی بنیاد پر ترجمہ کیا جانے لگا تو یہ ایک واضح تبدیلی تھی۔ تازہ ترتحقیقات میں جملے کے نحوی اور نیم نحوی اجزا (Quasi-syntactic structures) بھی شامل کیے جانے لگے ہیں۔

۱: لفظ کی بنیاد پر ترجمہ (Word-based translation): کچھ انسانی زبانوں میں ترجمے کی بنیادی اکائی لفظ ہوتا ہے۔ ترجمہ شدہ جملے میں لفظوں کی تعداد مختلف ہوتی ہے جس کی وجہ مرکبات، صرفی شکلیں اور محاورات ہوتے ہیں۔ ترجمہ شدہ الفاظ کی قطار (Sequence) کی لمبائی کو لفظی بڑھوتری (Fertility) کہتے ہیں جس سے یہ تناسب معلوم ہوتا ہے کہ ایک مقامی لفظ دوسری زبان کے کتنے الفاظ پیدا کرتا ہے۔ انفارمیشن تھیوری یہ بات لازماً فرض کر لیتی ہے کہ ہر لفظ ایک ہی خیال کو محیط ہوتا ہے۔ جیسے مثلاً لفظِ نسخہ کا ایک ہی معنیٰ ہوگا، یعنی کتاب یا طبیب کی تجویز کردہ دوا۔

لفظ کی بنیاد پر ترجمہ کرنے والے سادہ ترجمہ کار سافٹ ویئر لفظی بڑھوتری کی مختلف شرح رکھنے والی زبانوں کا آپس میں ترجمہ نہیں کر سکتے۔ اِن سادہ ترجمہ کاروں کو بڑی آسانی کے ساتھ زیادہ الفاظ پیدا کرنے والی زبان کا ترجمہ کرنے سے نمٹنے میں لگایا جا سکتا ہے۔ لیکن یہ کام یک رخا ہوتا ہے: ایک لفظ زیادہ لفظوں کے ترجمے کے لیے مخصوص ہو سکتا ہے، اِس کا الٹ البتہ ممکن نہیں ہوتا۔ یعنی اگر ہم اردو سے انگریزی ترجمہ کر رہے ہیں تو ہم اردو کے ایک لفظ کو انگریزی کے جتنے چاہیں الفاظ کے لیے مخصوص کر دیں، یا بعض اوقات کسی ایک لفظ کے لیے بھی نہیں، لیکن ہمارے پاس اس کا کوئی طریقہ نہیں ہے کہ انگریزی کے دو لفظوں کو اردو کا ایک ہی لفظ پیدا کرنے کے لیے مخصوص کر دیں۔ چنانچہ ہم اردو لفظ آیئے کے لیے تو یہ طے کر سکتے ہیں کہ ہمیں اس کا ترجمہ ہمیشہ Come here ملے، لیکن Come here بھی ہمیشہ آیئے ترجمہ دے، یہ ممکن نہیں ہے۔ بالکل اسی طرح اردو جملے: ابا جان آئے۔ کے لیے یہ مخصوص کر کے کہ اس کا ترجمہ ہمیشہ Father came ملے، پروگرام کو یہ سمجھایا جا سکتا ہے کہ

ترکیب: اباجان میں لفظِ جان کا ترجمہ نہیں کرنا ہے۔ وغیرہ۔ ++GIZA ایک مفت ترجمہ کرنے والا سافٹ ویئر ہے جو شماریاتی ترجمہ تکنیک پر کام کرتا ہے۔

لفظ کی بنیاد پر ترجمہ کرنے والے سافٹ ویئروں کے مقابلے میں آج کل فقروں کی بنیاد پر کام کرنے والے سافٹ ویئر زیادہ عام ہیں۔ فقروں کی بنیاد پر کام کرنے والے سافٹ ویئر کارپس سے ہم آہنگی کے لیے ابھی بھی زیادہ تر ++GIZA کو ہی استعمال کررہے ہیں۔ الفاظ کی یہ ہم آہنگی متن میں سے فقروں کو الگ کرنے اور نحوی اصول اخذ کرنے کے لیے استعمال ہوتی ہے۔

ایسے متن جن میں نچلی سطر میں ترجمہ چل رہا ہو(Bi-text)، ابھی تک یہ ایک مسئلہ ہے جس پر بحث ہوتی رہتی ہے کہ کون سا لفظ کون سے لفظ کا ترجمہ ہے۔ جہاں تک چھپے ہوئے متن کی بات ہے، پاکستان میں پروفیسر عبدالرحمٰن طاہر کے تین رنگوں میں کیے ہوئے ترجمۂ قرآنِ پاک (عربی-اردو) کی حد تک یہ کام ہوا ہے جس میں لفظ اور اُس کا معنٰی ایک ہی رنگ میں دے یہ مسئلہ سلجھانے کی کوشش کی گئی ہے، جو بہر حال کامیاب رہی ہے۔[۵] لیکن ظاہر ہے کہ اِس کوشش کا مشینی ترجمے کی کسی بھی قسم سے کوئی واسطہ نہیں۔ آن لائن متن کے کیے ہوئے ترجمے جو نچلی سطر میں چل رہے ہوں، اردو میں ابھی بالکل بھی خودکار نہیں ہوئے۔ پاکستان کے سبھی نیوز چینل یہ فیتے یا پٹی (Ticker) چلانے کے لیے اردو کو اِن پیج (Inpage) میں الگ سے ٹائپ کرواکے لگاتے ہیں۔ پٹی چلانے کے لیے سہ لسانی سافٹ ویئر بھی ملتے ہیں۔[۶] ++GIZA ایسا چھا گیا کہ اس کے سامنے اِس نوعیت کے کسی بھی سافٹ ویئر کا چراغ نہ جلا۔ آج اِس کی طرح کے بہت سے آن لائن پروگرام کام کررہے ہیں۔ اردو میں بھی اِس قسم کے (لیکن بہت ہی سادہ) پروگرام ملتے ہیں۔

۲: فقرے کی بنیاد پر ترجمہ (Phrase-based translation): شماریاتی مشینی تراجم کے ماڈل میں فقروں کی بنیاد پر کیے جانے والے ترجمے میں مقصد یہ ہوتا ہے کہ لفظوں کی پوری قطاروں کو، جن کی لمبائی مختلف بھی ہوسکتی ہے، ترجمہ کر کے لفظ کی بنیاد پر ترجمہ کرنے والے سافٹ ویئروں کی وجہ سے لگی پابندیوں کو کم کیا جائے۔ کمپیوٹر کی زبان میں لفظوں کی قطار کو بلاک یا فقرہ کہتے ہیں؛ یہ وہ فقرہ نہیں ہوتا جسے لسانیات فقرہ تسلیم کرتی ہے بلکہ یہ کارپسوں پر مختلف شماریاتی طریقے لاگو کرنے سے ملنے والی لفظوں کی قطار ہوتی ہے۔ تحقیق سے یہ بات سامنے آئی ہے کہ فقروں کو محض لسانی طور پر درست فقروں تک محدود کردینا ترجمے کے معیار کو گھٹا دیتا ہے۔

۳: نحوی اور نیم نحوی اجزا کا ترجمہ (Syntactic & Quasi-syntactic structures): شماریاتی مشینی تراجم کے ماڈل پر نحوی اجزا کا ترجمہ ابھی بالکل نئی چیز ہے۔ اِس میں لفظوں اور فقروں کے ساتھ ساتھ مطلوبہ زبان میں نحو کے اعتبار سے درست ترجمہ کرنے کی تحقیقات ہورہی ہیں۔

2.3.2: شماریاتی ترجمے کو درپیش چیلنج

شماریاتی مشینی ترجمے کو جن چیلنجوں کا سامنا ہے اُن میں مندرجۂ ذیل اہم ہیں:

۱: مرکبات (Compound Words): مشین کے لیے یہ طے کرنا مشکل ہے کہ مثلاً مرکب: دلبر داشتہ میں ترکیب کی اصل کیا ہے: آیا یہ دلبر+داشتہ ہے، دل+بر+داشتہ ہے، یا دل+بر داشتہ۔ یا مثلاً اِس شرارت سے بھرے فقرے: داشتہ آید بکار میں کہیں کار سے مراد car تو نہیں ہے۔

۲: محاورات (Idioms): محاورات کو محاورات سے ترجمہ کرنا ممکن نہیں ہوتا۔ اور اگر کیا جائے تو معنوں میں بعدالمشر قین ہو جاتا ہے۔ محاورات وہ لسانی پیکر ہیں جن کے ترجمے میں انسانی ہاتھ کی کارفرمائی کے بغیر تسلی بخش کام ہو ہی نہیں ہوسکتا۔ صرف اردو کی ایک مثال لیجیے کہ ایک محاورہ ہے: آڑے آنا۔ اِس کا معنٰی مدد کرنا بھی ہے اور روکنا بھی۔ مشین کو اِن دونوں یکسر مختلف معنی میں امتیاز کر سکنے جتنی مصنوعی ذہانت دینا کم سے کم فوری طور پر ممکن نہیں ہے۔

۳: تصریف (Morphology): صرفی قواعد تقریباً ہر زبان میں بہت رنگارنگی کے حامل ہوتے ہیں۔ مثلاً اردو میں اسمِ آلہ بنانے کی یہ صورتیں دیکھیے جو بغیر کسی گہرے مطالعے کے گھر میں سے صرف سونے کے کمرے اور باورچی خانے میں ایک سرسری نظر ڈالنے سے سامنے آئی ہیں (یاد رہے کہ یہاں صرف ایسے اَسمائے آلہ ذکر کیے جا رہے ہیں جو کسی نہ کسی لفظ سے مشتق ہیں):

سرمچو (سرمہ لگانے کا آلہ)،

نُہرنی (ناخن تراشنے کا آلہ)،

موچنا (بال اُکھیڑنے کا آلہ، اصلاً فارسی: مُوئے چینہ)،

تَلائی (تلے بچھانے کا ہلکا گدّا)،

تپائی (تین پایوں والا چیزیں دھرنے کا طاق)،

پنکھا / پنکھی (ہوا دینے کا آلہ)،

پیڑھی (بیٹھنے کی چوکی)،

چکلا (آٹے کا پیڑا پھیلانے یا اناج دَلنے کا گول پتھر)،

بیلن / بیلنا (روٹی یا پوری کو بڑھانے کا لکڑی یا پتھر کا گول ڈنڈا نما آلہ)،

کڑچھا / کڑچھی (کڑاہی میں پھیرنے کا آلہ)،

چمٹا / چمٹی (چیزیں پکڑنے کا لکڑی یا لوہے کا دو پھلوں کا آلہ)،

مدھانی (دودھ بلونے / رڑکنے یا دہی کو پھینٹنے / متھنے کا آلہ، جسے لغت میں متھانی بھی لکھا ہے)،

رکابی (سالن رکھنے کی تھالی)،

کچوکنی (قیمہ بنانے کا آلہ یا سِل)،

گھوٹنا (کونڈی میں چیزیں کوٹنے پیسنے کا ڈنڈا)،

کدو کش (سبزیوں کے چھوٹے چھوٹے ٹکڑے کرنے کا آلہ)،

دھونکنی (آگ کو پھونک مارکر بھڑکانے کا آلہ)،

پھونکی / پُھکنی / پُھکنا (آگ کو پھونک مار کر بھڑکانے کا آلہ)۔

یہ اسمائے آلہ جن الفاظ سے بنے ہیں وہ بالترتیب یہ ہیں: سرمہ، نہ (ناخن)، تلا، پنکھ، پاٹ بمعنٰی تختہ یا پٹ بمعنٰی ران کا موٹا حصہ سے مشتق، چکلان (وسعت، چوڑائی)، بیلنا، کڑاہی، چمٹنا، متھنا، رکاب، کچوکا (مصدر کچونا سے)، گھوٹنا، کش/کشیدن (فارسی)، دھواں، پھونکنا۔ معلوم ہوا کہ اردو میں اسمِ آلہ بنانے کی کوئی ایک صورت نہیں ہے: -چُو، -رنی، -ائی، -ا/ ی، -چھا/ چھی، -انی، -نا/ نی وغیرہ لاحقی صورتیں صرف اِنہی الفاظ سے نکل آئی ہیں۔ تحقیق سے اِن میں اضافہ بالکل ممکن ہے۔ لسانی خوبصورتی اور کم سے کم الفاظ میں زیادہ سے زیادہ مطالب کی سمائی کی انتہا دیکھیے کہ گھوٹنا اور بیلنا میں تو مصدر ہی بطورِ اسمِ آلہ کام دے رہا ہے۔ اِس کے بالمقابل انگریزی میں دیکھیے تو اسمِ آلہ کی عموماً صرف ایک ہی صورت (-er) ملتی ہے۔ اوپر کی مثالوں میں سے آخری لفظ — پھونکنا — کے انگریزی ترجمے سے اسمِ آلہ بنایئے تو لفظ Blow کے آخر میں صرف er- لگانا پڑتا ہے۔ اللہ اللہ خیر سلاّ۔

۴: الفاظ کی مختلف ترتیبیں (Different word orders): زبانوں میں لفظوں کی ترتیب مختلف ہوتی ہے۔ فاعل (Subject)، فعل (Verb) اور مفعول (Object) میں کی گئی بنیادی تقسیم کی بنا پر کچھ زبانوں کو SVO یا VSO وغیرہ کہا جاتا ہے۔ لفظوں کی ترتیب میں ہیر پھیر ان صورتوں کے علاوہ بھی ہوتا ہے جیسے مثلاً جب اسماء یا صفات کو بطورِ فعل استعمال کیا جائے (جیسے: حکومت نے بینکوں کو قومیا لیا۔ [قوم] ساقی فاروقی نے مضمون میں تنقید کو نرما دیا۔[نرم] اقبال کی شاعری نے قوم کے جذبات کو گرما دیا۔[گرم] فوج کی شکست نے اہلِ وطن کو شرما دیا۔ [شرم]) یا جب کبھی ایک ہی مجموعۂ الفاظ کو جملۂ خبریہ اور جملۂ استفہامیہ کی صورت میں ادا کیا جائے (جیسے: وہ آئے تھے؟/ وہ آئے تھے۔) وغیرہ۔

آواز کی پہچان اور ساتھ ساتھ ترجمے کا متن لکھنے کی صورت میں لفظوں کی قطاروں کے پارچے بنا کر ان کو بلاکوں کی صورت میں درست جگہ پر رکھا جا سکتا ہے۔ ہر جملہ ہر وقت اور ہر موقع پر ایک ہی معنی دے، یہ کوئی ضروری نہیں۔ مثال لیجیے کہ انگریزی جملہ: I see کبھی میں دیکھتا/ دیکھتی ہوں۔ کا معنٰی دیتا ہے، کبھی اب سمجھا/ سمجھی، اور کبھی خوب/ بہت اچھے۔ وغیرہ۔ بالکل اِسی طرح سے اردو میں بھی ہے کہ مثلاً فقرے: بڑی مہربانی۔ یا اچھی کہی۔ وغیرہ موقع کے لحاظ سے کئی مفہوم دیتے ہیں جن میں سے کچھ آپس میں بالکل متضاد معنٰی رکھتے ہیں۔ وغیرہ۔

اصولاً شماریاتی مشینی ترجمے میں مشینی ترجمہ کار تو لفظوں کے چھوٹے چھوٹے مجموعوں کے صرف ترجمے ہی سے

سروکار رکھتا ہے، ترجمہ شدہ الفاظ کو درست ترتیب میں لگانا تو پروگرام ڈیزائنر کے سوچنے کا کام ہے۔ ترجمے کے ماڈلوں کو ترتیب دینے کی ایسی کوششیں کی گئی ہیں جن میں اوپر متن اور نیچے ترجمے کی صورت میں ہر ترجمہ شدہ پارچے کو درست جگہ پر لگایا جاسکے۔ درست جگہ پر لگنے والے ایسے پارچوں کو لسانی ماڈل میں جگہ دی جاسکتی ہے اور بہترین کام کرنے والی صورت کو اپنایا جاسکتا ہے۔

۵: نحو (Syntax)۔ زبانوں میں جملوں کی نحو میں باہم بہت فرق ہے۔ مثلاً اردو SOV ہے اور انگریزی SVO۔

۶: ناموجود الفاظ (Out of Vocabulary Words): شماریاتی مشینی تراجم کے سسٹم لفظوں کی ایسی بہت سی صورتوں اور فقروں کو الگ تھلگ محفوظ کر لیتے ہیں جو اُنھیں تربیتی مواد (Training Data) میں دستیاب نہ ہوئے ہوں۔ لفظوں کی یہ عدم دستیابی تربیتی مواد کی کمی کی وجہ سے بھی ہوسکتی ہے، اُس دائرۂ عمل کے تبدیل ہوجانے کی وجہ سے بھی ہوسکتی ہے جس کے لیے سسٹم استعمال کیا جارہا ہو، اور الفاظ کے اِملاء اور تصریفی صورتوں کے بدل جانے کی وجہ سے بھی۔ یاد رہے کہ الفاظ کے بالائے لغت معنی (Ultra-dictionary meaning) بالکل مختلف عنوان ہے۔

2.3.3: شماریاتی ترجمے کے فائدے

شماریاتی ترجمے کے وہ فوائد جن کا ذکر مشینی ترجمے کے روایتی مثالی نمونوں (Paradigms) میں نمایاں طور سے کیا جاتا ہے:

۱: وسائل کا بہتر استعمال:

الف: انسانی زبانیں بڑی مقدار میں مشین ریڈایبل حالت میں ملتی ہیں۔

ب: شماریاتی مشینی ترجمے کے سسٹم عام طور سے کسی بھی دو زبانوں کے لیے مخصوص نہیں ہوتے۔

ج: قواعدی ترجمہ کار سسٹم کو دستی بنائے ہوئے لسانیاتی اصولوں کی ضرورت ہوتی ہے، جو مہنگا بھی پڑتا ہے اور جسے دوسری زبانوں کے لیے عمومی طور پر استعمال بھی نہیں کیا جاسکتا۔ قواعدی ترجمہ کار زبان سے ناوابستہ ہو ہی نہیں سکتا، اور اسی جانبداری کی وجہ سے ہر موقع پر استعمال بھی نہیں کیا جاسکتا۔

۲: اِس تکنیک سے بہتر یعنی انسانی زبان کے قریب قریب (Natural) ترجمہ حاصل ہوتا ہے۔

2.4: دوغلا مشینی ترجمہ (Hybrid MT)

دوغلا مشینی ترجمہ دراصل قواعدی اور شماریاتی طریقہائے مشینی تراجم کی قوتوں اور خصوصیتوں کو یکجا استعمال کرنے کا نام ہے۔ مشینی ترجمے کی سہولت فراہم کرنے والی کچھ کمپنیاں مثلاً Asia Online اور Systran قواعدی اور شماریاتی دونوں طریقوں کو ملا کر استعمال کرنے کا دعویٰ کر رہی ہیں۔ ترجمے کا یہ طریق کار کئی انداز میں کام کرتا ہے:

۱: قواعد کے بعد شماریات کا استعمال: اِس تکنیک میں قواعدی انجن کے ذریعے ترجمہ کرنے کے بعد اِس آؤٹ پُٹ کو سنوارنے بنانے کا کام شماریاتی ترجمہ کار سے لیا جاتا ہے۔

۲: شماریات کو قواعد کی رہنمائی میں استعمال کرنا: اِس تکنیک میں قواعد کو پہلے سے اِس انداز میں استعمال کیا جاتا ہے

کہ وہ شماریاتی انجن کو بہتر رہنمائی دے سکیں۔ شماریاتی طریقے سے ترجمہ ہوچکنے کے بعد دوبارہ سے قواعدی انجن کے ذریعے آؤٹ پُٹ کو فطری زبان میں ڈھالنے کا کام لیا جاتا ہے۔ ظاہراً یہ طریقہ گھماؤ پھیر والا ہے لیکن کام کے اعتبار سے یہ بہت طاقتور تکنیک ہے جس کے ذریعے ترجمہ کاری کی ہر سطح کو بآسانی اپنی ضرورت کے مطابق چلایا جاسکتا ہے۔

2.5: مشینی ترجمے کی جچائی (Evaluation)

مشینی تراجم کی کارکردگی کو جانچنے کے کئی ذریعے ہیں۔ ترجمے کے معیار کو جانچنے کا سب سے پرانا طریقہ ''انسانی ججوں'' کا ہے۔ انسانی تشخیص میں وقت اگرچہ زیادہ لگتا ہے لیکن قواعدی اور شماریاتی طریقوں سے کیے گئے مشینی ترجموں کو جانچنے کا یہی طریقہ سب سے زیادہ قابلِ اعتماد ہے۔ ترجمے کی جچائی کے مشینی ذرائع میں BLEU، NIST اور METEOR وغیرہ زیادہ معروف ہیں۔[۷]

مشینی ترجمے کو انسانی نگاہ سے گزارے بغیر استعمال کر لینا اِس حقیقت سے پہلوتہی ہے کہ انسانی گفتگو موقع کی مناسبت سے ہوتی ہے، اور جس میں سمجھنے سمجھانے کی غرض سے بسا اوقات معیاری لسانی ساختیات (Structuralism) سے بہت فاصلہ رکھنا ضروری ہوجاتا ہے۔ یہ درست ہے کہ انسان کے کیے ہوئے ترجمے میں بھی بعض اوقات غلطیاں در آتی ہیں۔ چنانچہ اِس امر کو یقینی بنانے کے لیے کہ مشین سے کیا ہوا ترجمہ لوگوں کے لیے قابلِ استعمال ہے اور قابلِ اشاعت بھی، یہ ضروری ہے کہ اِسے زبان کا جاننے والا کوئی آدمی دیکھ لے اور ایڈٹ کردے۔ علاوہ ازیں یہ بھی ضروری ہے کہ متن کو ترجمے کے لیے مشین کے سپرد کرنے سے پہلے بھی اِسے مناسب سی ایڈیٹنگ کے تمہیدی عمل سے گزار لیا جائے تاکہ مشینی ترجمہ کار سافٹ ویئر سے حاصل ہونے والا آؤٹ پُٹ الفاظ کا محض جھاڑ جھنکار نہ ہو۔

لغت کی بنیاد پر کام کرنے والے ترجمہ کار سافٹ ویئر کچھ قسم کے متون کا، جیسے کمپنیوں کی مصنوعات کی خصوصیات وغیرہ جو انتہائی سنبھلی ہوئی زبان میں تحریر کی جاتی ہیں، اطمینان بخش ترجمہ فراہم کرتے ہیں جنھیں اشاعت سے پہلے انسانی نگاہ سے گزارنے کی زیادہ ضرورت نہیں ہوتی۔[۳]

3: مشینی ترجمے کا سب سے بڑا عمومی مسئلہ: معنی میں ابہام

ایسے الفاظ کے جن کے مفہوم و معنٰی میں ابہام پایا جاتا ہے، درست مطلب کو پانے کے لیے مشینی ترجمہ کار کو تربیت دینا ایک دیرینہ مسئلہ ہے۔ یہ مسئلہ سب سے پہلے ۱۹۵۰ء کی دہائی میں Yehoshua Bar-Hillel نے اُٹھایا تھا۔ اُس نے کہا تھا کہ ایک ''عالمگیر انسائیکلوپیڈیا'' کے بغیر مشین کبھی ایک لفظ کے دو یا زائد معنوں میں سے سیاق و سباق کے اعتبار سے درست معنٰی میں امتیاز نہیں کرسکتی۔[۸] آج اِس مسئلے کے بہت سے حل مہیا ہیں جو بنیادی طور پر دو طریقوں سے کام کرتے ہیں: سطحی (Shallow) اور گہرا (Deep)۔

سطحی طریقوں سے کام کرنے میں متن کے مفہوم کو جاننے کی ضرورت نہیں ہوتی، مبہم لفظ کے اِردگرد کے الفاظ پر صرف چند شماریاتی اصول لگانا پڑتے ہیں۔ گہرے طریقوں سے کام کرنے میں لفظ کے بارے میں بہت سی معلومات چاہیے ہوتی ہیں۔ ابھی تک سطحی طریقے ہی زیادہ کامیاب رہے ہیں۔

زبان کے لہجوں کا اختلاف (Dialects) بھی بہت اہمیت رکھتا ہے۔ ایک ہی لفظ صرف صوتی شخصیت ہی میں نہیں بلکہ ترکیب و محاورہ میں بھی (مثلاً) امریکی انگریزی میں مختلف مفہوم رکھتا ہے اور برطانوی، آئرستانی، آسٹریلیائی، سنگاپوری، وغیرہ وغیرہ لہجوں میں مختلف۔

لہٰذا کسی بھی زبان سے انگریزی میں ترجمہ کرتے وقت اِن سب لہجوں کے اختلاف کو ملحوظ رکھنا ہوتا ہے۔ کوئی بھی زندہ زبان ایسی نہ ہوگی جس کے کئی لہجے نہ ہوں۔ ؎

جوگی کس سے بولے، دکھڑے من کے کس سے کھولے

بارہ کوس پہ بولی بدلے، تیرہ کوس پہ ریت

چنانچہ ترجمے کے مسائل کی کنہ تک صرف ایسے ہی لوگ پہنچ پاتے ہیں جو ایک زبان میں اعلیٰ تعلیمی قابلیت رکھنے کے ساتھ ساتھ کم سے کم ایک بیرونی زبان کو بحیثیتِ زبان پڑھانے یا اُس کے عملی استعمال کا طویل تجربہ رکھتے ہوں۔ ترجمے سے ہٹ کر صرف ادب میں بھی جھانکیے تو اردو میں پطرس بخاری، ابنِ انشا، کرنل محمد خان، مشتاق یوسفی، مختار مسعود، شان الحق حقی، شمس الرحمٰن فاروقی اور مشفق خواجہ وغیرہ کی تحریروں کے جاندار ہونے کی ایک بڑی وجہ یہ ہے کہ اِن میں کا ہر ایک صرف اکیلی اردو کا منتہی نہیں ہے بلکہ انگریزی میں بھی یکساں قدرت رکھتا ہے، اور اِن کے ساتھ ساتھ کم سے کم ایک ایک اور زبان میں بھی اعلیٰ مہارت کا حامل ہے۔ اقوامِ متحدہ اور عالمی ادارۂ صحت کے لیے سالہا سال تک مترجم کے طور پر کام کرنے والے ماہرِ لسانیات ونفسیات Claude Piron نے لکھا ہے کہ مشینی ترجمہ کار زیادہ سے زیادہ یہی کر سکتا ہے کہ مترجم کے کام کا آسان تر حصہ خودکار طریقے سے کر دے۔ مشکل حصہ تو وہی ہوتا ہے جس میں ماخذ متن کے معانی میں ابہام (Ambiguity) دور کرنے کے لیے تحقیق پر بہت سا وقت صرف کرنا پڑتا ہے، اور جسے مطلوبہ زبان میں ڈھالتے وقت گرامر اور لفظیات کے بارے میں بہت سے فوری اور ہنگامی فیصلے کرنا پڑتے ہیں۔[۱]

''معانی میں ابہام'' ایک کثیر سطحی اور کئی مفہوم رکھنے والی اصطلاح ہے۔ اب تک لفظوں کی صرف اندرونی کیفیت کی بات ہوئی۔ زبان معاشرے میں بولی جاتی ہے۔ معاشرہ لفظوں کی حیثیت کے بارے میں جو فتویٰ دے اُسے قبول کیے بنا چارہ نہیں۔ یاد کیجیے کہ جب آل انڈیا ریڈیو کی صبح کی نشریات بہاول پور میں سنائی دی جانے لگیں (یہ تقسیمِ ہند سے پہلے کی بات ہے) تو خبریں شروع کرنے سے پہلے اعلان ہوتا کہ ''یہ آل انڈیا ریڈیو ہے۔ اِس وقت یہاں صبح کے چھے بجے ہیں......'' یہ جملہ بہاول پور کی ریاستی زبان کے اعتبار سے شدید قابلِ اعتراض تھا۔ چنانچہ نواب صاحب بہاول پور نے ڈائریکٹر جنرل آل انڈیا ریڈیو کو باقاعدہ سرکاری چٹھی کے ذریعے یہ جملہ بدلنے کو لکھا۔ اور یہ تبدیلی کرنا بھی پڑی۔

اِس ضمن میں تیسری بات ماحول، معلوماتِ عامہ اور زمینی حقائق کی ہے، جن کا ترجمے ہی میں نہیں بلکہ بسا اوقات ماخذ متن میں بھی پھڈا ہو جاتا ہے۔ مثال لیجیے کہ شخصیت کا ذکر کرتے ہوئے آزادؔ سے محمد حسین آزاد اور مولانا ابوالکلام آزاد میں التباس ہو جایا کرتا ہے۔ چراغ حسن حسرت اور رئیس المتغزلین حسرت موہانی میں تمیز نہیں ہو پاتی۔ نظیر اکبر آبادی اور اکبر الٰہ آبادی کو جوڑ کر نظیر اکبر الٰہ آبادی بھی بنا مل جاتا ہے۔ جنابِ انور مسعود نے ایسی معلومات کے ملغوبے کی کیا خوب تصویر کھینچی ہے: ؎

اِک عطائی کر رہا تھا ایک مجمعے سے خطاب

یاد ہے بس ایک ٹکڑا مجھ کو اُس تقریر کا

بات یہ بانگِ درا میں شیخ سعدی نے کہی

''صبح کرنا شام کا لانا ہے جوئے شیر کا''

یہاں اقبالؔ کے ایک دیوان میں غالبؔ کا ایک مصرع ڈال کر یہ دیوان شیخ سعدیؔ کو دے دیا گیا ہے۔ نثر سے بھی ایک مثال لیجیے۔ یہ دو جملے دیکھیے:

قائدِ اعظم محمد علی جناح برطانوی راج کے خاتمے کے زبردست حامی تھے۔

مولانا ابوالکلام آزاد برطانوی راج کے خاتمے کے زبردست حامی تھے۔

اِن دونوں جملوں میں بظاہر صرف نام کا فرق ہے۔ مشین اندھے وار صرف نام بدل کر یہ ترجمہ کر سکتی ہے اور بس۔ زمینی حقائق یہ ہیں کہ پہلے جملے میں مذکور شخصیت تقسیمِ ہند اور قیامِ پاکستان کی علمبردار تھی جب کہ دوسرے جملے میں مذکور شخصیت آزادیِ ہند کی ہوا خواہ۔ چنانچہ اِن دونوں جملوں کا ترجمہ اصولاً ایک سا نہیں ہوگا۔ پس معلوم ہوا کہ درست ترجمہ بہت سے نواحی مطالعے کا مطالبہ کرتا ہے، جس کے لیے زبان کا محض باصلاحیت بولنے والا ہونا ہی کافی نہیں ہوتا بلکہ اِدھر اُدھر سے کئی باتیں پوچھ پُچھا کر پہلے اپنی معلومات کو درست اور مکمل کرنا ضروری ہو جاتا ہے۔ یہی وہ بات ہے جس کی طرف Claude Piron نے توجہ دلائی ہے۔

اِسی طرح ایک بڑا مسئلہ ایک زبان میں دوسری زبان کے الفاظ اور اُن کی تصریفی صورتوں کا استعمال (Code-mixing) ہے۔ مثلاً اِس جملے: وہ لیٹ گیا۔ میں لفظِ لیٹ اردو مصدر لیٹنا کی ایک تصریفی صورت بھی ہو سکتا ہے اور انگریزی لفظ Late بھی۔ اور اِس جملے: یہ پھٹیک بھی میرے ہی سر پڑی۔ میں لفظِ پھٹیک اردوائی ہوئی صورت ہے انگریزی (بلکہ فرانسیسی) لفظ Fatigue کی۔ یہ وہ چیزیں ہیں جنھیں متن کو ترجمے کے لیے مشین کے حوالے کرنے سے پہلے دیکھنا چاہیے ورنہ مشین کو دی گئی اشتقاقی معلومات کا دھواں نکل جائے گا۔

پھر کچھ ایسے مسائل بھی ہیں جنھیں کسی خاص عنوان کے تحت لکھنا بھی مشکل ہوتا ہے چہ جائیکہ ترجمہ کرنے کے کسی پروگرام کی چند لائنیں سوچی جائیں۔ کچھ عرصہ پہلے سے پاکستان میں ایک بنک KASB نے کام شروع کیا ہے۔ یہ لفظ KASB اصلاً مخفف ہے ''خادم علی شاہ بخاری'' (جو ایک خالص اردو ترکیب کا نام ہے) میں ہر لفظ کے انگریزی ہجوں کے پہلے حروف کے مجموعے کا (Khadim Ali Shah Bukhari)۔ اِس مخفف سازی کی مزید خوبصورتی بلکہ چکر بازی دیکھیے کہ اِس میں اِسے عربی لفظ کسب (جس کا معنٰی کمانا/کمائی وغیرہ ہے) کے ساتھ کمال مہارت سے ذومعنی کر دیا گیا ہے۔ اب بتائیے کہ مشین لفظِ کسب کو بطورِ اسمِ معرفہ نمٹائے، بطورِ فعل دیکھے، بطورِ مخفف برتے، یا کچھ اور؟ اِس لفظ (یا مخفف) کی قواعدی حیثیت کی تعیین کی تو ایک رہی، صرف یہی طے کرنا ایک اچھے بھلے لکھے پڑھے آدمی کے لیے گھما دینے والی ذہنی ورزش ہے کہ اِس کثیر النسل لفظ کو کس زبان کے ذخیرۂ الفاظ میں ڈالا جائے: عربی، اردو یا انگریزی — اور نتیجہ اِس کا مرغی اور انڈے کی بحث والا ہے۔

یہ وہ موضوعات ہیں جن سے مشینی ترجمہ نہ تو خود نبٹ سکتا ہے اور نہ ہی اُس سے اِس کی توقع کی جانی چاہیے۔ ''گہرے'' (Deep) طریقے سے کام کرنے والے ایک مثالی (Ideal) مشینی ترجمہ کار سے مطالبہ کیا جا سکتا ہے کہ وہ معانی میں ابہام کے دور کرنے کی اِس قسم کی ساری تحقیقات خود سے کرے۔ لیکن ظاہر ہے کہ اِس کے لیے مشین کی مصنوعی ذہانت میں بے حد اضافہ کرنا ہوگا، جو فوری طور پر ممکن نہیں ہے۔ چنانچہ ''سطحی'' (Shallow) طریقوں سے ماخذ متن کے جملوں میں پایا جانے والا ہر طرح کا ابہام، جس میں مترجم اپنے وجدان اور نواحی مطالعے سے ملنے والے انتباہات کے زیرِ اثر ذاتی طور پر کچھ معلومات حاصل کر کے یہ خلا دور کرتا ہے، نسبتاً آسان ہے۔

4: خاتمہ

مشینی ترجمے کی بنیادی طور پر چار تکنیکیں ہیں: مشین کو زبان کے قواعد سکھا کر مصنوعی طور پر ذہین بنا دینے سے، یا مشین پر ٹرانسلیشن میموری کا بوجھ لاد کر، یا شماریاتی طریقے سے لفظوں کو ناپ تول کر، یا کچھ تکنیکوں کی خوبیوں کو جمع کر کے ایک دوغلی تکنیک بنا کر مشینوں کو خودکار طریقے سے کسی بھی دوسری زبان میں ترجمہ کر دینے کے قابل بنا دینا۔ مشینی ترجمہ وقت کی ایک بڑی ضرورت اور مانگ ہے۔ اس مقالے میں اردو کے تناظر میں اس ضرورت اور ان تکنیکوں کا ایک مطالعہ کمپیوٹر اور اطلاعیات کی زبان میں پیش کیا گیا ہے۔

تحریر: ۲۷/فروری ۲۰۱۰ء، مطابق ۱۲/ ربیع الاولیٰ ۱۴۳۱ھ

مزید مطالعہ:

اِس مقالے کے قارئین سے درخواست ہے وہ مندرجۂ ذیل مقالات کو بھی توجہ سے پڑھ ڈالیں:

۱۔ بخاری، سید ذوالکفل؛ظہیر احمد، ڈاکٹر؛صفوان محمد چوہان، ڈاکٹر حافظ؛ ۲۰۰۹ء، اردو کارپس: تکنیکی تعارف، اہمیت، ضرورت اور دائرہ و لائحۂ عمل ۔مشمولہ: جرنل آف ریسرچ، بہاء الدین زکریا یونیورسٹی ملتان، شمارہ-۱۴۔

۲۔ صفوان محمد چوہان، ڈاکٹر حافظ؛ ۲۰۱۰ء، اردو اطلاعیات: آج اور کل۔ مشمولہ: سہ ماہی اثبات، شمارہ: ۵-۴۔

۳۔ صفوان محمد چوہان، ڈاکٹر حافظ؛ ۲۰۰۷ء، ترویج اردو کی ایک فوری ضرورت: اردو رسم الخط میں انگریزی-اردو لغات کی آن لائن فراہمی۔ مشمولہ: اورینٹل کالج میگزین۔ جلد-۸۳، عدد: ۲-۱۔

حواشی وحوالہ جات

حوالہ

بخاری، سید ذوالکفل؛ظہیر احمد، ڈاکٹر؛صفوان محمد چوہان، ڈاکٹر حافظ؛ ۲۰۰۹ء، اردو کارپس: تکنیکی تعارف، اہمیت، ضرورت اور دائرہ و لائحۂ عمل ۔مشمولہ: جرنل آف ریسرچ، بہاء الدین زکریا یونیورسٹی ملتان، شمارہ-۱۴۔

حواشی

۱۔ ملاحظہ کیجیے: www.absoluteastronomy.com/topics/Machine_translation

...machine translation, at its best, automates the easier part of a translator's job; the harder and more time-consuming part usually involves doing extensive research to resolve ambiguities in the source text, which the grammatical and lexical exigencies of the target language require to be resolved.

۲۔ String of alphabets embedded between two blank spaces or having a punctuation sign on its either side

۳۔ القرآن ۲:۳۱۔

۴۔ Europe Speaks Arabic by Dr V Abdul Rahim, Institute of the Language of the Quran Inc, Toranto, Canada, 2008.

۵۔ مصباح القرآن، عبدالرحمٰن طاہر، انسٹی ٹیوٹ آف ایجوکیشنل ریسرچ، قائداعظم کیمپس، پنجاب یونیورسٹی، لاہور۔

۶۔ ایک سہ لسانی فیتے والا پروگرام ملاحظہ کیجیے جس میں عربی اور انگریزی کے ساتھ ساتھ اردو بھی چلتی ہے:
urdu-arabic-english-text-ticker.smartcode.com/info.html

۷۔ ملاحظہ کیجیے:
www.chandos.ca/Metrics_for_Evaluating_Translation_Memory_Software.pdf

۸۔ ملاحظہ کیجیے: www.hutchinsweb.me.uk/Milestones-6.pdf

تشکر(Acknowledgement):

۱۔ یہ مقالہ لکھنے میں جناب وصی اللہ کھوکھر نے میری بہت مدد کی ہے اور بے حد اہم معلومات فراہم کی ہیں۔ اُنھوں نے اِس مقالے کو چھپنے سے پہلے جانچا بھی ہے۔ رسمی شکریے کے کوئی الفاظ اُن کی خدمات کا بدل نہیں ہو سکتے۔ جناب وصی اللہ کھوکھر اردو مشینی ترجمے میں پاکستان کے پرانے لوگوں میں سے ہیں اور کمپیوٹر اور اطلاعیاتی ٹیکنالوجی کی پندرہ سے زیادہ کتابوں کے مصنف/مترجم ہیں۔ وہ مائیکروسافٹ کے ساتھ اردو ترجمے کے کئی ایک پراجیکٹ کر چکے ہیں۔ کمپیوٹر سائنس کی اصطلاحات کو اردو۔انگریزی میں جس سہولت اور تخلیقی صلاحیت سے وہ استعمال کرتے ہیں، صرف پاکستان ہی نہیں بلکہ پوری دنیا میں بھی ایسے کم یاب لوگ کم ہی ہوں گے۔

۲۔ اِس مقالے کی تیاری کے دوران میں حوالے کی بہت سی باتوں کے ضمن میں ڈاکٹر شمس الرحمٰن فاروقی، ڈاکٹر گوپی چند نارنگ، ڈاکٹر خواجہ محمد زکریا اور ڈاکٹر خورشید رضوی صاحب کو بار بار زحمت دی گئی۔ ہر چار حضرات کا شکریہ واجب ہے۔

۳۔ یہ مقالہ لکھنے کے لیے www.wikipedia.org کے مختلف مقالات سے آزادانہ استفادہ کیا گیا ہے۔

فہرستِ اسنادِ محولہ

مآخذ

الف: کتابیات

۱۔ قرآنِ پاک۔ The Message of THE QURAN by Muhammad Asad, Dar al-Andalus Limited, 3 Library Ramp, Gibralter. 1980.

۲۔ اقبال، شاعرِ مشرق ڈاکٹر علامہ سر محمد، کلیاتِ اقبال، پانچواں ایڈیشن، اقبال اکادمی پاکستان، لاہور۔ ۲۰۰۰ء

۳۔ ذوق، محمد ابراہیم، کلیاتِ ذوق، مرتبہ ڈاکٹر تنویر احمد علوی، مجلسِ ترقیِ ادب لاہور۔ ۲۰۰۹ء

۴۔ عابد صدیق، پانی میں ماہتاب، دوسرا ایڈیشن، الحمد پبلی کیشنز، لاہور۔ ۲۰۰۶ء

۵۔ غالب،مرزا اسداللہ خاں،دیوانِ غالب۔غالب انسٹی ٹیوٹ،نئی دہلی۔فروری ۱۹۸۶ء

۶۔ Europe Speaks Arabic by Dr V Abdul Rahim, Institute of the Language of the Quran Inc, Toranto, Canada, 2008.

ب: رسائل اور تحقیقی جرائد

۱۔اردو اطلاعیات آج اور کل ۔مشمولہ: سہ ماہی اردو نامہ مجلسِ زبانِ دفتری حکومتِ پنجاب۔شمارہ اکتوبر ۲۰۰۸ء تا مارچ ۲۰۰۹ء۔ص ص ۱۶۳ تا ۱۷۳۔

۲۔اردو اور دنیا کی بڑی زبانوں کی شماریات ۔مشمولہ: اردو سائنس میگزین ۔شمارہ-۲، ۲۰۰۹ء۔ص ص ۱۰ تا ۲۳۔

۳۔اردو رسم الخط میں انگریزی-اردو لغات کی آن لائن فراہمی ۔مشمولہ: اردو سائنس میگزین ۔شمارہ-۴، ۲۰۰۷ء۔ص ص ۲۲ تا ۳۷ اور مشمولہ: اورینٹل کالج میگزین۔ جلد-۸۳،عدد: ۲-۱، ۲۰۰۷ء۔ص ص ۲۴۳ تا ۲۶۲۔

۴۔اردو لغت (تاریخی اصول پر) بدلتے لسانی تناظر میں چند تجاویز ۔مشمولہ: جرنل آف ریسرچ، بہاء الدین زکریا یونیورسٹی ملتان ۔شمارہ-۱۲(۲۰۰۷ء)۔ص ص ۲۷۵ تا ۲۸۶۔

۵۔مشین ریڈایبل اردو رسم الخط: حروف کی کشتیاں، اعراب، نقطے، شوشے اور کششیں۔مشمولہ: اردو سائنس میگزین۔شمارہ-۳، ۲۰۰۸ء۔ص ص ۲۹ تا ۴۳۔

ج: اِنٹرنیٹ سائٹس (چند منتخب سائٹس)

1. http://www.hutchinsweb.me.uk/Nutshell-2005.pdf

2. http://www.isi.edu/natural-language/projects/rewrite/mtsummit03.pdf

3. http://www.machinetranslations.org

تکنیکی مشاورت

۱۔ڈاکٹر خواجہ محمد زکریا،سابق پرنسپل،اورینٹل کالج،جامعۂ پنجاب،لاہور [اردو و انگریزی]

۲۔ڈاکٹر سید خورشید حسن رضوی،لاہور [اردو و عربی]

۳۔خواجہ غلام ربانی مجال، ۲۸-گلستان کالونی،لین نمبر-۲،نیشنل پارک روڈ،راول پنڈی [اردو و انگریزی]

۴۔راؤ صفدر رشید،ڈیٹابینک سپروائزر،مرکزِ فضیلت برائے اردو اطلاعیات،مقتدرۂ قومی زبان،اسلام آباد [اردو مشینی ترجمہ]

۵۔وصی اللہ کھوکھر،ایم جی ایچ سالوشنز،کاموکے [اردو مشینی ترجمہ]

۶۔ڈاکٹر پرویز احمد، Language Technologies Research Centre, International Institute of Information Technology, Gachibowli, Hyderabad. [اردو مشینی ترجمہ]