ڈاکٹر حافظ صفوان محمد چوہان/ قرۃ العین

اُردو مثال گھر (دی بینک آف اُردو): ایک تعارف

Urdu Data House : An Introduction

Dr Hafiz Safwan Muhammad Chohan, Chairman, Computer & Data Services Department, Manager Data Network, Telecommunication Staff College, Haripur, Pakistan.

Qurat ul Ain, Assistant Informatics Officer, Centre of Excellence for Urdu Informatics, National Language Authority, Islamabad, Pakistan.

Abstract:

The Bank of Urdu (TBU) is a proposed corpus of Urdu text of both written and spoken language. The idea is basically taken from The Bank of English which is being used as the backbone of English language engineering, corpus and lexicon development, and works of the same order.
Issues of the formation of TBU are discussed in this introductory paper. Where the Urdu corpus must deviate from the standard roadways of The Bank of English, is specially dealt with.

کلیدی الفاظ:

مشین ریڈایبل اُردو، لسانی انجینئرنگ، اُردو لسانیات، مال خانہ (Repository)، عام بول چال (Informal Spoken Language)، قاموس الامثال (Corpus)، سلینگ، جارگن۔

مخففات:

اُردو ڈیٹا بیس/ اُردو ڈیٹا بینک: UDB: Urdu Data Base/ Urdu Data Bank

اُردو مثال گھر/ اُردو کارپس/ دی بینک آف اُردو : TBU: The Bank of Urdu

برٹش نیشنل کارپس: BNC: British National Corpus

پرانی انگریزی کے نظائر: LEME: Lexicons of Early Modern English

0: تعارف

دنیا بھر میں زبانوں پر تحقیق کا کام اِس وقت زوروں پر ہے۔ زبان پر تحقیق سے مراد زبان کی ساخت پرداخت کا مطالعہ بھی ہے اور بین اللسانی تعلقات کا مطالعہ بھی۔ کمپیوٹر کی آمد کے ساتھ ہی زبانوں پر تحقیق کے علم میں نئی اور وسیع تر جہتیں سامنے آنا شروع ہوئیں اور خالص سائنسی انداز میں زبانوں کی بنت اور اثرات کا جائزہ لیا جانا شروع ہوا۔ اِس مطالعے اور تحقیق کے لیے زبان کے استعمالی نظائر کی بنیادی اہمیت ہے۔ اِس وقت کمپیوٹر پر دنیا کی کئی زبانوں کا ڈیٹا اِس مقصد کے لیے جمع کیا گیا ہے۔ یہ ڈیٹا وقت گزرنے کے ساتھ ساتھ کمیت میں بڑھتا اور کیفیت میں بہتر سے بہتر ہورہا ہے۔ لسانی انجینئرنگ (Language Engineering) اور لسانیات (Linguistics) کے سبھی شعبوں میں تحقیق کے لیے اِس ذخیرۂ امثال کی بنیادی ضرورت ہے۔ دی بینک آف انگلش کے نام سے انگریزی زبان کا سب سے بڑا مال خانہ (Repository) جس کے روح و رواں آنجہانی پروفیسر جان میک ہارڈی سنکلیئر تھے، اِس وقت برمنگھم یونیورسٹی برطانیہ میں قائم ہے اور اِن شعبوں میں جاری تحقیقات کے لیے موزوں لسانی متن محققین اور تحقیقی اداروں کو فراہم کررہا ہے۔ زیرِ نظر مقالے میں اُردو کے لیے ایسے ڈیٹا بینک کے قیام کی ضرورت کے بارے میں بتایا گیا ہے۔

مرکزِ فضیلت برائے اُردو اطلاعیات، مقتدرۂ قومی زبان اسلام آباد میں اِس موضوع پر کام کا آغاز ہورہا ہے۔ پہلے یہاں پر یہ شعبہ ''اُردو ڈیٹا بینک'' کے نام سے تھا لیکن یہیں پر کام کررہے ایک دوسرے شعبے ''اُردو ڈیٹابیس'' کے مخفف کے ہم آواز اور ہم صورت ہونے کی وجہ سے التباس (اِن دونوں شعبوں کا انگریزی مخفف UDB تھا) اور دی بینک آف انگلش کے نام سے بہت دور ہونے کی وجہ سے اِس کا نام بدلنے کی تجویز دی گئی، جسے ''اُردو اطلاعیات: آج اور کل'' کے عنوان سے ۷-۸ جون ۲۰۰۸ء کو ہونے والی ورکشاپ میں ماہرینِ اُردو و اطلاعیات اور حکامِ بالا نے قبول کیا؛ اور اِس کا اُردو نام ''اُردو مثال گھر'' بھی قبول کیا۔

اُردو مثال گھر بوجوہ دی بینک آف انگلش سے مختلف ذرائع سے بھی مواد لے گا۔ اِن وجوہ اور اِن ذرائع کے بارے میں بھی تفصیلات دی گئی ہیں۔

1: دی بینک آف انگلش: مختصر تعارف

دی بینک آف انگلش تقریباً ۴۵۰ ملین بولے اور لکھے جانے والے الفاظ کا مجموعہ ہے جسے کمپیوٹر میں زبان کے تحقیقی استعمال کے لیے اکٹھا کیا گیا ہے۔ یہ مواد ۱۹۹۰ء سے اب تک کے استعمال ہونے والے ذخیرہ کردہ متن سے لیا گیا ہے۔ ہر مہینے اِس مواد میں ۱۰ ملین الفاظ کا نیا ڈیٹا ڈالا جاتا ہے۔ اس مال خانے میں موجود بیشتر حصہ برطانوی انگریزی پر مشتمل ہے جب کہ امریکن انگریزی کے متون ۲۵٪ ہیں؛ پانچ فیصد مواد دنیا بھر میں بولی جانے والی انگریزی سے لیا گیا ہے جس میں آسٹریلیائی اور سنگاپور کی انگریزی بھی آتی ہے۔

دی بینک آف انگلش میں صرف اور صرف مشین ریڈایبل انگریزی متن ہے جو کمپیوٹر پر محفوظ ہے۔ یہ متن پلیٹ فارم سے غیر وابستہ (Platform Independent) ہوتا ہے اور کسی بھی پروگرام کے لیے استعمال ہونے کی صلاحیت رکھتا ہے۔

دی بینک آف انگلش میں صرف اور صرف آج کی بولی، لکھی اور استعمال کی جانے والی انگریزی زبان کے نظائر موجود ہیں۔ زبان کے کلاسیکل استعمال سے اِسے واسطہ نہیں۔

1.1: دی بینک آف انگلش کے مآخذ

دی بینک آف انگلش میں لیا گیا مواد اخبارات، میگزین، فکشن اور نان فکشن کتب، پمفلٹ اور خطوط وغیرہ سے لیا گیا ہے۔ [۱] دی بینک آف انگلش کا دو تہائی کارپس میڈیا کی زبان پر مشتمل ہے، یعنی اخبارات اور میگزینوں کے ساتھ ساتھ ریڈیو اور ٹی وی کی زبان۔ [۲] دی بینک آف انگلش کے تقریباً ۲۰ ملین الفاظ عام بول چال پر مشتمل ہیں۔ یہ ڈیٹا غیر رسمی ملاقاتوں، انٹرویوز اور مباحثوں کو ریکارڈ کر کے لکھا گیا ہے۔ [۳] یہ حصہ اِس لیے رکھا گیا ہے کہ انگریزی زبان کے زندہ استعمال اور تازہ ترین رجحانات سے باخبر رہا جا سکے۔

1.2: دی بینک آف انگلش میں کہاں سے متن نہیں لیا جا رہا؟

دی بینک آف انگلش ادبی زبان کو شامل نہیں کیا گیا، یعنی ناول اور شاعری کی زبان کو بینک کا حصہ نہیں بنایا گیا۔ اِسی طرح یہ بینک سائنسی مقالات اور مختلف شعبوں کی مخصوص زبان

(Jargon; slang) کو بھی محفوظ نہیں کرتا۔ وجہ یہ ہے کہ اِن سب جہات کی زبان عام زبان نہیں ہوتی بلکہ اہلِ علم یا مخصوص لوگوں/ شعبوں کی زبان ہوتی ہے۔ لسانی انجینئر نگ اِن پگڈنڈیوں کی نہیں بلکہ شاہراہ کی زبان پر وقت اور صلاحیتیں خرچ کرنے اور کرانے سے متعلق ہے۔ دی بینک آف انگلش چلتے ہوئے کاروبار میں اپنا مال لگاتا ہے نہ کہ محدود یا مخصوص چھوٹے کاروبار میں۔

1.3: دی بینک آف انگلش میں ڈیٹا جمع کرنے کے طریقے

مشین ریڈایبل متن مختلف طریقوں سے اکٹھا کیا جاسکتا ہے۔ مثلاً یہ کہ مختلف ویب سائٹوں پر سے مطلب کا ڈیٹا حاصل کیا جائے۔ اگر مواد متن (Text) کی شکل میں نہیں ہے تو اِسے متن کی صورت میں ٹائپ کرایا جائے۔ یہ ٹائپنگ معیاری ہونی چاہیے۔ آواز سے متن (Speech-to-text) اور تصویر سے متن (OCR) کی جدید ترین مشینوں کو استعمال کرکے یہ مواد جمع کیا جاتا ہے۔

مواد کی جمع آوری میں بنیادی بات یہ ہے کہ یہ کم وقت میں اور کم خرچ میں ہو، لیکن ابہام اور اغلاط سے پاک ہو۔

1.4: دی بینک آف انگلش کے استعمالات

دی بینک آف انگلش کو بنانے کی اہم ترین وجہ ایک ایسا قاموس الامثال (Corpus) تیار کرنا تھا جو محققین اور لغت نویسوں کو الفاظ کے متعلق زیادہ سے زیادہ اور معیاری معلومات فراہم کرسکے۔ وقت گزرنے کے ساتھ ساتھ اِس کے استعمالات کی نئی سے نئی صورتیں سامنے آتی رہی ہیں، اور اِن میں اضافہ روزافزوں ہے۔ چناں چہ دی بینک آف انگلش میں الفاظ کی سب تصریفی صورتوں کا استعمال موجود ہوتا ہے۔ اِس میں نظائر چونکہ روزمرہ کی زبان (یعنی آج کل کی بولی اور لکھی جانے والی زبان) سے لی گئی ہیں لٰہذا زبان کا تازہ ترین، زندہ استعمال سامنے لاکر اِس پر جاری تحقیقات میں سہولت فراہم کی جاتی ہے۔

2: دی بینک آف انگلش کیا ہے؟ کیا نہیں؟

دی بینک آف انگلش ذخیرۂ الفاظ یا مترا کمہ (wordlist) نہیں ہے بلکہ انگریزی الفاظ کی جملہ تصریفی صورتوں اور ہر سطحی استعمال کے نظائر (مثالوں) پر مشتمل متون کا مجموعہ ہے۔ اِس بینک میں فی الوقت موجود متون ۱۹۹۰ء اور اِس کے بعد کے ہیں۔ اِس وقت سے پہلے کے نظائر بینک سے نکال کر سرد خانے میں ڈال دیے گئے ہیں۔ ۴؎ یہ انگریزی زبان کا بینک ہے نہ کہ انگریزی ادب کا۔

دی بینک آف انگلش انگریزی لسانیات یا لسانی انجینئرنگ کا نام نہیں ہے۔ اِس کی مثال عام بینک کے طور سے لی جا سکتی ہے۔ عام بینک لوگوں کی رقوم جمع کرتا ہے اور محفوظ رکھتا ہے۔ بینک رقوم کو استعمال نہیں کرتا بلکہ مختلف کاروبار کرنے والوں کو استعمال کے لیے دیتا ہے؛ اُن سے اِس استعمال سے ملنے والے نفع میں شرکت کرتا ہے؛ وغیرہ۔ دی بینک آف انگلش بھی الفاظ اور اُن کے استعمالات کو صرف محفوظ رکھتا ہے، اور محققین اور لغت نویسوں وغیرہ کو اپنے پاس موجود متون پر نفع بخش کام کرنے کی اجازت دیتا ہے، اور اِن کاموں کے نتائج سے اپنی قدر و قیمت میں اضافہ کرتا رہتا ہے۔

3: کلاسیکل انگلش اور دی بینک آف انگلش

جیسا کہ اوپر عرض کیا گیا، دی بینک آف انگلش انگریزی زبان کے معاصر استعمال کے نظائر پر مشتمل ہے۔ سوال پیدا ہوتا ہے کہ پھر انگریزی کا کلاسیکل استعمال کہاں گیا؟ اِس کے بارے میں عرض ہے کہ کلاسیکل اور ابتدائی اَدوار کی زبان کو انگریزی میں الگ سے موجود ڈیٹا بینکوں (مثال گھروں) کی شکل میں محفوظ کیا جاتا ہے۔ اِس کی ایک مثال LEME ہے، جس میں موجودہ انگریزی کے ایسے نظائر جمع کیے گئے (اور کیے جا رہے) ہیں جنھیں ''پرانی'' انگریزی (Early Modern English) کہا جاتا ہے۔ ملاحظہ کیجیے: ۵؎

4: اُردو مثال گھر (دی بینک آف اُردو)

اُردو مثال گھر دی بینک آف انگلش کے ڈھب اور ہیولے پر تجویز کیا گیا اُردو متون کا ایسا خزانہ ہے جو الفاظ کے استعمال کے نظائر ذخیرہ کرے گا تاکہ اُردو پر بطور فطری زبان تحقیق

ہو سکے۔ اِس میں موجود سارے متون مشین ریڈایبل اُردو میں اور پلیٹ فارم سے غیر وابستہ حالت میں ہوں گے تاکہ کسی بھی مواجہ پر اُردو میں کام کرنے والے محقق یا مشین کو متن کی سہولت فراہم کی جا سکے۔

اُردو مثال گھر اپنے مال خانے میں بنیادی طور پر سب متون کو اُردو کے روایتی رسم الخط (Indo-Perso-Arabic Script) میں رکھے گا اگرچہ کسی وقت میں رومن اُردو اور دیوناگری کے لیے معاونت (support) بھی فراہم کی جا سکے گی۔ چونکہ یہ مثال گھر دائیں سے بائیں لکھے جانے والے اُردو کے متون فراہم کرے گا اِس لیے اُردو کی ہم رشتہ، تمام پاکستانی زبانوں کے متون ذخیرہ کرنے کے لیے اگر مثال گھرانے بنائے جائیں (جنھیں ضرور اور فوراً بنانا شروع کرنا چاہیے) تو اِن سب کو اِس اُردو مثال گھر کے ذیلی اداروں (Subsidiaries) کے طور پر رکھا اور استعمال کیا جا سکتا ہے۔

4.1: اُردو مثال گھر کے مآخذ

شان الحق حقی (۱۹۹۶ء) نے لکھا ہے کہ ''...... اُردو پنپنے اور پروان چڑھنے نہ پائی تھی کہ کلاسیکیت کی راہ پر چل نکلی۔ شعر گوئی کے کام کی رہ گئی۔'' بات درست ہے۔ اُردو پر بطورِ زبان کوئی تحقیق کرنے، علی الخصوص کمپیوٹر پر کوئی ایسا کام کرنے کا ارادہ کیا جائے تو معلوم ہوتا ہے کہ اُردو کا تحریری سرمایہ زیادہ تر زبان کے صرف ادبی استعمالات کے نظائر پر مشتمل ہے۔ اِس لیے اُردو مثال گھر کے منصوبے پر دی بینک آف انگلش سے خاصے فاصلے پر رہ کر کام کرنا ضروری ہو گیا ہے۔ چونکہ اُردو نے ادبی طور پر بہت ترقی کی ہے اِس لیے اُردو متون اکٹھا کرنے میں اُردو ادب سے مدد لینی پڑے گی۔ یاد رہے کہ دی بینک آف انگلش نے ادبی زبان کے متون کو اپنے مال خانے کا حصہ نہیں بنایا۔ بایں وجوہ اِس تجویز میں یہ بات قابلِ ذکر ہے کہ اُردو مثال گھر اُردو اخبارات، ادب، (کسی حد تک) شاعری، اِنٹرنیٹ اور دیگر شائع کردہ کتابوں سے اپنا راس المال (Capital) حاصل کرے گا۔ یہ بہت بنیادی فرق ہے جو دی بینک آف انگلش اور اُردو مثال گھر میں ہے۔

آغاز میں ایسا بہت سا متن جمع کر کے کام شروع کیا جا سکتا ہے۔ اُردو زبان کے معاصر استعمالات کو مشین ریڈایبل اُردو میں جمع کرنے کا کام بھی فوراً شروع کر لیا جائے تاکہ

اِس مد میں متن آ تا رہے۔ رفتہ رفتہ کلاسیکل متون اور پرانی اُردو کے نظائر کے ذخائر کو LEME کے طرز پر الگ (Denest) کر دیا جائے تا کہ یہ ایک مستقل ماخذ کے طور پر پنپ سکے اور اُردو زبان پر کلاسیکی تحقیق میں کام آ سکے۔

فوری طور پر کلاں اُردو لغت، شائع کردہ اُردو لغت بورڈ کراچی سے بھی مدد لی جائے تو انتہائی مناسب اور برمحل ہے کیوں کہ اُردو کے ذخیرۂ الفاظ کے کلاسیکل استعمال کے سب سے زیادہ نظائر صرف یہیں سے یکجا مل سکتے ہیں۔ اِس کی شکل یہ ہوگی کہ اُن ماٰخذ کی ایک ڈھیلی ڈھالی سی فہرست بنالی جائے جن پر اِس لغت نے زیادہ انحصار کیا ہے۔ اِن ماٰخذ کو مشین ریڈایبل اُردو میں، اور درست اور معیاری انداز میں، کمپیوٹر پر محفوظ کرنے کا آغاز کر دیا جائے۔

4.2: اُردو مثال گھر میں فی الوقت متون کہاں سے نہیں لیے جائیں گے؟

فی الحال اُردو مثال گھر کے مال خانے میں رکھے جانے والے متون بہت ہی مخصوص جگہوں سے لیے جانے کی تجویز ہے۔ یہ مثال گھر ریڈیو، ٹی وی اور میڈیا کی زبانوں سے (زیادہ تر) ڈیٹا نہیں لے گا۔ چنانچہ عام بول چال کی زبان اِس میں عموماً شامل نہیں کی جائے گی؛ لیکن اگر کہیں سے معیاری متن میں محفوظ کردہ بول چال کی زبان مل جائے تو اُسے ضرور لیا جائے گا۔ دی بینک آف انگلش کی طرح اِس میں سائنسی تکنیک سے متعلق زبان بھی شامل نہیں کی جائے گی، اگرچہ مستقبل میں اِس پر بھی کام کیا جا سکتا ہے۔ فوری طور پر کئی قسم کے انٹرویوز اور ٹاک شوز وغیرہ میں شامل ہونے والے الفاظ بھی اِس مال خانے کا حصہ نہیں بنیں گے۔ وغیرہ وغیرہ۔

اُردو مثال گھر کے ماٰخذ کو اِتنا محدود کرنے کے وجوہ بہت سے ہیں۔ سرِ دست اِس ضمن میں چند سامنے کی باتیں عرض کی جاتی ہیں۔

۱۔ معیاری متن کی جمع بندی کے لیے بہت وقت درکار ہوگا۔ موجودہ متون زیادہ تر معیاری نہیں ہیں۔ اِس لیے ادب کے متون کو معیاری بنا کر کام کرنا ضروری ہے۔ معیاری متون سے مراد اِملاء کے مسائل کا حل بھی ہے۔

۲۔ موجودہ اُردو متون زیادہ تر مشین ریڈایبل حالت میں نہیں ہیں۔ اِس ضمن میں مدیحہ اعجاز اور ڈاکٹر سرمد حسین (۲۰۰۷ء) نے تفصیل سے بتایا ہے کہ لوگ نہ صرف

متن فراہم کرنے میں فراخ دِلی کا ثبوت نہیں دیتے بلکہ اخبارات والے لوگ تو علی العموم متن کو مشین ریڈ ایبل بنانے کی ضرورت سے بھی آگاہ نہیں ہیں۔ مثال گھر صرف لکھے ہوئے متون کو اپنا حصہ بناتا ہے۔ اُردو مثال گھر میں ذخیرہ کیے جانے والے سارے متون کو پہلے مشین کے لیے قابلِ فہم بنانا ہوگا، جس کے لیے کافی وقت درکار ہے۔ چنانچہ فوری طور پر اُردو مثال گھر کا دائرہ کار اِتنا وسیع نہیں کیا گیا۔

۳۔ جیسا کہ اوپر اشارہ کیا گیا، اُردو مثال گھر میں شامل متون کے الفاظ تصدیق شدہ اور معیاری اِملاء میں ہونا ضروری ہیں، تا کہ تحقیق کرنے والے لوگ زبان پر تحقیق کریں نہ کہ اِملاء کے مسائل میں اُلجھ کر رہ جائیں۔ انگریزی نے مختلف الاملاء اور مختلف الہجا الفاظ کے لیے یہ ترتیب اختیار کی ہے کہ کسی ایک ہجے کو معیاری مان کر لغت میں درج کر کے کام شروع کر دیا ہے جب کہ اختلافی اِملاء کو ساتھ میں رہنے دیا ہے۔ انگریزی کے کسی بھی لغت کو دیکھ لیجیے، ہر صفحے پر اِختلافی اِملاء والے کچھ نہ کچھ لفظ ضرور نظر آئیں گے۔ اُردو میں ابھی تک ایسا نہیں ہوا۔ اِملاء اور ہجا کے مسائل کبھی مکمل طور پر حل نہیں ہوں گے؛ اور کسی لفظ کا ایک ہی اِملاء ساری دنیائے اُردو قبول کر لے، یہ بھی ممکن نہیں۔ انگریزی میں بھی ایسا نہیں ہوسکا۔ لہٰذا اختلافِ اِملاء و ہجا کے ساتھ ہی کام کرنا ہوگا۔ [۶]

4.3: اُردو مثال گھر کے استعمالات

اُردو مثال گھر کا استعمال موجودہ دور کے محققین اور اہلِ معانی کی اوّلین ضرورت ہے۔ اِس کے استعمالات کے امکانات علی الکلیہ وہی ہیں جو دی بینک آف انگلش کے ہیں۔ بلکہ کئی جہات میں یہ امکانات متنوع تر ہیں۔ لسانی اور لسانیاتی تحقیق کے لیے عام استعمال میں آنے والی اُردو کی مثالیں بنانا اور جمع کرنا، اور اِس کام کو مستقل طور پر کیے جانا، بے حد ضروری ہے۔ اِس قسم کے ذخیرۂ امثال کی ضرورت، تیاری اور لائحۂ عمل پر راقم الحروف (ڈاکٹر حافظ صفوان محمد چوہان: ۲۰۰۷ء) نے الگ سے بحث کی ہے اور اِس کے لیے مآخذ اور اُردو کے لیے ہمارے ماحول کے مطابق خاص انداز میں کام کرنے کی بابت کئی جہات پر تفصیلی گفتگو کی ہے۔ اُردو مثال گھر اُردو کے لیے بنائے جانے والے کارپس اور لسانی تحقیقات کے لیے ریڑھ کی ہڈی کی حیثیت رکھتا ہے۔

اِس ترتیب پر یعنی اُردو مثال گھر کے ڈھب پر اگر اُردو کے متون ذخیرہ کیے جائیں، اور اِن کو مستقل طور پر تازہ (update) بھی کیا جاتا رہے، تو اُردو لسانی تحقیق کا نیا منظر نامہ سامنے آتا ہے۔ اور اگر سب پاکستانی زبانوں کے متون ذخیرہ کرنے کی بات بھی چل پڑے تو اُردو کو وہ مقام ملنے کی امید کی جاسکتی ہے جس کی وہ جائز طور پر حق دار ہے۔ یہ سب مثال گھر انے مل کر دنیا میں جاری لسانیاتی تحقیقات کو بہت کام کے بنیادی متون فراہم کرسکتے ہیں۔

مشین ریڈایبل اُردو میں مہیا ایسا اُردو مثال گھر دنیا بھر کے لسانیاتی محققین کا ایک مشترک خواب ہے جو اُن کے چیلنج، اُردو زبان، پر تحقیقات کے لیے سب سے وقیع knowledge-base ہوگا۔ ڈاکٹر اینڈریو ہارڈی (۲۰۰۳ء) نے بھی اُردو کے بارے میں یہی کہا ہے۔

4.4: اُردو مثال گھر اور دی بینک آف انگلش میں فرق

جیسا کہ اوپر عرض کیا گیا، مجوزہ اُردو مثال گھر اپنی بنت اور ڈھب میں دی بینک آف انگلش سے بہت لحاظ سے مختلف ہے۔ اختلاف کے کچھ وجوہ بھی اوپر گزر چکے ذیلی عنوان 4.1 میں ضمناً بیان کردیے گئے ہیں۔ ذیل میں ایک جدول کی شکل میں یہ اختلافات واضح کیے جارہے ہیں۔

سب سے بڑا اختلاف جو انگریزی اور اُردو مثال گھرانوں میں ہے، یہ ہے کہ دی بینک آف انگلش ہر متن کو اپنے پاس یعنی اپنے مال خانے میں رکھتا ہے اور جو بھی کام کراتا ہے، اِسی متن کی بنیاد پر کراتا ہے؛ جب کہ اُردو مثال گھر فوری طور پر اُردو کے ہر اُس متن کو بھی قبول کرے گا جو مشین ریڈایبل حالت میں کہیں سے بھی مہیا ہوسکے۔

ماخذ	دی بینک آف انگلش	اُردو مثال گھر
ادبی (نثری) متون	–	+
شاعری	–	+
اخبارات	+	+
اِنٹرنیٹ	+	+
انٹرویوز	+	–
ریڈیو	+	–
ٹی وی	+	–
عام بول چال	+	–

5: خاتمہ

زبان کی ساخت اور سائنسی اصولوں پر لسانیاتی تحقیق کے لیے زبان کے استعمالی نظائر کی بنیادی اہمیت ہے۔ اِس وقت کمپیوٹر پر مشین ریڈایبل حالت میں دنیا کی کئی زبانوں کا متن اِس مقصد کے لیے جمع کیا گیا ہے۔ اُردو میں ایسے متون کی ضرورت اور اہمیت، اور اِس اہم کام کے لیے لائحۂ عمل اِس مقالے میں ذکر کیا گیا ہے۔

★ ★ ★

۱۳/ جولائی ۲۰۰۸ء

۱۰/ رجب المرجب ۱۴۲۹ھ

اِس مقالے کی تیاری میں Collins COBUILD لغت کے پہلے ایڈیشن (1987)، دوسرے ایڈیشن (1995) اور Collins COBUILD English Dictionary for Advanced Learners, Major New Edition (2003) سے آزادانہ استفادہ کیا گیا ہے۔ ذیل میں دیے گئے صفحات کے حوالے اِسی آخرالذکر ماخذ کے ہیں۔ مزید معلومات کے لیے رک:

www.cobuild.collins.co.uk

ڈاکٹر حافظ صفوان محمد چوہان برٹش نیشنل کارپس (BNC) اور Collins COBUILD لغات کے روح و رواں، برمنگھم یونیورسٹی برطانیہ کے شعبۂ جدید لسانیات کے پروفیسر John McHardy Sinclair (1933–2007) سے طالب علمانہ رابطے میں رہے ہیں۔ اُنھیں بہت افسوس ہے کہ یہ مقالہ (اُردو اور انگریزی دونوں زبانوں میں) اُن کی وفات سے پہلے نہیں لکھا جاسکا۔ تحقیقی مقالے کوکسی کے نام معنوَن کرنے (Dedication) کی روایت نہیں ہے، پھر بھی یہ مقالہ اعترافِ کمال کے طور پر آنجہانی پروفیسر جان میک ہارڈی سنکلیئر کے نام معنون کیا جاتا ہے۔

حوالے:

حقی، شان الحق (۱۹۹۶ء) ''اُردو الفاظ میں چھوت چھات'' مقالہ، مشمولہ لسانی مسائل و لطائف، مقتدرہ قومی زبان، اسلام آباد۔ ص ۲۷

سرمد حسین، ڈاکٹر و مدیحہ اعجاز (۲۰۰۷ء) "Corpus Based Urdu Lexicon Development" مقالہ، مشمولہ CLT07، شعبۂ کمپیوٹر سائنس، پشاور یونیورسٹی۔ ص ۸۷

صفوان محمد چوہان، ڈاکٹر حافظ (۲۰۰۷ء) ''اُردو لغت (تاریخی اصول پر): بدلتے لسانی تناظر میں چند تجاویز''، مقالہ، مشمولہ جرنل آف ریسرچ، بہاء الدین زکریا یونیورسٹی ملتان، شمارہ-۱۲۔ ص ۲۸۲

ہارڈی، ڈاکٹر اینڈریو (۲۰۰۳ء) "The Computational Analysis of Morphosyntactic Categories in Urdu"، غیر مطبوعہ مقالہ پی ایچ ڈی۔ لینکاسٹر یونیورسٹی، برطانیہ۔ ص ۴۳

حواشی

۱۔ Written texts come from newspapers, magazines, fiction and non-fiction books, brochures, leaflets, reports, and letters. (p-xiv)

۲۔ Two-thirds of the corpus is made up of media language: newspapers, magazines, radio and TV. Ibid.

۳۔ Informal spoken language is represented by recordings of everyday casual conversation, meetings, interviews and discussions. Ibid.

۴۔ دی بینک آف انگلش کی بنیاد پر بننے والا پہلا لغت Collins Cobuild English Dictionary ہے جس کا پہلا ایڈیشن ۱۹۸۷ء میں سامنے آیا۔ سرد خانے میں ڈالنے کا یہ مطلب نہیں ہے کہ اِسے اب استعمال نہیں کیا جارہا، بلکہ مراد یہ ہے کہ اِسے عام طور سے استعمال میں نہیں لایا جارہا اور مخصوص مواقع کے لیے رکھا گیا ہے۔

۵۔ http://ieme.library.utoronto.ca/

۶۔ یہ بات ڈاکٹر خواجہ محمد زکریا صاحب نے مجلسِ ترقیِ ادب لاہور میں منعقدہ ''اُردو اِملاء کے مسائل'' کے عنوان سے ہونے والے ایک اجلاس میں کہی۔ (۲۰۰۷ء)

کتابیات

۱۔ حقی، شان الحق، ''لسانی مسائل و لطائف''، مقتدرہ قومی زبان، اسلام آباد۔ ۱۹۹۶ء

رسائل

۱۔ جرنل آف ریسرچ، بہاء الدین زکریا یونیورسٹی ملتان۔ شمارہ-۱۲ (۲۰۰۷ء)

۲۔ Proceedings of the Conference on Language & Technology (CLT07) at Bara Gali Summer Campus، شعبۂ کمپیوٹر سائنس، پشاور یونیورسٹی (۲۰۰۷ء)

تکنیکی مشاورت

۱۔ ڈاکٹر خواجہ محمد زکریا، سابق پرنسپل، اورینٹل کالج، جامعۂ پنجاب، لاہور۔

۲۔ ڈاکٹر عطش درانی، پراجیکٹ ڈائریکٹر، مرکزِ فضیلت برائے اُردو اطلاعیات، مقتدرۂ قومی زبان، اسلام آباد۔

۳۔ سید محمد ذوالکفل بخاری، لیکچرر شعبۂ انگریزی، ام القریٰ یونیورسٹی، مکہ مکرمہ زادہا اللہ تشریفا، سعودی عرب۔

۴۔ ڈاکٹر ظہیر احمد، اسسٹنٹ پروفیسر شعبۂ مصنوعی سیارات (Multimedia ILab)، سرے یونیورسٹی، سرے، برطانیہ۔

O < ------ > O