Developing Standards and Linguistic Resources for Computational Research in Pakistani Languages

www.CLE.org.pk

Sarmad Hussain
Professor
Center for language Engineering (CLE)
Al-Khawarizmi Institute of Computer Sciences (KICS)
University of Engineering and Technology (UET) Lahore
sarmad.hussain@kics.edu.pk

1

Need

ICTs promise significant socio-economic impact
Impact dependent on size of population which can use ICTs

180 Million citizens need access
66+ languages
10% understand English
58% literate
11% have access to computers
70% have access to mobile phones
ITU IDI: Pakistan ranked 127 of 155 nations

Human Language Technology necessary to bridge the gap

Languages of Pakistan

	Urdu	Punjabi	Sindhi	Pushto	Balochi	Saraiki	Others (60)
Total	**7.57**	**44.15**	**14.1**	**15.42**	**3.57**	**10.53**	**4.66**
Rural	1.48	42.51	16.46	18.06	3.99	12.97	4.53
Urban	20.22	47.56	9.20	9.94	2.69	5.46	4.93

Percent Population of Pakistan by Mother Tongue

Languages of Pakistan

Economic

Socio-cultural

	Urdu	Punjabi	Sindhi	Pushto	Balochi	Saraiki	Others (60)
Total	**7.57**	**44.15**	**14.1**	**15.42**	**3.57**	**10.53**	**4.66**
Rural	1.48	42.51	16.46	18.06	3.99	12.97	4.53
Urban	20.22	47.56	9.20	9.94	2.69	5.46	4.93

Percent Population of Pakistan by Mother Tongue

Languages of Pakistan

Economic

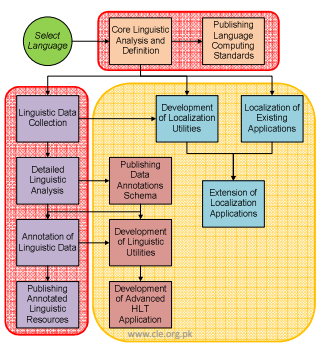
Socio-cultural

	Urdu	Punjabi	Sindhi	Pushto	Balochi	Saraiki	Others (60)
Total	**7.57**	**44.15**	**14.1**	**15.42**	**3.57**	**10.53**	**4.66**
Rural	1.48	42.51	16.46	18.06	3.99	12.97	4.53
Urban	20.22	47.56	9.20	9.94	2.69	5.46	4.93

Percent Population of Pakistan by Mother Tongue

Languages of Pakistan in Danger (UNESCO) (23)
Vulnerable
definitely endangered
severely endangered

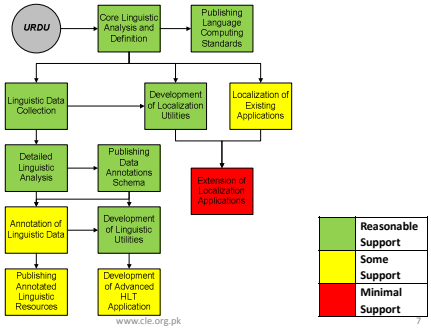
Development Process of Human Language Technology



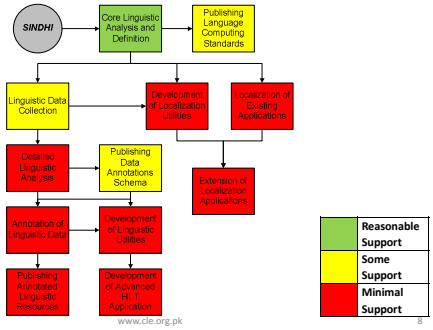
Standards & Linguistic Resources

Human Language Technology

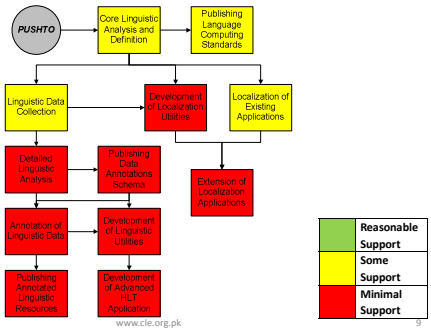
Status of Human Language Technology



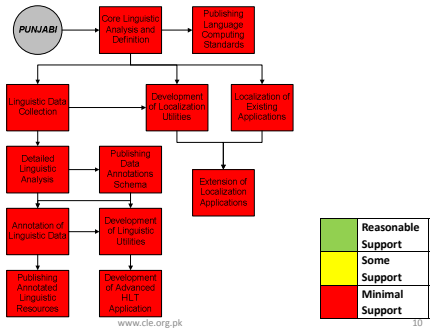
Status of Human Language Technology



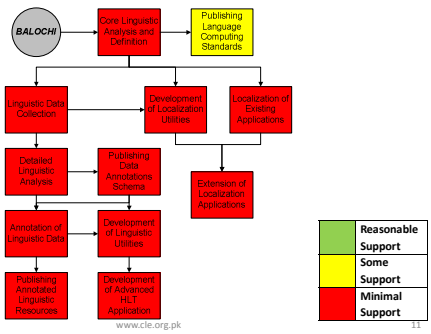
Status of Human Language Technology



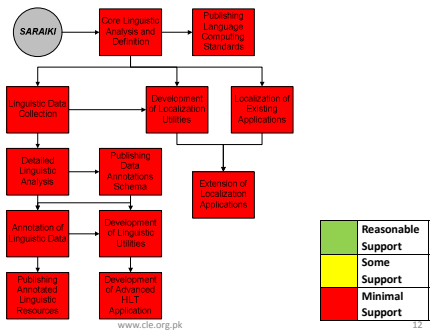
Status of Human Language Technology



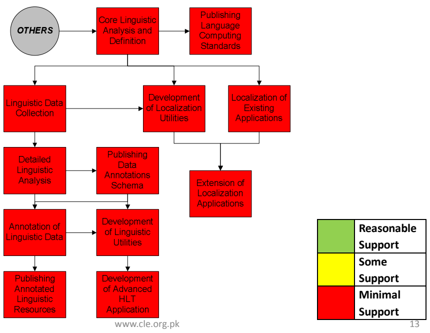
Status of Human Language Technology



Status of Human Language Technology



Status of Human Language Technology



Core Linguistic Analysis and Definition

- Linguistic details incomplete or unavailable
- Relevant cultural conventions rarely documented
- Need analysis of language and cultural conventions
  - Should be driven by linguists and community
- Precise definitions of relevant aspects
  - Involvement of technologists to iron out ambiguities

Core Linguistic Analysis and Definition

- Character-set
  - Alphabet – letters & aerab
  - Punctuation marks
  - Special symbols, etc.
- Sorting or collation of characters
- Cultural conventions for representing
  - Numbers
  - Time
  - Calendar
  - Currency
- Translation of terms for software interface

Standards

- Character Encoding – ISO 10646 - Unicode
- Locale – ISO 639 – ISO 3166 – Unicode-CLDR
  - Keyboard
  - Collation Element
- Localization Terminology

- Develop national standards
- Follow up with international standards

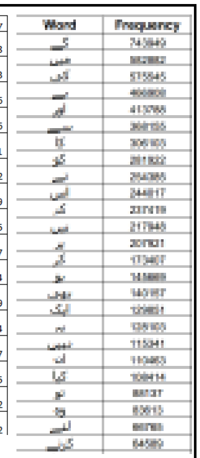
Lexica

- Word lists
- Annotated Lexical
  - Pronunciation
  - Agreement information (num, gender, case, respect)
  - Part of Speech (POS)
  - Translation
  - Sub-categorization frame/argument structure
  - Root
  - Morphological information
- Wordnet

Word Lists

- Word list – surface forms
- Lexeme list
- Closed class words
- Proper Names
- Frequency based word lists

Ligature	Frequency
ا	2885748
ر	1716383
و	1350586
د	896166
کے	746801
میں	582882
کی	581999
ہے	508375
ں	452167
کر	448244
کا	447729
نے	424334
سے	388097
کو	362535
ن	348302
ی	342672



Phonetic Lexicon

Attock	A TT A K
Bahawalpur	B A H AA V A L P U R
Bhakkar	B_H A K A R
Chakwal	T_SH A K V AA L
Faisalabad	F AY S L A B AA D_D
Gujranwala	G U D_ZZ R AAN V AA L AAN
Gujrat	G U D_ZZ R AA TT
Jhang	D_ZZ_H A NG
Jehlum	D_ZZ AE H L A M
Kasur	K A S UU R
Khushab	X UU SH AA B
Lahore	L AA H O R
Mianwali	M II AAN V AA L I
Multan	M U L T_D AA N
Rawalpindi	R AA V A L P I N DD II
Sahiwal	S AA H II V AA L
Sargodha	S A R G OO D_D_H AA
Sheikhupura	SH AE X UU P U R AA
Sialkot	S I A L K OO TT

Urdu ID	English ID	English Word	Category	Concept	Example	Synset
100853	6854161	English	Noun	انگریزوں کی بولی، انگلستان کی زبان	اسے عربی اور فارسی کے ساتھ ساتھ انگریزی پر بھی عبور حاصل تھا	انگریزی، انگلش
103196	617535	reading	Verb	مطالعہ کرنا	آج یہ کتاب پڑھنے کا موقع ملا ہے	دیکھنا، پڑھنا، مطالعہ کرنا
101288	4977386	intensity	Noun	شدید ہونے کی خصوصیت	ہوا اتنی زور سے چلی کہ خیمہ اڑگیا	جوش، زور، شدت، تندی
103802	2179176	give	Verb	روپیہ پیسہ امانت رکھنا	اس نے اپنے روپے اس کو سونپے کہ بعد میں لوں گا	دینا، سونپنا
102950	145831	certainly	Adverb	بے شک کسی شبہ کے بغیر	یقیناً اس کا کام مکمل ہوگیا ہوگا	بالیقین، بلا شبہ، بے شک، ضرور، یقیناً
104727	9771320	child	Noun	کم سن، جس کی عمر تھوڑی ہو، لڑکا، طفل، چھوکرا	اسٹیشن پر بم دھماکے کا زخمی سات سالہ بچہ اسپتال میں دم توڑ گیا	بچہ، طفل

Text Corpora

- Text corpus
  - Size
  - Time
  - Genre
- CLE Urdu Digest corpus
  - Size
    - 100k
    - 500k
    - 1M
  - Time
    - 2003-2012
  - Genres
    - Balanced

Category	Sub-category	Percentages
1. Informational (80%)		
a) Informal	Letters	10%
	Interviews	10%
a) Formal		
	Press	8%
	Religion	8%
	Sports	8%
	Culture (travel, history)	8%
	Entertainment	4%
	Health	8%
	Science (education, technology)	16%
1. Imaginative (20%)		
	Short Stories	8%
	Translation of foreign literature	4%
	Novels	4%
	Book reviews	4%

Text Corpora

- State of the Art
  - English: Web 1T 5-gram Version 1
  - Published by Google
    - Number of tokens: 1,024,908,267,229
    - Number of sentences: 95,119,665,584
    - Number of unigrams: 13,588,391
    - Number of bigrams: 314,843,401
    - Number of trigrams: 977,069,902
    - Number of fourgrams: 1,313,818,354
    - Number of fivegrams: 1,176,470,663

English: Web 1T 5-gram Version 1

- Sample Trigrams and Frequencies
  - ceramics community for 61
  - ceramics companies . 53
  - ceramics companies consultants 173
  - ceramics company ! 4432
  - ceramics company , 133
  - ceramics company . 92
  - ceramics company 41
  - ceramics company facing 145
  - ceramics company in 181
  - ceramics company started 137
  - ceramics company that 87
  - ceramics component (76
  - ceramics composed of 85
  - ceramics composition as 41
  - ceramics computer graphics 51
  - ceramics computer imaging 52
  - ceramics consist of 92

Annotated Corpora

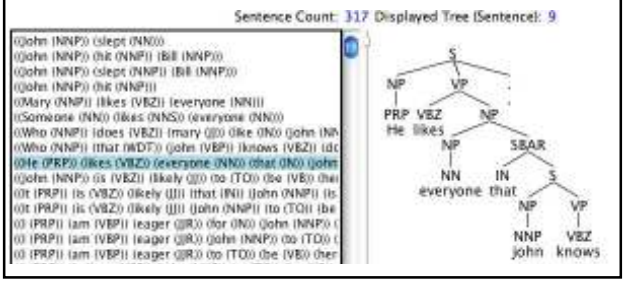
- Text Corpus
  - POS tagged
  - Chunk tagged
  - Sense tagged
- Tree Bank
  - Syntax
  - Grammatical relations
  - Dependency relations

دنیا کا ہر فرد کامیابی کا آرزومند ہے۔ ناکامی سے سب گھبراتے ہیں۔ عزت، دولت، راحت اور عافیت کی زندگی کے بھی شیدائی ہیں۔ لیکن اصل کامیابی کیا چیز ہے؟ اور حقیقی عزت وراحت کس طرح نصیب ہوتی ہے؟ اس بھید سے بہت کم لوگ واقف ہیں۔ اگر آپ حقیقی کامیابی کے گر جاننا چاہتے ہیں تو ڈاکٹر زاہد منیر عامر کی تازہ تصنیف آئینہ کردار پڑھیے۔ ۱۱۲ صفحوں کی اس کتاب کا ایک ایک حرف بصیرت کے دریچے کھولنے پر مامور ہے۔

دنیا/NN کا/PSP ہر/JJ فرد/NN کامیابی/NN کا/PSP آرزومند/NN ہے/VBF ۔/PU ناکامی/NN سے/PSP سب/JJ گھبراتے/VBF ہیں/AUXT ۔/PU عزت/NN ،/PU.NN دولت/NN ،/PU.NN راحت/NN اور/CC عافیت/NN کی/PSP زندگی/NN کے/PSP بھی/PRP شیدائی/NN ہیں/VBF ۔/PU لیکن/SC اصل/JJ کامیابی/NN کیا/RB چیز/NN ہے/VBF ؟/PU اور/CC حقیقی/JJ عزت/NN ،/CC راحت/NN اس/PDM طرح/NN نصیب/NN ہوتی/VBF ہے/AUXT ؟/PU اس/PDM بھید/NN سے/PSP بہت/Q کم/Q لوگ/NN واقف/NN ہیں/VBF ۔/PU اگر/SCP آپ/PRP حقیقی/JJ کامیابی/NN کے/PSP خواہاں/VBI ہیں/VBF تو/AUXT ،/SC ذاتی/NN زاویہ/NNP منیر/NNP عامر/NNP کی/PSP تازہ/JJ تصنیف/PU ،/PU آئینہ/NN کردار/NN پڑھیے/VBF ۔/PU ۱۱۲/CD صفحوں/NN کی/PSP اس/PDM کتاب/NN کا/PSP ایک/CD ایک/CD حرف/NN بصیرت/NN کے/PSP دریچے/NN کھولنے/VBI پر/PSP مامور/NN ہے/VBF ۔/PU

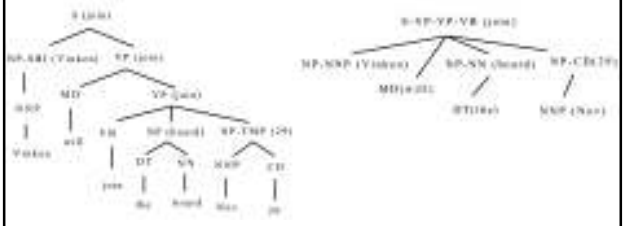

```
( (S
   (NP Battle-tested industrial managers
       here)
    always
   (VP buck
       up
       (NP nervous newcomers)
       (PP with
           (NP the tale
               (PP of
                   (NP (NP the
                           (ADJP first
                                 (PP of
                                     (NP their countrymen)))
                           (S (NP *)
                              to
                              (VP visit
                                  (NP Mexico))))
                       ,
                       (NP (NP a boatload
                               (PP of
                                   (NP (NP warriors)
                                       (VP-1 blown
                                           ashore
                                           (ADVP (NP 375 years)
                                                 ago)))))
                           (VP-1 *pseudo-attach*)))))))
  .)
```


- ((S (NP Battle-tested industrial managers here) always (VP buck up (NP nervous newcomers) (PP with (NP the tale (PP of (NP (NP the (ADJP first (PP of (NP their countrymen))) (S (NP *) to (VP visit (NP Mexico)))) , (NP (NP a boatload (PP of (NP (NP warriors) (VP-1 blown ashore (ADVP (NP 375 years) ago))))) (VP-1 *pseudo-attach*)))))))) .)



Arabic (Quran)

Phrase Structure vs. Dependency Treebanks



Sense Tagged Corpus

Text Corpora

- Parallel Text corpus
  - Language
  - Size
  - Genre
  - Time
  - Alignment
- Spelling Error corpus

Parallel Corpus

- English

Pierre Vinken , 61 years old , will join the board as a nonexecutive director Nov. 29 .

Mr. Vinken is chairman of Elsevier N.V. , the Dutch publishing group .

- Urdu

NNP>>پیئرونکن JJ>>سالہ CD>>اکسٹھ
CM>>کو NNP>>نومبر CD>>۲۹
JJ>>نان ایگزیکٹو RB>>بطور
CM>>میں NN>>بورڈ NN>>ڈائریکٹر
VBL>>ہوں JJ>>شامل
AUXT>-<SM>>گے

NNP>>ونکن NNP>>مسٹر
CM>>کے NNP>>ہالینڈ
NN>>گروپ JJ>>اشاعتی
NNP>>ایلس ویئر
CM>>کے NNPC>>این۔وی۔
SM>>۔ VBT>>ہیں NN>>چیئرمین

- Nepali

६१<CD> वर्षीय<JJ>
पियरे<NNP>
भिन्केन<NNP>
नोभेम्बर<NNP> २९<CD>
बाट<POP>
सल्लाहकार<NN>को<PKO
> रूप<NN> मा<POP>
सञ्चालक<NN>
समिति<NN>मा<POP>
आउनुहुनेछ<VBX> ।<YF>

श्री<NN> भिन्केन<NNP>
डच<NNP> प्रकाशन<NN>
समूह<NN>
एल्सेभियर<NNP>
एन.भी.<FB> को<PKO>
अध्यक्ष<NN>
हुनुहुन्छ<VBF> ।<YF>

Corpora

- Speech Corpus
  - Duration
  - Speakers
  - Channels
  - Environments
  - Contents
  - Accent(s)
  - Gender
  - Style (read/spontaneous)
  - Vocabulary

Annotated Corpora

- Speech Corpus
  - Textual transcription
  - Segmental IPA transcription
  - Aligned transcription
  - Syllable boundaries
  - Stress
  - Intonation
  - POS

Speech Tagging Schema

p1ˆp2-p3+p4=p5@p6 p7
/A:a1 a2 a3
/B:b1-b2-b3@b4-b5&b6-b7#b8-b9$b10-b11!b12-b13;b14-b15|b16
/C:c1+c2+c3
/D:d1 d2 /E:e1+e2@e3+e4&e5+e6#e7+e8 /F: f1 f2
/G:g1 g2 /H:h1=h2ˆh3=h4|h5 /I:i1 i2

p_1	the phoneme identity before the previous phoneme
p_2	the previous phoneme identity
p_3	the current phoneme identity
p_4	the next phoneme identity
p_5	the phoneme after the next phoneme identity
p_6	position of the current phoneme identity in the current syllable (forward)
p_7	position of the current phoneme identity in the current syllable (backward)

a_1	whether the previous syllable stressed or not (0: not stressed, 1: stressed)
a_2	whether the previous syllable accented or not (0: not accented, 1: accented)
a_3	the number of phonemes in the previous syllable
b_1	whether the current syllable stressed or not (0: not stressed, 1: stressed)
b_2	whether the current syllable accented or not (0: not accented, 1: accented)
b_3	the number of phonemes in the current syllable
b_4	position of the current syllable in the current word (forward)
b_5	position of the current syllable in the current word (backward)
b_6	position of the current syllable in the current phrase (forward)
b_7	position of the current syllable in the current phrase (backward)
b_8	the number of stressed syllables before the current syllable in the current phrase
b_9	the number of stressed syllables after the current syllable in the current phrase
b_{10}	the number of accented syllables before the current syllable in the current phrase
b_{11}	the number of accented syllables after the current syllable in the current phrase
b_{12}	the number of syllables from the previous stressed syllable to the current syllable
b_{13}	the number of syllables from the current syllable to the next stressed syllable
b_{14}	the number of syllables from the previous accented syllable to the current syllable
b_{15}	the number of syllables from the current syllable to the next accented syllable
b_{16}	name of the vowel of the current syllable
c_1	whether the next syllable stressed or not (0: not stressed, 1: stressed)
c_2	whether the next syllable accented or not (0: not accented, 1: accented)
c_3	the number of phonemes in the next syllable

d_1	gpos (guess part-of-speech) of the previous word
d_2	the number of syllables in the previous word
c_1	gpos (guess part-of-speech) of the current word
c_2	the number of syllables in the current word
c_3	position of the current word in the current phrase (forward)
c_4	position of the current word in the current phrase (backward)
c_5	the number of content words before the current word in the current phrase
c_6	the number of content words after the current word in the current phrase
c_7	the number of words from the previous content word to the current word
c_8	the number of words from the current word to the next content word
f_1	gpos (guess part-of-speech) of the next word
f_2	the number of syllables in the next word
g_1	the number of syllables in the previous phrase
g_2	the number of words in the previous phrase
h_1	the number of syllables in the current phrase
h_2	the number of words in the current phrase
h_3	position of the current phrase in this utterance (forward)
h_4	position of the current phrase in this utterance (backward)
h_5	TOBI endtone of the current phrase
i_1	the number of syllables in the next phrase
i_2	the number of words in the next phrase

Speech Annotated File Sample

p1^p2-p3+p4=p5@p6 p7
/A:a1 a2 a3
/B:b1-b2-b3@b4-b5&b6-b7#b8-b9$b10-b11!b12-b13;b14-b15|b16
/C:c1+c2+c3
/D:d1 d2 /E:e1+e2@e3+e4&e5+e6#e7+e8 /F: f1 f2
/G:g1 g2 /H:h1=h2^h3=h4|h5 /I:i1 i2

1807500 3432500 pau^pau-ao+th=er@1_2/A:0_0_0/B:1-1-2@1-2&1-7#1-4$1-3!0-2;0-4|ao/C:0+0+1/D:0_0/E:content+2@1+5&1+2#0+3/F:in_1/G:0_0/H:7=5^1=2|L-L%/I:7=3/J:14+8-2

3432500 4557500 pau^ao-th+er=ah@2_1/A:0_0_0/B:1-1-2@1-2&1-7#1-4$1-3!0-2;0-4|ao/C:0+0+1/D:0_0/E:content+2@1+5&1+2#0+3/F:in_1/G:0_0/H:7=5^1=2|L-L%/I:7=3/J:14+8-2

Corpora

- Document Image Corpus
  - Resolution (DPI)
  - Color/Grayscale/BW
  - Print quality
  - Paper quality
  - Paper transparency
  - Genre
  - Publisher

Annotated Corpora

- Document Image Corpus
  - Textual transcription
  - Line
  - Text Area
  - Figures and page frames
  - Heading
  - Header and Footer
  - Main bodies
  - Diacritics and their association

جبکہ عربی اور کسی حد تک اردو زبان بھی وہ بول لیتا تھا اگرچہ وہ ان زبانوں میں لکھ یا پڑھ نہیں سکتا تھا۔

اس کی اس خصوصیت کے انکشاف نے یکدم ہی اسے اپنی کلاس اور کسی حد تک اسکول میں پاپولر کر دیا تھا۔ لینگویج کی کلاس میں ایک دن اتفاقاً اس کے ٹیچر کو اس بات کا پتا چلا تھا کہ وہ جرمن زبان پر بھی دسترس رکھتا ہے۔

''تو ڈینیل تم دو زبانوں کو استعمال کر سکتے ہو؟'' ٹیچر نے اسے سراہتے ہوئے کہا۔

''دو نہیں چار..... عربی اور اردو بھی۔ اگرچہ میں انھیں لکھ پڑھ نہیں سکتا مگر اس میں گفتگو کر سکتا ہوں۔'' مدہم آواز میں کہے گئے جملے نے یک دم ہی پوری کلاس کو سر موڑ کر اس کی طرف متوجہ ہونے پر مجبور کر دیا۔ ان کی آنکھوں میں حیرت کے ساتھ ساتھ ستائش بھی تھی۔

''چار زبانیں..... زبردست۔ مگر چار زبانیں کیسے؟ میرا مطلب ہے عربی اور اُردو؟''

''میرے ڈیڈی بہت عرصے سے مڈل ایسٹ اور ایشیا کے ممالک میں کام کرتے رہے ہیں، میری پیدائش بھی مراکش میں ہوئی اس لیے عربی بولنا آ گئی اور پچھلے دو سال سے ہم لوگ انڈیا میں تھے۔ وہاں لوگوں سے بات چیت انگلش یا اردو میں ہی ہوتی تھی، اس لیے اس کو بھی استعمال کرنا آ گیا۔''

113 ایمان، اُمید اور محبت

جبکہ عربی اور کسی حد تک اردو زبان بھی وہ بول لیتا تھا اگرچہ وہ ان زبانوں میں لکھ یا پڑھ نہیں سکتا تھا۔

اس کی اس خصوصیت کے انکشاف نے یکدم ہی اسے اپنی کلاس اور کسی حد تک اسکول میں پاپولر کر دیا تھا۔ لینگویج کی کلاس میں ایک دن اتفاقاً اس کے ٹیچر کو اس بات کا پتا چلا تھا کہ وہ جرمن زبان پر بھی دسترس رکھتا ہے۔

''تو ڈینیل تم دو زبانوں کو استعمال کر سکتے ہو؟'' ٹیچر نے اسے سراہتے ہوئے کہا۔

''دو نہیں چار..... عربی اور اردو بھی۔ اگرچہ میں انھیں لکھ پڑھ نہیں سکتا مگر اس میں گفتگو کر سکتا ہوں۔'' مدہم آواز میں کہے گئے جملے نے یک دم ہی پوری کلاس کو سر موڑ کر اس کی طرف متوجہ ہونے پر مجبور کر دیا۔ ان کی آنکھوں میں حیرت کے ساتھ ساتھ ستائش بھی تھی۔

''چار زبانیں..... زبردست۔ مگر چار زبانیں کیسے؟ میرا مطلب ہے عربی اور اُردو؟''

''میرے ڈیڈی بہت عرصے سے مڈل ایسٹ اور ایشیا کے ممالک میں کام کرتے رہے ہیں، میری پیدائش بھی مراکش میں ہوئی اس لیے عربی بولنا آ گئی اور پچھلے دو سال سے ہم لوگ انڈیا میں تھے۔ وہاں لوگوں سے بات چیت انگلش یا اردو میں ہی ہوتی تھی، اس لیے اس کو بھی استعمال کرنا آ گیا۔''

جبکہ عربی اور کسی حد تک اردو زبان بھی وہ بول لیتا تھا اگرچہ وہ ان زبانوں میں لکھ یا پڑھ نہیں سکتا تھا۔
اس کی اس خصوصیت کے اِنکشاف نے یکدم ہی اسے اپنی کلاس اور کسی حد تک اسکول میں پاپولر کر دیا تھا۔ لینگویج کی کلاس میں ایک دن اتفاقاً اس کے ٹیچر کو اس بات کا پتا چلا تھا کہ وہ حرمس زبان پر بھی دسترس رکھتا ہے۔

''تو ڈنیل تم دو زبانوں کو استعمال کر سکتے ہو؟'' ٹیچر نے اسے سراہتے ہوئے کہا۔

''دو نہیں چار..... عربی اور اردو بھی۔ اگرچہ میں انھیں لکھ پڑھ نہیں سکتا مگر اس میں گفتگو کر سکتا ہوں۔'' مدھم آواز میں کہے گئے جملے نے یک دم ہی پوری کلاس کو سرموڑ کر اس کی طرف متوجہ ہونے پر مجبور کر دیا۔ ان کی آنکھوں میں حیرت کے ساتھ ساتھ ستائش بھی تھی۔

''چار زبانیں..... زبردست۔ مگر چار زبانیں کیسے؟ میرا مطلب ہے عربی اور اردو؟''

''میرے ڈیڈی بہت عرصے سے مڈل ایسٹ اور ایشیا کے ممالک میں کام کرتے رہے ہیں، میری پیدائش بھی مراکش میں ہوئی اس لیے عربی بولنا آ گئی اور پچھلے دو سال سے ہم لوگ انڈیا میں تھے۔ وہاں لوگوں سے بات چیت انگلش یا اردو میں ہی ہوتی تھی، اس لیے اس کو بھی استعمال کرنا آ گیا۔''

ایمان، اُمید اور محبت 113
جبکہ عربی اور کسی حد تک اردو زبان بھی وہ بول لیتا تھا اگرچہ وہ ان زبانوں میں لکھ یا پڑھ نہیں سکتا تھا۔
اس کی اس خصوصیت کے اِنکشاف نے یکدم ہی اسے اپنی کلاس اور کسی حد تک اسکول میں پاپولر کر دیا تھا۔ لینگویج کی کلاس میں ایک دن اتفاقاً اس کے ٹیچر کو اس بات کا پتا چلا تھا کہ وہ حرمس زبان پر بھی دسترس رکھتا ہے۔
''تو ڈنیل تم دو زبانوں کو استعمال کر سکتے ہو؟'' ٹیچر نے اسے سراہتے ہوئے کہا۔
''دو نہیں چار..... عربی اور اردو بھی۔ اگرچہ میں انھیں لکھ پڑھ نہیں سکتا مگر اس میں گفتگو کر سکتا ہوں۔'' مدھم آواز میں کہے گئے جملے نے یک دم ہی پوری کلاس کو سرموڑ کر اس کی طرف متوجہ ہونے پر مجبور کر دیا۔ ان کی آنکھوں میں حیرت کے ساتھ ساتھ ستائش بھی تھی۔
''چار زبانیں..... زبردست۔ مگر چار زبانیں کیسے؟ میرا مطلب ہے عربی اور اُردو؟''
''میرے ڈیڈی بہت عرصے سے مڈل ایسٹ اور ایشیا کے ممالک میں کام کرتے رہے ہیں، میری پیدائش بھی مراکش میں ہوئی اس لیے عربی بولنا آ گئی اور پچھلے دو سال سے ہم لوگ انڈیا میں تھے۔ وہاں لوگوں سے بات چیت انگلش یا اردو میں ہی ہوتی تھی، اس لیے اس کو بھی استعمال کرنا آ گیا۔''