

# The Micro-structure of an Encoding Dictionary

Mahmood Ahmad

*The paper aims to discuss the format of the entry in an Urdu-English dictionary geared to meet the encoding needs of learners of English in Pakistan. The content and structure of the existing Urdu-English dictionaries suffer from serious inadequacies. Drawing upon various schemes proposed for the construction of an entry in a bilingual dictionary schemes, the desiderata of an entry in an Urdu-English dictionary are proposed. The paper stresses on the need to improve reference skills of the users so that they can make full use of the information contained in the dictionary entry.*

**Keywords:** dictionary, encoding, Urdu

## 1. Introduction

The present article is an attempt to highlight the user-perspective in lexicography in Pakistan with special reference to the encoding needs of advanced learners of English in Pakistan. It is hoped that it would lead to an improvement in the quality of the Urdu-English dictionaries which otherwise fall short of the needs of the users. This is due to the fact that the compilers/publishers of dictionaries in Pakistan adopt an ivory tower approach. They seldom take into account the needs of the users. This indifference on the part of the compilers/publishers is detrimental both to them and to the users since it undermines the usefulness (effectiveness) as well as the usability (i.e. learners' preference) of the dictionaries. What is all the more regrettable is the fact that dictionary research is not considered an area worthy of academic pursuit in the characteristic Pakistani context. This apathy has led to the maintenance of status quo.

### 1.1. *The format of an entry in the existing dictionaries*

It seems advisable to begin by spelling out what the structure of an entry in the Urdu-English dictionaries looks like:

- i. *Orthographic form* of the head word in Persian script.
- ii. *Transliteration in the Roman script*: The headword is transliterated in Roman script. Interestingly, the Roman script does not adequately represent Urdu letters in many cases. For example, there is no Roman letter to represent the following eight Urdu letters: ا, ب, پ, ت, ث, ج, ح, خ. Similarly, letter 'z' is used to represent four Urdu letters ز, ذ, ظ, ض. Consequently, the difference among Urdu words such as ظن (zan) i.e. 'opinion' vs. زن (zan) i.e. 'woman' is blurred; Letter 's' is used to represent three Urdu letters س, ص, ش, and Letter 'h' is used to represent two Urdu letters ه, ح.
- iii. *Part-of-speech tag* – e.g. noun, adjective, verb, etc.
- \*iv. *Subclass tag* – e.g. gender and/or number of a noun.
- v. *Equivalents in English*: In most cases, the entry contains a string of translation equivalents - with some rudimentary punctuational structuring (commas versus

semicolons, for example, or numbers such as (1), (2) etc. However, the discrimination among different senses is hardly made clear in an adequate manner. Iannucci's (1957) remarks about major weakness - the hamartia -of the bilingual dictionaries 'even in the best dictionaries ... meaning discrimination is very spotty' holds for Urdu-English dictionaries as well.

- vi. *Word-combinations* e.g. collocations and idioms.
- vii. *Origin tag* – Arabic, Hindi, Sanskrit, Persian, English, etc.
- \*viii. *Cross-references* to other entries.
- \*ix. *Abbreviations*
- x. *Encyclopaedic notes*

*Note:* Items marked above with an asterisk do not appear in all the dictionaries.

## 2. Theoretical Considerations

Various schemes have been put forward to improve the quality of bilingual dictionaries. In section (§2), I shall describe these theories briefly. The aim is to draw upon these theories in order to lay down the format of the entry in an Urdu-English dictionary which can address the encoding needs of advanced learners:

Shcherba (1940) conceived the idea of an 'active' or encoding dictionary. He proposed the 'translation principle' to this end which implied two things: (a) the dictionary should give the real translation of the L1 lexical item into the L2 equivalent, preferably inserted within a phrase (b) in cases of polysemy, the senses should be disambiguated as to how and when to use which translation.

According to Haas (1962) a truly ideal bilingual dictionary 'would anticipate every conceivable need of the prospective user'. So it would:

- i. provide for each word or expression in the source language just the right translation in the target language
- ii. contain all the words, locutions, circumlocutions, and idioms that any user might ever want to look up
- iii. contain all the inflectional, derivational, syntactic, and semantic information that any user might ever need
- iv. contain information on all levels of usage
- v. contain all proper names and any other name that any user might want to look up
- vi. contain special vocabulary items of all the registers
- vii. contain all the information on spelling aids and alternative spellings
- viii. include all the information on exact pronunciation
- ix. should be oriented to speakers of both languages
- x. should be well-adapted to the purposes of machine translation as it is to human translation
- xi. should be compact
- xii. include illustrations to picture unfamiliar items

Haas was aware of the fact that her vision of the bilingual dictionary was utopian. She herself pointed out that a number desiderata proposed were either impracticable, e.g. all-

inclusiveness, adaptability to machine as well as human translation, orientation to speakers of both languages or not indispensable e.g. inclusion of illustrations. She also pointed toward the misconception that a single dictionary can cater for the reference needs of speakers of the two languages. She makes an interesting observation in this regard: 'Thinking they are preparing a dictionary for speakers of both languages, they may end up producing a dictionary which is not as useful as it should be to speakers of both languages'.

Zgusta's (1971) criteria as summarized in Klapicova (2005) for the construction of an entry in a bilingual dictionary are as follows:

- i. the presence of the entry word in its canonical form
- ii. grammatical information
- iii. indication of pronunciation
- iv. equivalents in the target language in their canonical form
- v. indication of the whole lexical meaning of the entry word by partial equivalents of the target language
- vi. encyclopaedic information
- vii. etymology of the entry word
- viii. the lexicalised and the verbatim meaning of different morphemic and word combinations.

Zgusta stresses that regarding the information to be included in a bilingual dictionary the most powerful factor is the intention of the dictionary. He proposes several devices such as the addition of labels, glosses and examples in order to disambiguate the multiple meanings of the translation equivalents of the entry-word (cf. *ibid*).

According to Bartholomew and Schoenhals (1983: 25), an entry should have the following general structure:

- i. Entry form in the vernacular (classified in a psychologically natural manner)
- ii. Enumeration of the different senses (classified according to translation equivalent)
  - grammatical part of speech
  - gloss (translation equivalent(s), or explanation of usage)
  - qualifying comment on the gloss (where necessary)
  - diagnostic illustrative of the entry word in context
  - natural translation of the sentence into the target language
- iii. Essential linguistic information on conjugation type, principal parts of verbs, etc.
- iv. Cross references
  - iv a. synonym or antonym references
  - iv b. Cross references to elements of a compound or a basic stem of derivatives
- v. Subentries that relate the word to other lexical units (entries sharing the same root and other words semantically related to them)
  - v a. derivatives: nouns from verbs, transitive verbs from intransitive, etc.
  - v b. Compound words utilizing the same root
  - v c. Idioms, idiomatic phrases.

They give the following graphic representation of the lexical unit with several possible sense discriminations (ibid.: 12):

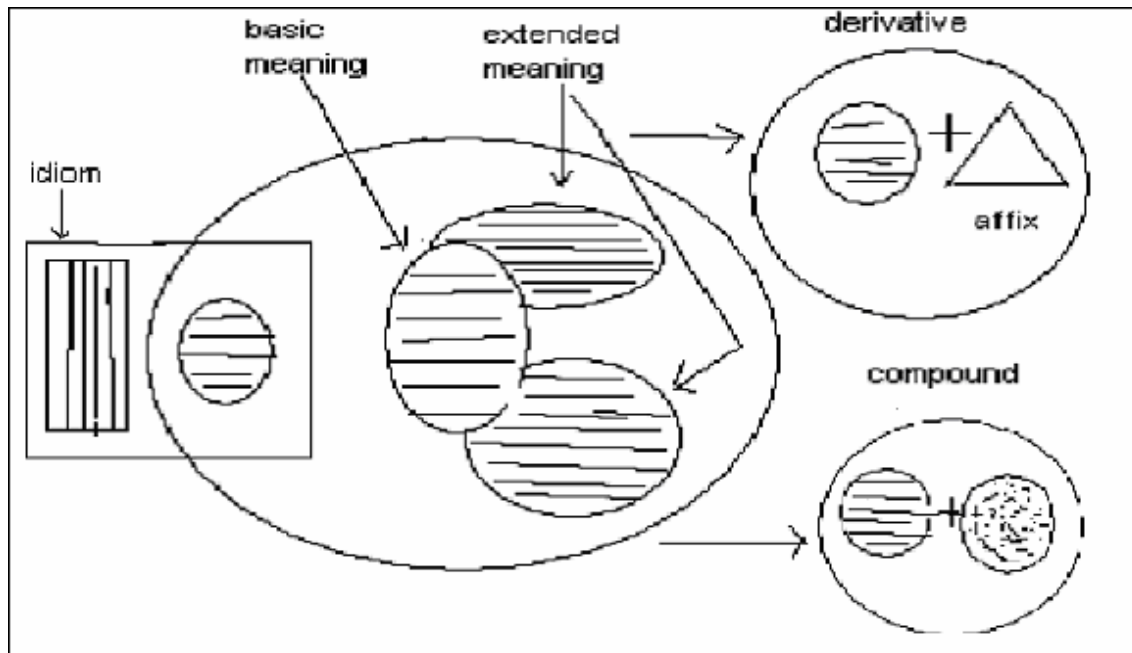


Figure 1 Bartholomew & Shoenhal's (1983) model of the lexical unit and related words

The schema proposed by Bartholomew and Shoenhals is seminal, especially due to the stress it lays on sense discrimination. They point out that the native speaker can easily perceive the discrete senses of a lexical item in specific contexts. However, the task of sense discrimination is an arduous one for foreign learners. According to them,

The rationale for including sense discrimination in a bilingual dictionary is the basic principle that very seldom can one word in the target language cover exactly the same area of meaning as the source language word for which it is designated the gloss. The common area of meaning may be great or small but in only a small proportion of the entries is it exactly the same (ibid: 70).

Sekaninova (1993) cited in Klapicova (2005) emphasises seven parameters for the construction of an entry in a bilingual dictionary:

- i. phonetic information
- ii. grammatical component
- iii. stylistic parameter
- iv. lexical equivalence
- v. lexical stability
- vi. lexical-semantic connection potentiality
- vii. context applicability.

The scheme put forward by Sekaninova (1993) is discerning. The use of various ingenious devices to disambiguate the multiple meanings of translation equivalents is believed to contribute to improve the effectiveness of the dictionaries meant for production (cf. Zgusta 1971; Bartholomew and Schoenhals 1983; Laufer 1995; and Laufer and Levitzky-Aviad 2006).

Laufer (1995) suggested to create a special L1-L2-L2 (i.e. a semi-bilingual) dictionary for production. The proposed dictionary should include four elements:

- i. *L1-L2 translations*. The bilingual dictionary plus will start with an L1 lexical item which will be first translated into L2. In case of polysemy, all the translation equivalents will be provided.
- ii. *Information about the L2 translation options*. Information about the L2 translation will consist of words' phonological, grammatical and semantic specifications (all provided in L2), followed by a definition, and examples of use.
- iii. *Semantically related words*. This part resembles a thesaurus. Words semantically related to the L2 equivalent of the looked up L1 word will be listed with their definitions and L1 translations. This component will enable the user to select the most suitable word out of several words in the semantic area.
- iv. *Additional meanings of the L2 translation*. Additional L1 meanings of the L2 translation will be provided in those entries where the English translation equivalent is polysemous or homonymous while the L1 word is not. For example, the Hebrew word 'MOFSHAT' is 'abstract' (the opposite of 'concrete'). But 'abstract' has also an additional meaning 'summary'. This meaning has nothing to do with MOFSHAT. This additional meaning could be translated and illustrated. This component will prevent the user from assuming that each time 'abstract' appears, it will mean MOFSHAT.

Laufer and Levitzky-Aviad (2006) renamed it *bilingual dictionary plus* since it was to contain more information than a standard bilingual dictionary. They carried out an empirical study to check its effectiveness. The results confirmed the value of the proposed dictionary in terms of usefulness as well as learners' preference. However, it was revealed that the third and fourth components of the proposed entry i.e. thesaurus information and additional meanings of the L2 translation were considered to be unnecessary by the subjects.

In spite of these findings, Laufer and Levitzky-Aviad (2006) maintain that the thesaurus component is very much relevant to free writing where learners often look for a precise shade of meaning, or for a synonym in order not to repeat the same word. Moreover, the fourth component i.e. additional meanings of the L2 translation may serve as a basis for vocabulary exercises distinguishing between the lexicalisation of various concepts in learners' L1 and the foreign language (*ibid*).

Haensch and Omenaca (2004: 240) as cited in Klapicova (2005) laid down the following structure of an entry in a general bilingual dictionary:

- i. statement of the lemma
- ii. orthographic variants
- iii. parts of speech
- iv. pronunciation and accentuation

- v. syntactic information such as gender, irregular plural forms, verb inflections and ideally, verbal valencies, etc.
- vi. lexicographical marks
- vii. usage restrictions
- viii. examples of application
- ix. in very few cases: illustrations.

We see that, like various other theorists, they also consider grammatical information, labels, examples and illustrations to be useful features which can ensure the effectiveness of the dictionary for the learners. These are the devices which have been used efficiently in the compilation of EFL dictionaries to improve their effectiveness (Stein 2002).

Warsch (2005) gives an account of the innovations Langenscheidt publishing house introduced into bilingual learner's lexicography to improve *Didaktisiertes Schulwörterbuch* (DSW). The proposed dictionary was to have the following design features:

- i. The number of headwords was to be reduced in favour of an extended user-oriented microstructure.
- ii. Instead of swung dashes, each compound or derivative, each headword that is repeated in a phrase or example sentence, was to be given in full.
- iii. Each headword was to be given full lemmatization and nesting of compounds or derivatives was to be avoided.
- iv. The colour of the head words was to be blue.
- v. Entries that changed considerably due to the new German orthography were to be given in a blue frame.
- vi. Cryptic abbreviations for grammar were to be avoided.
- vii. Usage notes were to be given.
- viii. Contextualized, lively illustrations were to be included.
- ix. Info boxes containing information on language phenomena, cultural or political topics etc were to be given.

Warsch (2005) claims that, due to these innovations, a major breakthrough in bilingual learners' lexicography took place. In the light of the feedback from other European dictionary publishers, Langenscheidt established *Power* family, including a range of bilingual learner's dictionaries for the main European languages. It published *Power Wörterbuch Französisch* (French) in 1999; *Power Wörterbuch Spanisch* (Spanish) in 2004 and *Power Wörterbuch Italienisch* (Italian) in 2005.

The above schemes have valuable insights to offer in connection with the micro-structure of Urdu-English dictionaries tailored to meet the encoding needs of the learners. In the following section, drawing upon the above schemes, the desiderata of the micro-structure of an entry in an encoding dictionary are laid down.

### 3. The proposed format of an entry in an Urdu-English learner's dictionary

- i. *Head word*: Orthographic form of the head word in Persian script.
- ii. *Cross-reference*: A word or symbol to refer the user to relevant information covered in other entries
- iii. *Variant spellings and forms* of the headwords, e.g. تک & تک 'tak' & talak' (i.e. 'up to, till')
- iv. *Pronunciation*: As noted in (1.1.2), the Roman letters do not adequately represent Urdu letters. Hence, it is proposed that the International Phonetic Alphabet should be used to help resolve inconsistencies/inadequacies like the ones mentioned above through the use of special letters and diacritical marks (cf. Humayoun 2006). The use of the IPA will necessitate that the target users i.e. advanced Pakistani learners are trained to handle it. The skills thus acquired in the use of the IPA will help learners use EFL dictionaries more efficiently since these dictionaries make use of the IPA to represent the sounds of English.
- v. *Part of speech*: The indication of the part of speech to which the entry word belongs has been found to be pretty useful for the users and that is why it is an essential feature of the dictionaries aimed at the learners. In many cases, the lexical items may behave in various ways and as such belong to different categories. For instance, the Urdu lexical item 'پر' 'par' can be a preposition, a conjunction and a noun.
- vi. *Lexical equivalence*: In order to establish equivalence between the lexical items of the two languages which is the core function of a bilingual dictionary, a comparative analysis of the structures of the two languages becomes indispensable. Such an analysis reveals the characteristic features of the lexical items of the two languages. The ultimate purpose is to determine translation equivalent(s) of the Urdu lexical items in English. Some Urdu lexical items will have several equivalents in English, e.g. the Urdu word مادہ 'maadah' is either 'matter' or 'feminine' depending upon context. Similarly, the postposition میں 'mein' requires different translation equivalents in different contexts:

(1)

بستر میں	bister <b>mein</b>	<b>in</b> the bed
ٹکڑوں میں کاٹنا	TukRon <b>mein</b> kaTna	Cut <b>into</b> pieces
راستے میں	raastay <b>mein</b>	<b>on</b> the way
کے بس میں	kay bus <b>mein</b>	<b>within</b> one's power
دونوں لڑکوں میں	donon laRkon <b>mein</b>	<b>between</b> the two boys
دوستوں میں	doston <b>mein</b>	<b>among</b> friends
سات بجنے میں دو منٹ	saat bajnay <b>mein</b> paanch	five <b>to</b> seven (time)
کی عمر میں	ki oomar <b>mein</b>	<b>at</b> the age of
باتوں باتوں میں	baton baaton <b>mein</b>	<b>during</b> the conversation
میں مقبول	<b>mein</b> maqbool	popular <b>with</b>

- vii. *Synonyms*: In a bilingual dictionary, the indication of synonyms in the target language helps the user to find variant possibilities of expression, if only for purely stylistic variation (Zgusta 1971). Though there is no such thing as true synonymy, the purpose

is to give the user an idea of the possible shades of the various senses of the headword in the target language.

- viii. *Pictorial illustrations*: Gous (1994: 61) as cited in Gangla (2001) argues that pictorial illustrations can be employed as a micro structural type to enhance semantic disambiguation. The following figure shows how the translation equivalents of Urdu lexical item پر **'par'** 'wing', 'quill' and 'feather' may be illustrated in an Urdu-English dictionary:

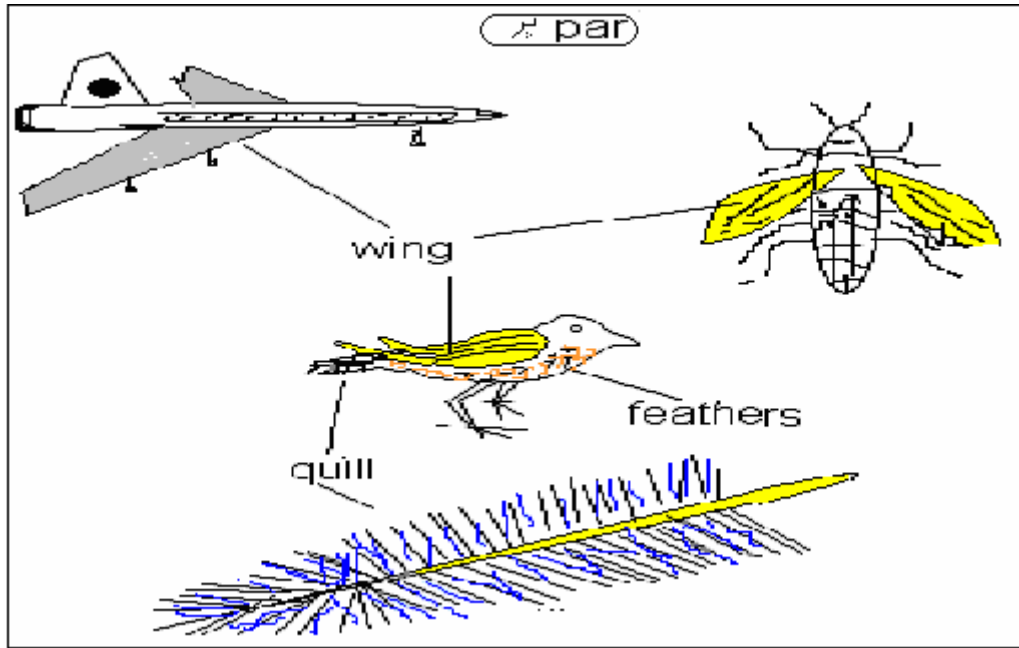


Figure 2 پر *'par'*

From the above illustration it is clear that:

- The use of colours draws the users' attention to the specific details of the translation equivalents.
  - a **'wing'** is an organ of insects and birds as well as a tool for flight
  - only birds have quills and feathers as opposed to wings which insects do have
  - **quills** and **feathers** are much smaller in size as compared with wings.
- ix. *Grammar*\_Grammatical properties of the target language equivalents, in most cases, differ from those of the source language headwords. For example, the Urdu word 'baal' may have plural number whereas its translation equivalent in English i.e. 'hair' is non-count (a feature unknown in the Urdu language) and is used in singular sense.
- x. *Word-combinations* The lexical plight of non-native learners is aggravated in case of fixed expressions, such as collocations and idioms. The problems that fixed expressions pose derive as much from internal features of the target language as from interlingual interference. For example, the translation equivalents of the word 'gehra' will vary depending on the word with which it occurs:



- (2) گہری پانی \* 'gehra pani' deep water; گہرا رنگ \* 'gehra rung' fast colour, but گہری نیند  
\* 'gehri neend' sound sleep.

\*Being an adjective the word *gehra* inflects according to the noun which follows it- a characteristic feature of Urdu.

Below are some of the cases which show how idioms are difficult to translate:

- برا بھلا کہنا *bura bhala kehna* (i.e. to call sb names) literally means  
[say + bad + good];  
- دل باغ باغ ہونا *dil bagh bagh hona* (i.e. 'to be overjoyed') literally means  
[heart + become + garden + garden].

- xi. *Usage labels and notes* The use of labels prevents the users from using the lexical items in an inappropriate context which may cause embarrassment. For example, the lexical item 'goli' means 'tablet' in medicine and 'bullet' in war. The indication of usage restrictions helps the users to use a word or lexical item in an appropriate manner and hence a useful element for encoding purposes.
- xii. *American vs. British variants* American and British Englishes are chief varieties of English. In many instances the differences in the form and the behaviour of the words are significant. (cf. McCreary 2003). At present Urdu-English dictionaries concentrate solely on the British variety. Given the increasing influence of the USA in Pakistan in every field in the wake of advancement in technology, it is necessary that the Urdu-English dictionary does not ignore the US variety of English if it is to be a good guide for the users.
- xiii. *Examples* :The use of verbal illustrations is one of the key features of learner's dictionaries. Toope (1996) claims that in general, the translation of a source language example demonstrates the collocational, stylistic, syntactic, morphological, cultural and semantic characteristics of the target language equivalent to the same extent as the source language example does for the lemma. Let us consider the following instances where various senses of the word 'par' have been disambiguated with the help of the example sentences:

پر **par**

### (3) Noun

(i) wing: *the wing of a bird, aeroplane, or an insect* (ii) feather: *the feathers on the breast/back of the bird* (iii) quill: *the quills on the tail of the bird*

### Conjunction

(i) yet, still, however, but: *People protested yet/still/however/but the government did not change the policy.*

### Preposition:

(i) at: *Who is at the door?* (ii) in: *There are clouds in the sky.* (iii) on: *He put the books on the table.* (iv) over: *A fan hung over her head.* (v) upon: *We depend upon plants for food.*

- xiv. *Encyclopaedic notes* Some lexical items cannot be translated e.g. the names of cultural items, e.g. تعزیه ta'zia. Such items are just taken into the target language in their original form and the corresponding explanation is given in the target language:

تعزیه ta'zia (n. m): A model of the tomb of Hazrat Imam Hussain, the grandson of the Prophet of Islam, carried in a procession by the Muslims of the Shi'a sect to observe his martyrdom

#### 4. Conclusion

It is regrettable that the notion that dictionaries should respond to the linguistic needs, prejudices and abilities of the target users is little known in Pakistani academic milieu. The publishers prefer to pay scant regard to these vital issues. They do so because their product i.e. dictionaries, finds a ready market. Tied to the format of an entry are the issues of the need to teach reference skills and the use of information technology in the form of corpora. The involvement of the EFL teachers can prove a welcome step in view of the fact that they can help raise dictionary awareness and to impart reference skills which are, on the whole, very low (cf. Iqbal: 1987). Also, despite the presence of a number of state-run institutions dedicated to the publication of dictionaries in Pakistan, no corpus of Pakistani Urdu is available so far. The availability of such a corpus will certainly improve the quality of dictionaries. Last but not least, it is high time that we start in-depth reflections and discussions on the theory of lexicography in peculiar Pakistani context in order to explore new vistas of dictionary making.

#### References

- BARTHOLOMEW, Doris, SCHOENHALS, Louise. 1983. *Bilingual Dictionaries for Indigenous Languages*. Summer Institute of Linguistics, 1983.
- GANGLA, Lilian. 2001. *Pictorial illustrations in dictionaries*. M.A. Thesis. Pretoria: University of Pretoria, 2001.
- GOUS, Rufus. 1994. Ostensieve adressering in vertalende wooreboeke. In *Lexikos 4 Afrilex Series*, 1994, vol. 4, pp. 61-85.
- HAAS, Mary. 1962. What belongs in a bilingual dictionary? In HOUSEHOLDER, F. W., SAPORTA, S. (eds.). *Problems in Lexicography*. Bloomington, IN: Indiana University Press, 1962.
- HAENSCH, G., OMENACA, Carlos. 2004. *Los dictionaries del espanol en el siglo XXI*. Salamanca: Universidad de Salamanca, 2004.
- HUMAYOUN, Muhammad. 2006. *Urdu Morphology, Orthography and Lexicon Extraction* [online]. MA Thesis. Chalmers University of Technology, 2006. Available at: <[http://www.lama.univ-savoie.fr/~humayoun/UrduMorph/downloads/HumayounThesis\\_Web.pdf](http://www.lama.univ-savoie.fr/~humayoun/UrduMorph/downloads/HumayounThesis_Web.pdf)>.

IANNUCCI, James. 1957. Meaning Discrimination in Bilingual Dictionaries: A New Lexicographical Technique. In *The Modern Language Journal*, 1957, vol. 41, no. 6, pp. 272-281.

IQBAL, Zafar. 1987. *Aspects of the Learner's Dictionary with special Reference to Advanced Pakistani Learners of English*. Unpublished Dissertation. Birmingham: University of Aston, 1987.

KLAPICOVÁ, Edita. 2005. Composition of the Entry in a Bilingual dictionary. In *Skase Journal of Theoretical Linguistics* [online]. 2005, vol. 2, no. 3 [cit.2007-03-31], pp. 57-74. Available at: <<http://www.skase.sk/Volumes/JTL04/index.htm>>.

LAUFER, Batia. 1995. A Case for Semi-Bilingual Dictionary for Productive Purposes. In *Kernerman Dictionary News* [online]. 1995, no. 3 [cit. 2007-07-01]. Available at: <[www.kdictionaries.com](http://www.kdictionaries.com)>.

LAUFER, Batia, LEVITZKY-AVIAD, Tamar. 2006. Examining the effectiveness of 'Bilingual dictionary Plus' – A Dictionary for Production. In *International Journal of Lexicography*, 2006, vol. 19, no. 2, pp. 135-155.

McCREARY, Don. 2003. British and American English. In *MED Magazine* [online]. 2003 [cit. 2008-05-18]. Available at <<http://www.macmillandictionary.com/MED-magazine/September2003/11-Language-Awareness-British-American-UK.htm>>.

SEKANINOVÁ, Ella. 1993. *Dvojjazyčná lexikografia v teórii a praxi*. Bratislava: Veda, 1993.

SHCHERBA, Lev. 1940. *Opyt obscej teoril lekskografii*. (An attempt towards General theory of Lexicography). In *Izvestija, AN SSR.*, 1940, pp. 89-117.

TOOPE, Michael. 1996. *Examples in the Bilingual Dictionary*. M.A. Thesis. Ottawa: University of Ottawa, 1996.

WORSCH, Wolfgang. 2005. Power to the Learner: An Approach towards Pedagogically – Oriented Bilingual Dictionaries. In *Kernerman Dictionary News* [online]. 2002, no. 10 [cit. 2007-07-01]. Available at: <[www.kdictionaries.com](http://www.kdictionaries.com)>.

ZGUSTA, Ladislav. 1971. *Manual of Lexicography*. Mouton: Academia, 1971.

*Mahmood Ahmad*  
*Ph. D. Scholar*  
*Bahauddin Zakariya University*  
*Multan,*  
*Pakistan*

*mahmoo.szb@yahoo.com*

In *SKASE Journal of Theoretical Linguistics* [online]. 2008, vol. 5, no. 1 [cit. 2008-06-24]. Available on web page <[http://www.skase.sk/Volumes/JTL11/pdf\\_doc/4.pdf](http://www.skase.sk/Volumes/JTL11/pdf_doc/4.pdf)>. ISSN 1339-782X.