CHAPTER 1

PHONOLOGICAL AND MORPHOLOGICAL PROCESSING IN URDU: AN EYE-TRACKING STUDY
Chapter 1

Phonological and Morphological Processing in Urdu: An Eye-tracking Study

Introduction and Review of literature

Eye tracking as a tool for investigation becomes important because it is through the eye that an individual gains information about the text being read. It enables researchers to study the underlying processing strategies when the information being read from the fovea where vision is sharpest and the extent to which partial information from the parafovea facilitates the reader.

Vision is said to be accurate only at the foveal region which is 2° in the center of the retina. The area around the foveal region, i.e. 5° on either side of fixation is the parafoveal region. This region helps in obtaining partial information from the text. The area around the parafovea is said to be the peripheral vision where vision is very poor and extracting information is difficult to impossible (Lima & Inhoff, 1985; Rayner, Well, Pollatsek, & Bertera, 1982).

Eye tracking as a tool, in reading research, has been successful in obtaining accurate, moment-to-moment, online, spatial and temporal record for every word irrespective and independent of a word being used.
in isolation or in context (Just & Carpenter, 1980; McConkie & Zola, 1979; Rayner, 1978; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989)

A large body of research in eye movement suggests lexical processing to be rapid (less than 50ms) (Rayner & Duffy, 1986), automatic and autonomous (Seidenberg, Tanenhaus, Leiman, & Beinkowski, 1982) and post lexical effects to be slow (125-200ms) (Seidenberg et al., 1982; Tanenhaus, Leiman, & Seidenberg, 1979).

However, as the foveal region is limited, an individual has to make eye movements to access a particular part of the text while reading. This movement from one word or a part of the text to another is called a ‘saccade’. The duration of a saccade depends on its length and takes around 20-30ms hence information is hardly extracted (Matin, 1974). After a saccade the eye lands on an area of the text which is termed as a ‘fixation’. The duration of a fixation helps in extracting information from the text being read. The fixation duration can be short, long, skipped, or regressed according to the frequency, word length, linguistic features of a language, ease or difficulty of text, visual acuity and the skill of the reader (Rayner, 1998) and the limitation of visual acuity results in an increase in the frequency of saccades and longer fixation duration.
Research on word recognition has been central to analysis of different codes like grapheme, phoneme, morpheme, and semantics in language processing. Additionally, word recognition research has proposed theories of automatic and attentional processes (Healy & Drewnowski, 1983) and recognition patterns. Research in several languages like English (Rayner, Fischer, & Pollatsek, 1998), French (Vitu, O'Regan, & Mittau, 1990) Finnish (Hyona, 1995), German (Radach & Kempe, 1993) and Hebrew (Deutsch & Rayner, 1999) confirm that most first fixations fall just left/right (Hebrew) of the center of a word (Rayner, 1979) and were independent of word length.

Therefore, the position the eye initially fixates seem to have a strong influence on the ease/ difficulty with which a word is recognized (O'Regan, 1992). The measures to understand ease in word recognition can be the overall time spent inspecting the word (O'Regan & Levy_Schoen, 1987), word naming latencies (O'Regan & Jacobs, 1992), lexical decision latencies and errors (Nazir, O'Regan, & Jacobs, 1991) and perceptual identification errors (Nazir, Heller, & Sussmann, 1992).

Nevertheless, it is understood that fixation duration measures reveal early and late processing stages. Whereby, first fixation duration measure is said to be a window to early processing stage and Dwell time, which is the sum of total fixations and saccades on an area of interest irrespective
of it being the only fixation or the first of multiple fixations, is considered a late processing measure (Inhoff, 1989; Rayner & Pollatsek, 1987). Eye movement studies suggest fixation duration lasts for about 200-250ms with 7-9 letter spaces saccades (Rayner & Bertera, 1979; Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981).

To understand the reading processes in different languages and to postulate models it is important to first understand the basic components of a language and its role in reading and comprehension. Hence, one has to know that the major elements of any language are phonology (sound system), lexicon (vocabulary), semantics (meaning) and morphology (structure) and the process of integrating these linguistic features in print (orthography) for comprehension is said to be ‘Reading’. Also, every language has distinct linguistic features and these features make the task of word recognition and reading easy or difficult. Additionally, researches in language studies confirm that languages with common linguistic features usually show similar results enabling researchers to postulate reading models and that morpho-phonemic, semantic and orthographic feature of a language reasons out the differences in processing during lexical access. Most of the studies, however, are done in English.
Therefore, models of visual word recognition propose that skilled readers bypass phonological information until meaning is acquired (Jared & Seidenberg, 1991) and phonology seem to be accessed via a top-down process initiated by semantic code activation (Patterson & Coltheart, 1987). In lieu with these studies, the connectionistic approach also assumes that printed word recognition is exclusively mediated by orthographic codes while phonological codes being activated automatically in a parallel distributed network (Seidenberg & McClelland, 1989). Furthermore, eye movement research suggests activation of orthographic codes rather than phonological codes to access meaning in English (Daneman & Reingold, 1993) and Hebrew (Frost, Katz, & Bentin, 1987) concluding that phonological structure may not necessarily mediate semantic access as it is integrated automatically during word recognition processes (Van Orden, 1991). Nevertheless, studies in Arabic report word recognition processes to be mediated via phonological codes (Perfetti & Bell, 1991) and activation of phonological codes to be early in case of English (Inhoff & Topolski, 1994; Rayner, 1998; Rayner, Pollatsek, & Binder, 1998). Furthermore, a replication study on low target word predictability suggested strong evidence of orthography being used to activate word meaning (Russel Wood: (Daneman & Reingold, 1993) while high target word predictability text
showed phonological codes to be used to activate word meaning (replication expt: Black Queen: (Daneman, Reingold, & Davidson, 1995). While studies on visual recognition of isolated words in various languages suggest word recognition to be influenced by the languages’ phonological, orthographic, semantic and syntactic features, contemporary linguists suggest morphological structure and productivity to play a more dominant role in shaping the cognitive strategies of a reader as distinct morphological levels are represented in the mental lexicon (Taft & Forster, 1975) with orthographic depth to be concurrent to processing morphological information.

Accordingly, Anderson’s (1992) theory of morphology states “derivational processes to constitute sources for lexical stems and inflectional processes introduce inflectional material into the surface forms of words” and in case of Indo-European languages, the non-neutral derivational processes are less productive and less effective as compared to derived forms with neutral affixes that do not have a stem or affix allomorphy (Alegre & Gordon, 1999; Silva, 2008). Findings in Finnish suggest derived words are accessed as monomorphic words (Hyona, 1995). Additionally, investigation on the interplay of whole-word level and morphologically structured representation during derived word
processing showed different semantic properties to influence performance (Clahsen & Ikemoto, 2012).

With regard to behavioral research on the role of morphology, studies suggest morphological factor to play an independent role and to be distinct from form and meaning. Studies using different paradigms report priming of morphologically related primes in the absence of form and meaning (Bentin & Feldman, 1990; R. Frost, Forster, & Deutsch, 1997) while other factors are blinded (Longtin, Segui, & Halle, 2003; Rastle, Davis, Marslen-Wilson, & Tyler, 2000).

Hence, Rastle and Davis’s (2008) postulated a two-stage model of morphological processing. They state that morphological processing proceeds first in a preliminary stage of morpho-orthographic decomposition. In this stage the surface form of the word is broken down. And in the second stage of morpho-semantic analysis, the meaning of the constituent morpheme is verified (Rastle & Davis, 2008).

Contrastingly, studies in Semitic languages (Bentin & Feldman, 1990) reveal markedly different morphological processing patterns. Studies in these languages report robust morphological priming among morphologically opaque pairs in Hebrew. Additionally, Frost and colleagues (Deutsch, Frost, & Forster, 1998; R Frost & Grainger, 2000; R. Frost, Kugler, Deutsch, & Forster, 2005) demonstrated morphological
priming to arise early and generalized across word-class and degrees of morphological productivity. Therefore, the authors (Forster, Forster, & Deutsch, 1997) concluded that the different patterns were attributable to the non-concatenative feature of Hebrew.

Alternatively, the different emerging patterns in morphological processing across languages, apart from morphological structure, has been attributed to the richness, productivity and consistency in relationship between morphemic units within a language (Plaut & Gonnerman, 2000). Accordingly, they argued that the differences in the morphological productivity between English and Hebrew may be the reason for different patterns. Specifically the inconsistency observed between orthographic (form) and semantic (meaning) units in English as compared to Hebrew.

A reconciliation of the above discussion may be found in ‘Orthographic depth hypotheses’ proposed by (Frost et al., 1987). The authors assume that readers may follow two routes to lexical access based on the phonological transparency of the given writing system. Readers may access an indirect phonological assembly route for languages with shallow orthographies and a direct addressed lexical route for deep orthographies. This assumption is based on the evidence that readers of deep orthographies use a direct route (Bentin, Bargai, & Katz, 1984; R.
Frost, Katz, & Bentin, 1987) and readers of shallow orthography rely on an indirect route prior to lexical access (Feldman & Turvey, 1983; Frost et al., 1987; Lupker, Brown, & Colombo, 1997). Additionally, similar results were found in a comparative study with Hindi, a language with shallow orthography, and Urdu, a language with deep orthography, (Vaid, Rao, & Chen, 2007; Vaid, Rao, Chen, Kar, & Sharma, 2008). Furthermore, studies in compound word processing reports readers to follow a dual route model to encode multimorphemic words (Pollatsek & Hyona, 2006).

Moreover, studies by Rao, (2011) suggested that in case of Urdu, the printed words seemed to activate an internal visual representation of a coarse-grained whole-word which is morphologically structured and permits faster computation of missing information. The authors found different form-priming effect at 48ms and morphological priming effect at long 240ms in Urdu and both form-priming and morphological priming effect in Hindi at 48ms.

Additionally, Stanner’s et al., (1979), presented a long-term morphemic priming to be involved in word recognition (Stanners, Neiser, Hernon, & Hall, 1979) and Lima, (1987), reported that mere orthographic overlap does not produce long term priming effects (Lima, 1987b).
However, the PDP perspective argues that morphemic effects emerge as a result of interactions between orthographic, phonology and semantics (Gonnerman, Seidenberg, & Andersen, 2005), while Rastle, Davis and New (2004) report morphological effects to be independent of semantics and conclude that morphological relationship only facilitates priming irrespective of semantic relatedness (Dominguez, De Vega, & Barber, 2004; Rastle, Davis, & New, 2004) supporting the DRC model.

Nevertheless, masked priming (Forster & Davis, 1984) and eye tracking (Rayner, McConkie, & Zola, 1980) studies report orthographic priming effects during word recognition. Studies in English derived words suggested a principled distinction between orthographic and morphological priming (Morris, Grainger, & Holcomb, 2008). ERP and fMRI data report semantic and orthographic relationship to account for morphological effects (Diependale, Sandra, & Grainger, 2005). And studies in Spanish inflected words suggest a different time course for morphological and semantic information to be integrated. Moreover, segmentation into stem and affix may not be purely based on orthography (Dominguez et al., 2004) as they found distinct ERP profiles for orthography, semantic and morphological priming. The data was reported to be consistent with a large body of research (Dominguez et al., 2004; McQueen & Cutler, 1998) for morphological organization in the lexicon.
However, Taft and Forster, (1975) argue that polymorphemic words are initially decomposed into its constituent morphemes to access lexical files listed under the root morpheme supporting the notion of morphemic levels of analysis in visual word recognition of printed words (Taft & Forster, 1975).

Furthermore, research in the field of parafoveal processing with regard to role of linguistic features have reported both orthographic (Balota, Pollatsek, & Rayner, 1985; Drieghe, Rayner, & Pollatsek, 2005) and phonological (Ashby & Rayner, 2004) preview yield preview benefit. While, no evidence of morphological preview benefit was observed in English (Kambe, 2004; Lima, 1987a) and Finnish (Bertram & Hyona, 2007) robust preview benefit was evident in Hebrew (Deutsch, Frost, Pollatsek, & Rayner, 2005). Recently, semantic preview benefits have been reported in Chinese (Hohenstein & Kliegl, 2014) but not in English (Rayner & Schotter, 2014).
Additionally, (Balota et al., 1985) reported three major findings in parafoveal processing of words.

1. Fixation durations were shorter for orthographic similar preview and base-words.

2. No facilitative effect was observed for semantic similarity between preview and Target words (Rayner, Balota, & Pollatsek, 1986).

3. Interaction effect between predictability and preview information between preview and target.

Therefore, it seems that parafoveal facilitation depends on the languages’ linguistic features and it may be assumed that preview benefit usually occurs when there is some kind of a match between prime-target pairs and the nature of writing system.

The above mentioned studies mainly focus on European languages on one hand and Semitic language on the other. Both the branches of language systems stand in contrast to each other with regard to their linguistic features, directionality, origin and even the results provide a window to differences in cognitive processing. It would be interesting to investigate cognitive processing in a language that has features of both these language systems. Urdu, an Indo-European language, lesser studied though, seems to be sandwiched between European and Semitic
languages. The investigation might reason out the differences in results obtained in the two language systems. Urdu, as a language for investigation seems necessary because of its unique features, with similarities and differences in European as well as Semitic languages. It has similarities and differences to Hindi, an Indo-European language on one hand and Arabic, a Semitic language on the other. Hence, the following discussion presents salient features of Urdu.

**Genesis of Urdu**

The socio-cultural contact between the native population speaking different dialects (Khariboli, Haryanvi, and Brij Bhasha) and the Muslim soldiers belonging to different nationalities (Arab, Iranian, Afghan, Turkey and South Asians) in India during 12th century AD resulted in the emergence of a new language namely Urdu. Urdu lexicon is basically an amalgamation of native Urdu, Persian and Arabic loans. The word “Urdu” itself is a Turkish word meaning camp. As per a rough estimate, the Urdu lexicon has 75% words from Indic sources and 23% of words borrowed from sources like Arabic, Persian and English (Khan & Alward, 2000). These borrowings have enabled Urdu with an unusual capacity of assimilating and accommodating loan words by processes of nativization. Along with whole words, Urdu has borrowed derivational affixes from Arabic and Persian. The following section discusses more on
the impact of borrowings on phonology, semantics and morphological structure of Urdu language.

**Urdu Phonology**

Urdu phonology includes Arabic and Persian sounds along with Indic sound system. The loanwords undergo a phonological change of the base to suit the native/Indic sound system (Bauer, 2003). Various phonological changes in the tri-consonantal root of Arabic loanword are seen in Urdu (Islam, 2011). In other words, the tri-consonantal root undergoes internal vowel change, ex. Noun – qatāl قتل from the tri-consonantal root q t l ق ت ل

undergoes internal vowel change in the root to form another noun – qatıl قتل (Islam, 2011). The Persian alphabets included in Urdu are g گ, ch چ, p پ, žh ژ. And alphabets to suit the Indic phonemes included were ٹ, ڑ, ٹئ, Te, de, dal, and zal (Khan, Anwar, Bajwa, & Wang, 2012).
The vowels are represented by alif ‘ا’, for ‘a’, ‘e’ and ‘i’ sounds, bariyе’ے’ and choti ye ‘ي’ for ‘e’ and ‘i’ and wa’o ‘و’ for ‘u’ and ‘o’ sounds. There are diacritical marks called the zaber (Arabic Fatha) written above the letter to denote /a/ phoneme, zer (Arabic Kasra) written below the letter to denote /e/ or /i/ phoneme and pesh (Arabic damma) marked above the letter to denote /u/ and /o/ phoneme. These markers are optional and are context sensitive short vowels. The long vowels are alif madd ‘ا’ alif ‘ا’, choti ye ‘ي’ and wa’o’s’. The diacritics act as a depending vowel when used along with consonants and as independent vowels for stress or to prolong the phoneme. Native speakers are tuned for correct pronunciation by virtue of their knowledge by looking at the context.

**Urdu Orthography**

Urdu orthography is a modified form of Persio-Arabic script. It is written from the right to left. Urdu has 35 unaspirated consonants and 15 aspirated consonants. The aspirations are formed by adding the regular consonants to a special letter called Heh Do-Chashmee (ھ). The Urdu characters change their shape according to the position they occur in the word. The alphabet acquires different forms in isolated,
initial, medial and final positions of a word. The Urdu alphabets are
divided into joiners and non-joiners. There are 9 non-joiners in Urdu alphabet set, they are —ا، ر، ڑ، ز، ذ، ڈ، د، و، ذ، ۔ and do not change their shapes in words. These non-joiners are responsible for space insertion within the words resulting in ambiguous word boundary ex ہل badal. Readers tacitly use the rules and know the differences in space within the words and word boundaries (Rahman, 2012). The joiners, on the other hand, change their shapes as per their occurrence. There are a set of rules that are followed when the letters acquire shapes as per their occurrence in the word.

A joining letter takes-

- Initial form with a joiner letter ﮑ
- Initial form with a non-joiner letter ﮝ
- Final form when it comes after a joiner ﺟ
- Isolated form when it comes after a non-joiner ﺔ
- Medial form when it is already in the final form and is followed by a joiner ﮣ
• Medial form when it is already in the final form and is followed by a non-joiner.

A non-joiner alphabet follows the following set of rules-

• Isolated when a joiner or non-joiner follows.

• Isolated form when it is followed by a non-joiner.

• Final form when it is followed by a joiner.

The script is written in the ‘Nastalique’ font from Naskh and Talīq; also anglicised as (Nastaleeq) is one of the main calligraphic hands used in writing the Persian script, and traditionally the predominant style in Persian calligraphy. It was developed in Iran in the 14th and 15th centuries. It is sometimes used to write Arabic-language text where it is known as Taliq or Persian and is mainly used for titles and headings, but its use has always been more popular in the Persian, Turkic and Urdu sphere of influence. Nastalīq has extensively been (and still is) practised in Iran, Pakistan, India, Afghanistan and other countries for written poetry and as a form of art similar to Arabic. It moves upwards in a diagonal style. The cursive nature renders the text a compact visually complex appearance.
**Urdu Morphology**

The root word/morpheme is that morpheme that does not undergo any change in any morphological derivations (Sabzwari, 2002). While the Urdu verb and noun show agreement for number, gender and case, the verb shows inflections for infinitive, past and imperative forms and the noun shows derivational changes into adjectives and other nouns with or without semantic change (Hussain, 2004).

The root word has various morphological variants. When the structure of the root has similarity with its actual and undergoes grammatical changes only, then it is termed as inflectional stemming “e.g. actual word – barasaat (noun) root word – baras (noun) both having same category” and when the root and word are varied to form new words it is termed as derivational stemming e.g. actual word- khamoshi (noun); root word – khamosh (adjective) here noun is derived in adjective format” (Gupta & Joshi, 2015). Thus, Urdu is very rich in morphological productivity.
Morphological processes in word formation

Words are derived in the following main categories:

- **Affixation**: A major process in word formation is through affixation. Affixes are bound morphemes and are usually added to another morpheme to form a word. Urdu has its own native affixes and also borrows Persian and Arabic affixes in its lexicon. There are various patterns from each source and many variations within each pattern. Derivations are usually rule-bound but there are also deviations when affixation is the criteria.
The derivations thus formed from all the major categories of words—nouns, verbs, adjectives, adverbs and pronouns by affixation. Adverbs have inflected and uninflected formations, inflected forms require grammatical information while uninflected forms do not (Gupta & Joshi, 2015).

Native Urdu Affixes- Most of the derivations are based on native Urdu affixes, albeit only few - a, i and negative markers -ə, -ən. These affixes form the basis of Urdu morphology. Derivations with native Urdu affixes are systematic and rule-bound. When the base noun is native the derivations expresses an inherent gender and are mainly adjective that can also be used as nouns, while for loan words only the native affix –i is attached, expresses no gender and forms adjectives only.

Prefixes- the native prefixes like la-, bila-, or –bina and Persian prefixes in Urdu - be-, ba-, bad result in semantic change but have intact lexical category.
Arabic Infixes – Infixes are the insertion of morphemes within a word. This result in the non-concatenation in words thus formed. Urdu borrows this feature salient in Arabic in word formation and is restricted to Arabic loanwords in Urdu and morphological possibility seems to be exploited to a great extent. Words of Indic origin do not show infixation.

Suffixes- The suffixes in Urdu change the lexical category during derivations irrespective of the suffix being a loan Persio-Arabic suffix or native Urdu suffix.

Zero-modification- when the two morphemes are identical but differ in meaning. These words do not permit inflectional change ex. Akhbaar اخبار, mard مرد.

Suppletion: the original word form is entirely replaced by a new form. It usually occurs in tenses and gender marking ex mard مرد – aurat اورت

Word formations by compounding

Compound word formations may be tatsam or tatbhav, meaning compound words are derived from Sanskrit (Rahman, 2012). Urdu, like English, is a head-final language. A large source of compound words has
been borrowed from Persio-Arabic sources which follow the same rules as native compounds of Urdu. Compounding is a very rich phenomenon. It is inherited in Urdu via Persian (1). The compounds usually occur in three formations – the XY, X-o-Y, and X-e-Y. Compounds are formed by the combination of two free morphemes.

There are four types of XY combinations based on certain set of rules.

a. Dvanda- are pure compounds and combine two ways. When two morphemes are identical in meaning and differ in meaning. They form noun-noun or verb-verb combinations. 

\[ \text{kam-kaj, dekh-bhal or maa-baap, khelna kudna} \]

b. Tatpurusa- the compound word formations where the type of Y is related to X corresponding to one of the grammatical cases of X. ex

\[ \text{ghur-dourh} \]

c. Karnadharya – the compound word formation denotes a relation between the first and second morpheme which could be attributive, appositional or adverbial “e.g. bd kanna \( बड़ा \) one with big ears;

\[ \text{bdh-bolna \( बड़ा \) to exaggerate (Durrani, 2007)} \].
d. Divigu- the compound formation has the first morpheme as a numeral denoting quantity of the second morpheme. “e.g. adh-
muah।

The Tatpurusa, Karmadharya and Divigu are also called the determinative compounds.

X-o-Y and X-e-Y compound formations

X-o-Y formations are coordinating compounds involved in nominal constructions. It contains a linking –o- morpheme meaning ‘and’. The X-e-Y compounds contain an enl ic tic short vowel /e/ known as zer-izafat or hamza-e-izafat. Izafat means increase or addition. It forms noun-e-noun or noun-e-adjective compounds signifying possessor relationship (Schmidt, 1999). Zer-e-izafat is written as –a subscript zer or as hamza over bari Yeh when it follows a word ending in long vowels, as hamza over choti Heh and to denote silent feature when it follows a word ending with badi yeh.

Reduplication

Compound words are formed by the process of repetition of either a part or whole root (Pyne, 2006). It is used to put emphasis or when the Y component is a nonsense word.
Linguistic borrowings and loanword morphology

Urdu has always been considered to be similar to Arabic because of its close association with religion, the Arabic Quran and Muslims (followers of Islam). The lexicon of Arabic entered Urdu via Persian and hence undergoes changes in Persian before entering Urdu. It is interesting to note that the loanwords undergo phonological, morphological and semantic changes with orthography intact when they enter the recipient language.

The loan words entering Urdu from Arabic or Persian sources are limited and not all derivations are seen in Urdu. Also, the root/base in Urdu may be modified to form new words unknown to Arabic lexicon. E.g. the root ظ لم z l m forms the word ظالم, zalim found both in Arabic and Urdu; ظلمان zulmana found in Persian and Urdu not known to Arabic; ظلمī zulmi found only in Urdu not known to Persian or Arabic.

The borrowings never had any criteria for accommodation and depended upon the ease of inclusion, i.e. Arabic singular words were easy to borrow and hence borrowed and Arabic plurals, especially duality that is unknown to Urdu, could also be accommodated without changes ex. وَالدِينُ waledain (parents) duality in Arabic has been included in Urdu in
its original form to mean parents. Furthermore, all the loanwords under one family either entered Urdu with minor changes, an individual word of a family was borrowed or the whole family of loan words entered Urdu along with semantics on one hand and with phonological variations, which may be attributed to pragmatic influences, and subsequent semantic change on the other (Hopper & Traugott, 1993, 2003; Levinson, 2000).

Also, Urdu seems to have borrowed three major features from Arabic-

1. Feminine gender marker –a(h),


3. Various patterns of broken plurals.

Morphological changes of loanwords occur both at inflectional and derivational levels. At inflectional level, changes are based on native patters while at derivational level adaptability is restricted irrespective of Persian, Arabic or Native Urdu patterns.

*Morphology of Urdu noun*

The three sources of vocabulary –Native, Arabic and Persian makes the analysis of morphology of nous arbitrary. This is so because loans from each source behave differently. Though the loanwords follow the source language morphology, they are nativized for gender, number and case.
Native Urdu has two genders – masculine and feminine while there is no
gender distinction in Persian, two numbers as against 3 in Arabic and
there are three cases – oblique, nominative and vocative.

Moizuddin (1989) defines certain criteria for morphological changes –

a. Declinable and indeclinable nouns may or may not undergo
   structural changes.

b. Changes may be related to the case markers.

c. Structural changes are correlated to gender and number agreement.

Inflections in Urdu are based on suffixation which is fussional and
contains single vowel or syllable to mark gender, number or case
(Moizuddin, 1989).

Where Persian loans are marked for gender and plural marking as per
native Urdu morphological patterns, Arabic loans are nativised under
three criteria-

1. Loans adopt native Urdu gender and number markers.

2. Both loan singular and plural forms are borrowed and are intact.

3. Affixation based on vowel changes.

The basic masculine gender is inherent in Urdu and the feminine forms
are derived mostly from the masculine forms and nativized.
The foregoing literature provides evidence that loanwords from any source usually adopt the native Urdu morphological patterns. Where there are similarities, there are deviations in the morphological patterns. Urdu morphology is assumed to be an amalgamation of native Urdu, Arabic and Persian morphological structures. The adaptations may be partially or fully integrated in the native form because most Arabic loans enter Urdu via Persian. Hence loans first undergo morphological changes in Persian and then further undergo morphological changes in Urdu (Riaz, 2007).

Research in European language and Semitic language has shown contrasting results. If orthography and phonology facilitated lexical access in the former, morphology played a more crucial role in the latter. With the above literature we planned a series of experiments with Urdu as a language of investigation that seems to be related to both the language systems. Since there has been little research in Urdu the investigation could highlight features that would aid further research. We designed five experiments to understand the underlying cognitive processes and the influence of phonology, orthography and morphology during lexical access. The first experiment explores the effect of phonology and morphology in a visual world paradigm and recorded eye movements. We planned this experiment in the light of homophony and the complex morphology. Additionally, we wanted to see if skill in reading had an
effect in processing. The experiment touches upon effect of phonology because of many letters representing a single sound and mainly on role of morphology during lexical access. The next set of experiments focused on morphological processing in the parafovea during word recognition in the fovea. The last but not the least we designed an experiment to see differential processing across word type because the process of word formation in Urdu is similar to Arabic. The results have indeed opened new avenues for further research in Urdu as a language that would link results in both the language systems. Each experiment is well explained in the second chapter.

**Summary of Chapter**

The chapter highlights the use of eye tracking as a tool for understanding cognitive processes in reading research and linguistic feature of Urdu, the language for investigation. Research on eye tracking has enabled researchers to understand the limitations of visual acuity in the foveal region and that readers obtain partial information like word length, frequency etc from the parafovea while it is difficult or impossible to gain information from the peripheral region.

Findings in word recognition processes in several languages report linguistic features of a language to influence reading processes. Research using different paradigms like priming, ERP, fMRI, e-prime and eye
tracking has resulted in postulating reading models based on the specific linguistic features and the type of orthography of a language. Therefore, suggest that readers of shallow orthographies like Serbo-Croatian, Hindi and the like to follow a phonological assembled route and readers of deep orthographies like English, Hebrew and Urdu to follow a direct lexical addressed route during reading.

The second part of the chapter explains the linguistic structure of Urdu, an Indo-European language, socio-cultural context and the nativization processes of loanwords from Arabic.

The loan words from sources like Arabic and Persian undergo nativization and follow the native Urdu morphological patterns. The masculine gender and duality features seen in Arabic are well accommodated in Urdu without any change. Though the tri-consonantal root of Arabic is evidenced in Urdu, the concept of root and word pattern morphology seen in Semitic languages is alien to Urdu.

Moreover, derivations from the tri-consonantal Arabic root undergo phonological change and modification of the base as it enters Urdu. The words in Arabic are pronounced with an emphatic sound while the same in Urdu it undergoes phonological changes with native sounds. Hence, it is imperative to conclude that irrespective of the source of loanwords almost all loanwords entering Urdu follow the native Urdu morphological
patterns. Therefore, derivations are no more than loans as the process of derivational patterns of tri-consonantal root and word pattern in Arabic is unknown to Urdu. Furthermore, the loans are restricted to a few patterns which contain only few derived word. Therefore, Urdu, as a language, stands as an explicit example of integration of loanwords from complete different morphologies into its native morphological patterns. Studying the impact of such amalgamation of morphological structures on cognition would not only be interesting but also pave way to new fields in research.

The upcoming chapter deals with the series of experiments conducted with different set of stimuli, different paradigms and investigating tools to understand lexical processing in Urdu and compare the results with European languages on one hand and Semitic languages on the other. The five experiments are well planned with utmost care for precise results and the last chapter discusses the outcome of the five experiments and future directions for further research.