

## Appendix 2 : Glossing system

In this thesis, where I have given an example drawn from Schmidt (1999), the translation is Schmidt's; I have supplied a gloss.

The gloss given is not a morpheme-by-morpheme explanation as might be found in an in-depth study of Urdu morphology and syntax. Instead, the gloss is word-by-word. This is appropriate in a work on part-of-speech tagging, since the word is the level of language which is being dealt with. The glosses consist of one or more items, either meanings or grammatical features. Where a single word must be given more than one item in its gloss, these items are joined together with hyphens. For example, the word *hai* is frequently glossed as “be-PRES-3-SING”. No attempt is made to separate out the morphemes in the original Urdu, nor to indicate which part of the word corresponds to which part of the gloss; for the purposes of this thesis, it is irrelevant.

Meanings (i.e. lexical content expressed by means of the closest possible English equivalent) are given in lowercase letters in the glosses. Grammatical features such as agreement, tense or aspect are indicated using abbreviations in uppercase letters. Some words with minimal semantic content such as the postposition *nē* or the future auxiliary are given a grammatical feature as their “meaning”, since even a roughly equivalent English word is hard to find. Some grammatical features are described with two abbreviations; when this is the case the two are separated by a full stop, for example “PERF.PART” as a gloss indicating the perfective participle form of a verb. The full list of abbreviations used is tabulated below.

Not every grammatical feature has been marked in every single example, for reasons of space. For instance, it is unnecessary, in an example focussing on the

aspectual properties of the verb, to list the case, gender and number properties of every noun in the sentence: they make no difference to the point under discussion.

Nouns and verbs are particularly likely to require glosses containing several grammatical features. The following is the order in which the grammatical features are listed in such a case; of course not all these categories will be relevant in any one gloss: verb form (i.e. tense, aspect etc.)– person – gender – number – case.

The table that follows lists and expands all the abbreviations used in glosses in this thesis.

Abbreviation	Signification
1	First person
2	Second person
3	Third person
ERG	ergative (i.e. the “meaning” of the postposition <i>nē</i> ) <sup>1</sup>
FEM	Feminine
IMPERF	Imperefect
MASC	Masculine
NOM	Nominative
OBL	Oblique case

---

<sup>1</sup> I am not particularly happy with using ERG to indicate the meaning of *nē* since this comes perilously close to suggesting that Urdu does display ergativity, a claim I am not willing to make (see 1.1.5.4). However, I have not been able to find a translation for *nē*. Schmidt (1999) does not suggest one; Haq (2001) states explicitly that the word is not translated into English. This being the case, ERG seems the best way to state the signification of *nē*.

PART	Participle
PAST	Past tense
PERF	Perfective
PLRL	Plural
PRES	Present tense
SING	Singular

Note that a verb in the root form is given with no grammatical features, just a meaning, in its gloss.