

---

## Word order and question type: A corpus analysis of Bollywood scripts

---

Miriam Butt

*Universität Konstanz*

[miriam.butt@uni-konstanz.de](mailto:miriam.butt@uni-konstanz.de)

Farhat Jabeen

*Universität Konstanz*

[farhat.jabeen@uni-konstanz.de](mailto:farhat.jabeen@uni-konstanz.de)

Tag  
Datum  
Zeit  
Raum

Urdu/Hindi has a number of interesting properties with respect to questions. It is a language with default SOV word order, but where major constituents can scramble. The major determinant of word order seem to be information-structural concerns. With respect to constituent questions, the most natural position for wh-phrases is the immediately preverbal position, a position that is also associated with focus. However, wh-phrases can in principle appear anywhere else in the clause and have the same scrambling possibilities as NPs. The constraints and licensing conditions on the scrambling possibilities of wh-phrases in Urdu/Hindi remain under investigated, however there is a marked dispreference for the clause initial position. The reason for this appears to be due to the structure of polar questions.

Polar questions in Urdu/Hindi have a declarative syntax, but also allow for the presence of a question marker *kya* ‘what’. The clause initial position for this polar *kya* has been characterized as being its default position. However, just like wh-phrases in constituent questions, the polar *kya* can in principle appear anywhere in the clause, with a marked dispreference for the immediately preverbal position — the default position of the wh-phrase.

While some progress has recently been made on understanding the distribution and meaning of polar *kya*, the constraints and licensing conditions associated with where it appears in a clause are still understudied.

In this talk, we present work from a corpus study of 10 Bollywood film scripts. We investigate the position of wh-phrases and polar *kya* and how this correlates to interpretational possibilities of questions. For example, clause final polar *kya* is more often associated with the expression of rhetorical questions, while immediately postverbal wh-phrases seem to be used to express secondary focus. With our results, we seek to inform the current theoretical linguistic discussion with data that very closely approximates spoken language and in turn draw on the existing insights coming from theoretical linguistics in order to form the hypotheses that form the basis of our corpus study.