The Syntax and Interpretation of Urdu/Hindi Polar *kya*

Miriam Butt
University of Konstanz

María Biezma
University of Massachusetts, Amherst

Proceedings of the LFG'22 Conference

Miriam Butt, Jamie Y. Findlay and Ida Toivonen (Editors)

2022

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper focuses on the syntactic distribution and corresponding interpretative possibilities of Urdu/Hindi polar *kya* 'what'. Building on previous work on the prosodic and pragmatic properties of polar *kya*, this paper takes on the complex interaction between interpretational scope, prosodic prominence and the syntactic distribution of *kya*. The analysis posits one underlying wh-item *kya* 'what' that plays out as either a particle used in polar questions or as a constituent wh-word. The targets of application of the *kya* are determined via an interaction of syntactic position with prosodic prominence, which is modeled via f-precedence and scope marking at f-structure.

1 Introduction

The work presented here is part of a larger effort to understand non-canonical questions in Urdu/Hindi. We define non-canonical questions as being all those which are not straightforwardly information seeking (like rhetorical questions, for instance). This paper focuses particularly on the role and syntactic analysis of a wh-word that has been dubbed *polar kya* by Bhatt and Dayal (2020).

(1) **kya** ye mer-a pɑhl-a pɑhl-a pyar hɛ?
what this.Nom I.Gen-M.Sg first-M.Sg first-M.Sg love.M.Sg.Nom be.Pres.3.Sg
'Could this be my very first love?'
(From song Suraj Hua Maddham in *Kabhi Khushi Kabhie Gam*)

As (1) illustrates, this *kya* 'what' can appear in polar questions. However, its presence is not obligatory (see (12)). In (1), the *kya* is in clause-initial position, where it is found most frequently, together with the clause-final position. It can also be found distributed throughout the clause. In seminal work Bhatt and Dayal (2020) suggest that *kya* is a Polar Question Particle (PQP) and see it as residing in the ForceP of a clause, with its surface position signaling the scope over what is possible to be questioned.

In contrast, in previous joint work investigating the prosodic, syntactic and interpretational properties of *kya*, we argued that polar *kya* serves as a marker of uncertainty: the speaker uses *kya* to signal uncertainty about the propositional content of the polar interrogative (Biezma et al. 2022). Because of its meaning contribution, *kya* is also very suitable in utterances in which the speaker wants to emphasize disbelief, or surprise, as in (1), but also in rhetorical questions when the speaker wants to add a sarcastic note, as in (2).

[†]Very many thanks go to Rajesh Bhatt and Veneeta Dayal for the original inspiration and some further discussions, Saira Bano and Ghulam Raza for help with the data, suggestions, general pointers and interesting discussions. Bettina Braun, Regine Eckardt, Gillian Ramchand and Maribel Romero have each played important parts in helping us to come to grips with the phenomena and two anonymous reviewers provided invaluable comments. Mary Dalrymple helped with f-precedence (implementation vs. theory). Finally, we thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for funding within projects BI 1836/1-1 and BU 1806/9-2 ("Information Structure and Questions in Urdu/Hindi") of the FOR2111 *Questions at the Interfaces*.

(2) Context: A is telling B how to behave in a situation. B says (with sarcasm):

B: tʊm mer-i ɑmmɑ̃ ho **kya**?
you.Nom I.Gen-F.Sg mother.F.Sg.Nom be.Pres.2.Sg what
'Are you my mother?'

Because the distribution of *kya* in the clause is variable and because the different positions are correlated with different interpretational possibilities, we analyze *kya* as a focus sensitive operator whose precise interpretation differs according to its position in a clause (cf. also Syed and Dash 2017 for related languages). In terms of prosodic realization, we have found that the polar *kya* is prosodically flat or falling (Butt et al. 2017), unlike its wh-constituent counterpart, which always carries a high tone so that prosodic cues can serve to disambiguate a single underlying entry for *kya* 'what' (Butt et al. 2020).

Section 2 first provides background on Urdu/Hindi question formation, moving on to polar *kya* in section 3. A syntactic LFG analysis is provided in section 4, in which f-precedence and scope marking are invoked to model the complex distribution of polar *kya*, which is determined not only by syntactic position, but also by prosodic prominence. Section 5 concludes.

2 Constituent Question Formation in Urdu/Hindi

The material in this section draws on Butt et al. (2017) and Butt et al. (2020) and provides background on constituent question formation in Urdu/Hindi. The basic word order is SOV and major constituents can scramble, yielding information structural effects (Gambhir 1981, Butt and King 1996, 1997, Kidwai 2000).

Urdu/Hindi has traditionally been characterized as a wh-in-situ language (see Bayer and Cheng 2015 for some discussion). An example is in (3), where the *kɪs=ko* 'whom' appears in the same position as the object *ram=ko* 'Ram'.

(3) a. sita=ne dʰyan=se **ram=ko** dekʰ-a tʰ-a
Sita.F=Erg attention.M=Inst Ram.M=Acc see-Perf.M.Sg be.Past-M.Sg
'Sita had looked at Ram carefully' (based on Mahajan 1997)

b. sita=ne dʰyan=se **kɪs=ko** dekʰ-a tʰ-a?
Sita.F=Erg attention.M=Inst who.Obl=Acc see-Perf.M.Sg be.Past-M.Sg
'Who had Sita looked at carefully?' (based on Mahajan 1997)

Another example is (4), where the wh-constituent word *kya* 'what' appears in the same position as the questioned object *gõglu gʰoʃt* 'turnip-meat-dish'.

(4) a. norina=ne maze=se **gõglu gʰoʃt** kʰa-ya
Norina.F=Erg enjoyment.M.Obl=Inst turnip meat.M.Nom eat-Perf.M.Sg
'Norina ate a turnip-meat dish with great enjoyment.'

b. norina=ne maze=se **kya** kʰa-ya?
Norina.F=Erg enjoyment.M.Obl=Inst what eat-Perf.M.Sg
'What did Norina eat with great enjoyment?'

However, the default position for wh-constituents is actually not the in-situ position, it is the immediately preverbal position, as shown in (5) where the *kɪs=ne* 'who' appears immediately preverbally even though the default position for the subject 'Sita' is clause initial in the corresponding declarative.

(5) a. **sita=ne** ram=ko dek^{h}-a t^{h}-a
Sita.F=Erg Ram.M=Acc see-Perf.M.Sg be.Past-M.Sg
'Sita had seen Ram.'

b. ram=ko **kɪs=ne** dek^{h}-a t^{h}-a?
Ram.M=Acc who.Obl=Erg see-Perf.M.Sg be.Past-M.Sg
'Who had seen Ram?'

This immediately preverbal position is also the default position for focus. This has been established in the descriptive and formal literature (Gambhir 1981, Butt and King 1996, 1997, Kidwai 2000) as well as by experimental data: Féry et al. (2016) conducted a study in which they asked informants questions as in (6).

(6) a. Questioning the Subject: *In front of the well, who is pushing the car?*
b. Questioning the Object: *In front of the well, what is the man pushing?*

Féry et al. (2016) tallied the word orders used by Hindi respondents and while results varied, by far the most frequent word orders were OSV and SOV, as shown in (7). These results are in line with the immediately preverbal position being the default position for focus: when the subject was questioned, the overwhelming number of responses placed the subject immediately before the verb; when the object was questioned, the object was placed in immediately preverbal position.

(7)

	SOV	OSV
Subject Questioned (n=28)	6	22
Object Questioned (n=26)	26	—

These results are confirmed by our own corpus study of 12 Bollywood scripts,[1] in which we extracted all possible wh-words and then investigated the distribution of wh-words in the clause. Table 1 shows the word order distribution, where we did not include *kya* 'what' in the counts because of its multifunctional nature. The distribution of wh-words again identifies the immediately preverbal position as the one found most frequently, confirming the experimental data and the descriptive as well as theoretical analyses that posit the the immediately preverbal position as the default position for focused items and therefore as the preferred position for wh-words (which are inherently focused; Rooth 1985, 2016).

[1]The scripts are of the following movies and were originally compiled by Farhat Jabeen: Ankhon Dekhi (2014), Dedh Ishqiya (2014), Dum Laga Ke Haisha (2015), Jab We Met (2007), Lootera (2013), Masaan (2015), NH10 (2015), Queen (2014), Socha Na Tha (2005), Talvar (2015), Titli (2014), Udaan (2010).

Position	**Core Arguments** (without *kya*)	**Adjuncts** ('where', 'when')	**Total**
Single Word	28	14	42
Initial	9	10	19
Medial	2	12	14
Preverbal	**118**	**209**	**327**
In Verbal Complex	0	5	5
Postverbal/Final	6	7	13
Embedded	12	17	29
No Verb	14	5	19
Total	189	279	468

Table 1: Distribution of wh-words in Bollywood scripts

The data in Table 1 also illustrates that in addition to the default position immediately before the verb, wh-words can in principle appear anywhere in the clause, including within the verbal complex (Butt et al. 2016). Manetta (2012) establishes that wh-constituents have exactly the same kind of scrambling possibilities as standard NPs. And just like with standard NPs, the varying positions in the clause are associated with different information structural effects. This is shown in some detail for example with respect to immediately postverbal wh-constituents within the verbal complex by Butt et al. (2016), who analyze these as expressions of secondary focus. Much work remains to be done on the scrambling possibilities of Urdu/Hindi wh-constituents, with some other recent work having been done by Manetta and Gribanova (2016) and Dayal (2017).

3 Polar *kya*

The material in this section again draws on Butt et al. (2017) and Butt et al. (2020) to provide the necessary background on the properties of polar *kya*.

3.1 Question Formation via Intonation

As illustrated in (8), polar questions in Urdu/Hindi (as in Indo-Aryan languages in general) are string-identical to their corresponding declaratives.

(8) a. ʃahina=ne norina=ko mara$_{\text{L-L\%}}$
Shahina.F=Erg Norina.F=Acc hit-Perf.M.Sg
'Shahina hit Norina.' (Declarative)

b. ʃahina=ne norina=ko mara$_{\text{L/H-H\%}}$
Shahina.F=Erg Norina.F=Acc hit-Perf.M.Sg
'Did Shahina hit Norina?' (Polar Question)

Question vs. declarative status is signaled via intonation. A declarative has an intonational phrase boundary that ends on a low tone: L-L%. In contrast, a polar question is signaled by a high intonational boundary tone: L/H-H%. The F_0 contours for the examples in (8) are provided in Figure 1.

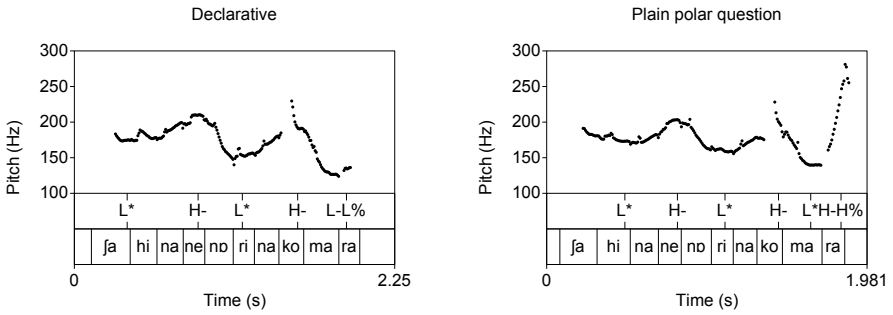


Figure 1: F_0 contour of a string identical declarative and polar question.

As already discussed above, polar questions can optionally be expressed with *kya* 'what'. As shown in (9), the boundary tone does not change.

(9) (**kya**) ʃahina=ne norina=ko mara$_{\text{L/H-H\%}}$
what Shahina.F=Erg Norina.F=Acc hit-Perf.M.Sg
'Did Shahina hit Norina?' (Lit. 'What did Shahina hit Norina?)

Our investigations did, however, bring to light a prosodic difference between polar *kya* and its wh-constituent counterpart. While the wh-constituent *kya* always carries a high tone, just like all the other wh-words in the language, the polar *kya* shows a flat or falling contour. Examples as in (10) and (11) are thus in principle ambiguous. However, access to prosody allows for disambiguation: if the *kya* carries a high tone, it is a constituent question; else it is a polar question.

(10) mɛ̃ **kya** bol-ũ?
I.Nom what speak-1.Sg
Constituent Question: 'What should I say?'
Polar Question: 'Should I say (something)?' Script, *Ankhon Dekhi*

(11) **kya** tɑklif ho rɑh-i hɛ [...]?
what bother.Nom be Prog-F.Sg be.Pres.3.Sg
Constituent Question: 'What's bothering (you)?' Script, *Ankhon Dekhi*
Polar Question: 'Could something be bothering (you)?'

3.2 Clause Internal Syntactic Distribution

Grammars and previous literature report polar *kya* as appearing only clause initially in Urdu/Hindi (Platts 1884, Glassman 1977, Masica 1991, Montaut 2004). However, as already illustrated by (2) and (10) above, the polar *kya* can appear anywhere in the clause, a point established systematically by Bhatt and Dayal (2020). This

is illustrated in (12), which also shows that the polar *kya* is strongly dispreferred (marked with the %) in the immediately preverbal position.

(12) (kya) ɑnu=ne (kya) uma=ko (kya) kɪtab (%kya) d-i
what Anu.F=Erg what Uma.F=Dat what book.F.Sg.Nom what give-Perf.F.Sg
(kya)?
what
'Did Anu give a/the book to Uma?

We see the dispreference for the immediately preverbal position as falling out from the establishment of this position as the default focus position of the clause. This position is eminently compatible with the inherently focused wh-constituent words (and their high tone), but is marked with the polar *kya* and its level/falling tone.

3.3 Interpretation of Polar *kya*

This section is based on Biezma et al. (2022), which offers an analysis of the semantics and pragmatics of polar *kya*. We provide the key ingredients of the analysis below. The reader is referred to Biezma et al. (2022) for further details and a full comparison with other proposals in the literature.

Although the presence of *kya* appears optional most of the time, there are contexts in which *kya* cannot be used. For example, polar *kya* cannot appear in contexts in which the speaker is seeking to confirm something they already believe to be true ("confirmation" questions), see (13) (small caps indicate prosodic prominence).

(13) Context: I had to work and couldn't attend Amra's birthday party but you were there. Ravi is a well known scrooge and while I don't have evidence of this...

Me: I'm sure Ravi didn't bring anything for Amra this time either. Come on, confirm what happened yesterday, . . .

(i) ✓ rɑvi=ne amra=ko kʰILONA di-ya?
Ravi.M=Erg Amra.F=Dat toy.M.Sg.Nom give-Perf.M.Sg
'Did Ravi give a toy to Amra?'

(ii) #/? rɑvi=ne amra=ko **kya** kʰILONA di-ya?
Ravi.M=Erg Amra.F=Dat what toy.M.Sg.Nom give-Perf.M.Sg
'Did Ravi give a toy to Amra?'

This contrasts with contexts in which the speaker is agnostic about the answer, as in (14). The only difference between (13) and (14) is that in (13) the speaker is certain of the answer (that Ravi did not give anything to Amra), while in (14) the speaker is agnostic about the answer. Polar *kya* is only licensed in the latter.

(14) Context: It is equally possible that Ravi gave a toy to Amra for her birthday or that he didn't bring anything. It all depends on whether he had time to pass by a toy store on his way to the party.

Me: So, what happened in the end?

(i) ✓ rɑvi=ne amra=ko kʰILONA di-ya?
Ravi.M=Erg Amra.F=Dat toy.M.Sg.Nom give-Perf.M.Sg
'Did Ravi give a toy to Amra?'

(ii) ✓ rɑvi=ne amra=ko **kya** K^{h}ILONA di-ya?
Ravi.M=Erg Amra.F=Dat what toy.M.Sg.Nom give-Perf.M.Sg
'Did Ravi give a toy to Amra?'

In embedded contexts polar *kya* can appear easily with rogatives like 'wonder', but is in general not licensed with responsives like 'know'. Bhatt and Dayal (2020) provide an analysis of the distribution of *kya* based on syntactic selection explaining that clauses with *kya* could not embed under responsives. However, Bhatt and Dayal (2020) acknowledge that their proposal leaves unexplained cases in which *kya* was actually possible embedded under responsives. Crucially, those contexts are contexts with negation, future tense or an imperative marker. All these are cases in which the attitude holder does not know/is uncertain; see (15).

(15) a. *jan-na ho-ga* 'will have to come to know'
ɪs=ke liye ye jan-na ho-g-a [ke **kya**
this.Obl=Gen.Obl for this.Nom know-Inf.M.Sg be-Fut-M.Sg that what
sacmuc koi nɑhĩ a-ya]
really someone not come-Perf.M.Sg
'For this, one needs to determine whether it is really the case that no one came.'

b. Neg + 'know'
koi nɑhĩ jan-t-a [ke **kya** tito stalɪn=se
someone not know-Impf-M.Sg that what Tito.M.Nom Stalin.M=Com
mɪl-e t^{h}-e]
meet-Perf.M.Pl be.Past-M.Pl
'Nobody knows whether Tito had met with Stalin.'

c. Imperative + 'know'
jan-ẽ [ke **kya** ap=ke bɑtʃtʃe=ke pas imel
know-Imp.2 that what you.Hon=Gen.Obl child.Obl=Gen.Obl near email
ɑkaunt hɛ]
account be.Pres.3.Sg
'Find out whether your child has an email account.'

Biezma et al. (2022) thus argue that polar *kya* is a marker of uncertainty, explaining why it is not licensed in contexts in which the speaker is certain (/knows the answer). In fact, as expected if *kya* expresses the attitude holder's uncertainty, *kya* is also possible with antirogatives with progressive marking (in which the attitude holder is still in the process of making up their mind and is thus still uncertain), see (16). The contrast in (16a) and (16b) supports the claim that when the attitude holder can conventionally convey uncertainty, polar *kya* is licensed.

(16) a. rɑvi sotʃ rɑh-ɑ t^{h}-a, ke **kya** vo
Ravi.M.Nom think Prog-M.Sg be.Past-3.Sg that what Pron.3.Sg.Nom
kɪtab amra=ko d-e-g-a?
book.F.Sg.Nom Amra.F=Dat give-3.Sg-Fut-M.Sg
'Ravi was thinking whether he will give a book to Amra.'

b. #rɑvi=ne sotʃ-a, ke **kya** vo kɪtab
Ravi.M.Nom think-Perf.M.Sg that what Pron.3.Sg.Nom book.F.Sg.Nom

amra=ko d-e-g-a?
Amra.F=Dat give-3.Sg-Fut-M.Sg
'Ravi thought whether he will give a book to Amra.'

That *kya* is a marker of uncertainty also explains why it is often claimed that *kya* does not contribute any meaning: in the unmarked situation, the speaker uttering a question is agnostic about the answer and, hence, can optionally use *kya*. To detect the meaning contribution of *kya* one needs to observe exactly those cases in which *kya* seems not to be available, as shown above, or when it contributes an additional meaning, as in (2), repeated below, where the speaker adds *kya* to convey sarcasm.

(2) Context: A is telling B how to behave in a situation. B says (with sarcasm):
B: tʊm mer-i ɑmmã ho **kya**?
you.Nom I.Gen-F.Sg mother.F.Sg.Nom be.Pres.2.Sg what
'Are you my mother?'

If *kya* were absent in (2), the rhetorical sense would not disappear, but speakers report that *kya* adds an additional layer of sarcasm. The analysis of *kya* as an expression of uncertainty explains this intuition: the sarcasm arises as a result of the speaker flouting quality by conventionally conveying that they are agnostic about the answer in a context in which it is mutually accepted by participants that they all know what the answer is.

Bhatt and Dayal (2020) argue that the distribution of polar *kya* provides information about what the question is ultimately about. That is, in asking a polar question regarding whether Ram gave a book to Sita yesterday, one can ultimately be asking whether it was a book or a toy, whether it was to Sita or someone else, whether it was Ram who gave the book to Sita or someone else or whether that Ram gave a book to Sita happened yesterday or any other day. One can also ask whether it happened at all or something completely different happened. Crucially, the position of *kya* helps to indicate which of these interpretations is intended:

(17) S IO *kya* Adv DO Verb
ram=ne sita=ko **kya** kɑl kɪtab d-i
Ram.M=Erg Sita.F=Dat what yesterday book.F.Sg.Nom give-Perf.F.Sg
'Did Ram give a/the book to Sita yesterday,...
(i) #ya mina=ne?
or Mina.F=Erg
'or had Mina?'
(ii) #ya vina=ko?
or Vina.F=Dat
'or to Vina?'
(iii) ya parsõ?
or day before yesterday
'or the day before yesterday?'
(iv) ya mɛgɛzin?
or magazine.F.Sg.Nom
'or a magazine?'

The pattern above indicates that *kya* is related to focus: the distribution of *kya* closely correlates with the consituent evoking alternatives relevant for the interpretation, the focus constituent. Constituents to the right of *kya* are available as targets for *kya* (and thus evoke alternatives relevant for the interpretation). With regular prosody, in which the last element on the utterance is prominent, constituents to the left of *kya* cannot be targets for expressions of uncertainty, but Biezma et al. (2022) show that with the right prosodic marking, elements to the left of *kya* can also be interpreted as targets for *kya*. Thus, in (18) the prosodically prominent indirect object to the left of *kya* can be the constituent that the speaker wonders about:

(18) Me: I know that Ravi gave a toy to someone . . .
. . . rɑvi=ne AMRA=KO **kya** kʰɪlona di-ya?
Ravi.M=Erg Amra.F=Dat *kya* toy.M.Sg.Nom give-Perf.M.Sg
'Did Ravi give a toy to Amra?'
ya mina=ko?
or Mina.F=Dat
'or to Mina?'

Biezma et al. (2022) show that by default *kya* appears to the left of the focus constituent, either immediately to its left or in sentence initial position. In the latter case the main predicate is in focus by default. However, *kya* can also appear to the right of a constituent if that constituent is marked as prominent via prosodic means. Ultimately, *kya* behaves like a focus sensitive particle, e.g., like English *only*. In fact, Biezma et al. (2022) show that, in the right context, *kya* can also appear in the otherwise marked preverbal position. For example, in (19) the context allows us to understand that the relevant alternatives are alternatives to the main predicate, as such the verb is the focus, and *kya* can appear immediately before it.

(19) Sita: I'm not sure whether Ravi lent or gave a toy to Amra,
. . . (✓*kya*) rɑvi=ne (%*kya*) amra=ko (%*kya*) kʰɪlona
what Ravi.M=Erg what Amra.F=Dat what toy.M.Sg.Nom
(✓*kya*) DI-YA (?*kya*)?
what give-Perf.M.Sg what
'Did Ravi give a toy to Amra?'

Biezma et al. (2022) conclude that *kya* is a focus sensitive operator and model this within the Roothian tradition (Rooth 1985, 2016), as shown on a very general level in (20). The symbol Φ stands for a syntactic expression and ⟦Φ⟧ for a set of propositions corresponding to that syntactic expression. The '~' is the Roothian operator linking the utterance to the context of utterance. The meaning of Q is from Biezma and Rawlins (2012) and Biezma (2020), which handles question interpretation via the identification of salient alternatives to the proposition.

(20) ⟦[$_Q$[***kya*** [~ Φ]]]⟧= ⟦[$_Q$~ Φ]⟧
defined only if the speaker believes that there is more than one live salient alternative in the context of the utterance in the set of possible answers.

The idea is that an interrogative with *kya* has the same denotation as the interrogative without *kya*. However, *kya* imposes the condition that there are different possible answers compatible with the attitude's holder doxastic alternatives (i.e., the attitude holder is 'uncertain'). This is captured in the definedness conditions in (20). The live alternatives that are answers to the question are also focus alternatives (following standard Roothian assumptions system). In Biezma et al.'s (2022) analysis *kya* is like any other focus sensitive operator (e.g., *only*) and they take an approach to focus sensitive particles in which these may appear in different places in the clause but their surface appearance simply signals the pressence of an operator in the clause that takes propositions as arguments, as in (20). This allows Biezma et al. (2022) to keep a single semantics for *kya*. Crucially, given *kya*'s subtle meaning contribution, the semantic effect of *kya* goes often unnoticed, but the proposal in Biezma et al. (2022) uncovers why it is not licensed in contexts in which the speaker is confirming what they believe to be the answer, and why it helps to convey surprise (with *kya* the speaker conveys that they still do not believe the answer) or sarcasm (*kya* is used to flout quality; the speaker uses *kya* to convey that they do not know the answer while it being mutually accepted that they do).

4 Syntactic Analysis

The analysis by Biezma et al. (2022) assumes a syntax that will deliver the right structures for a calculation of the semantics and the different possibilities for focus association. In this section we provide those syntactic underpinnings, using LFG's multidimensional architecture and lexicon. Analyses within other syntactic frameworks are, of course, possible and not precluded. However, we hope to demonstrate that LFG's projection architecture allows for a comparatively straightforward account of polar *kya*. In particular, we integrate crucial information from prosody via Bögel's (2015) conception of the prosody-syntax interface.

We follow the syntactic analyses as established as part of the Urdu ParGram grammar (Butt et al. 1999, Butt and King 2007), working with a flat exocentric structure where the main clause is represented by an S. There is no conclusive evidence for a VP in Urdu/Hindi (Butt 1995), but Urdu/Hindi does have a fairly complicated verbal complex consisting of main verbs or complex predicates in combination with modals and auxiliaries (Bhatt et al. 2011, Butt et al. 2016). The exocentric structure models the constituency of a clause in that all major constituents within an S may be scrambled among themselves, giving rise to information structural effects (Gambhir 1981, Butt and King 1997, Kidwai 2000, Manetta 2012).[2] Case clitics are treated as heads of noun phrases, resulting in a KP when a noun phrase is overtly marked (Butt and King 2004).

[2]Urdu/Hindi phrase structure constraints actually resemble those described for Ossetic quite closely (Belyaev 2022). The insights from Ossetic need to be integrated into future work.

4.1 Analysis of *kya*

We begin with the treatment of *kya* itself. With respect to this multifunctional item, we have several options. One option is to posit a separate entry for each of the uses of *kya*, e.g., wh-constituent vs. polar *kya*. However, it is more parsimonious to work with just one underlying lexical entry. We so far have no evidence for independent entries, even when considering the other uses of *kya*.[3] Prominently among these is the wh-counterpart of the *scope marking* construction (Dayal 1996, 2000), where *kya* has been analyzed as licensing the matrix scope of the wh-in-situ in an embedded clause. Butt (2014) provides a brief analysis of this within LFG that is based on Dayal's (1996, 2000) original insights.

Further uses of *kya* are as: 1) an adjunct, roughly translating to 'what's the point' as in (21); 2) in alternative questions as in (22) (Han and Romero 2004).

(21) ɑb mẽ us=se **kya** mıl-ũ?
now I.Nom Pron.Obl=Inst what meet-Subj.1.Sg
'What's the point of meeting with him/her now?'
(lit. What should I meet with him/her?)

(22) (**kya**) candra=ne kofi p-i ya cai
what Chandra.F=Erg coffee.F.Nom drink-Perf.F.Sg or tea.F.Nom
'Did Chandra drink tea or coffee?'

Biezma et al. (2022) analyze the interaction of *kya* and alternative questions as part of their overall analysis of polar *kya*. We believe that interpretations as in (21) will also follow from that overall analysis of polar *kya* as a marker of uncertainty so that no special information or entries for *kya* are necessary for (21) and (22).

(23) kyA Q **Disjunct 1** (wh-kya)
(↑ PRED) = 'kyA'
(↑ QUESTION-TYPE = constituent
(↑ NTYPE NSYN) = pronoun
(↑ PRON-TYPE) = int
(↑ CASE) = nom

Disjunct 2 (polar kya)
(↓ PRED) = 'kyA'
(↑ UNCERTAINTY-OP) = ↓
(↑ QUESTION-TYPE) = polar

Following the XLE grammar development convention (Crouch et al. 2017) of one head word per lexical entry, we propose the lexical entry in (23) to cover both polar *kya* and wh-constituent *kya*. This lexical entry consists of a head word 'kyA', which is our 'what' spelled according to the computationally motivated translit-

[3] A question arises with respect to potential language change and whether the different uses of *kya* would eventually lead to the loss of polar *kya* being associated underlyingly with the meaning of 'what'. In current Urdu/Hindi, native speakers consistently identify/translate *kya* as 'what' in each of its uses, indicating that the connection of polar *kya* with the wh-constituent remains strong.

eration scheme of the Urdu Perseo-Arabic based script (Malik et al. 2010). We use this transliteration as we are implementing our analysis computationally in line with the Urdu ParGram grammar (Butt et al. 1999, Butt and King 2007). However, the lexical entry for *kya* contains at least the two disjuncts in (24) (plus one for the scope marking use, not shown in the context of this paper). These account for the differences in their functionality. We treat all the uses of *kya* in terms of a Q at c-structure (roughly in line with Slade (2011) for Sinhala questions).[4] The wh-words in Urdu/Hindi have a very distinct syntactic distribution from the determiner *ek* 'one', quantifiers and other nominal specifiers, causing us to assign wh-words a dedicated part-of-speech analysis: Q.

In (23), the Q has a semantic predicate whose value is 'kyA' in all its instantiations, but there are differences across the two disjuncts. The first disjunct is just the wh-constituent option. The *kya* carries the functional information that this is an interrogative pronoun in the nominative case.

In contrast, the second disjunct registers that this is part of a polar question and that the *kya* functions as an uncertainty operator. This f-structural information provides the basis for the semantic interpretation sketched in section 3.3. We do not here provide the 'glue' between the final semantic interpretation and the f-structural information, but the calculation of the semantics from the f-structural information should follow straightforwardly, particularly within glue semantics (Dalrymple 1999, Asudeh 2022).

4.2 Prosodic Disambiguation

The disjunctions in the lexical entry for *kya* inherently allow for ambiguity and, as we have seen, there are examples which in principle allow for two readings (cf. (10) and (11)). We here illustrate the disambiguation via prosodic information with respect to (24), in which the two readings shown in (25) depend on whether the *kya* is interpreted as a constituent question where the 'what' is part of the NP headed by 'present', or whether it functions as a marker of uncertainty in a polar question. In the polar use, we analyze the *kya* as a direct daughter of S.[5]

(24) ʃahina=ne naz=ko **kya** tofa di-ya
Shahina=Erg Naz=Acc what present.M.Sg give-Perf.M.Sg
a) Constituent Question: 'What gift did Shahina give to Naz?'
b) Polar Question: 'Did Shahina (actually) give a gift to Naz?'

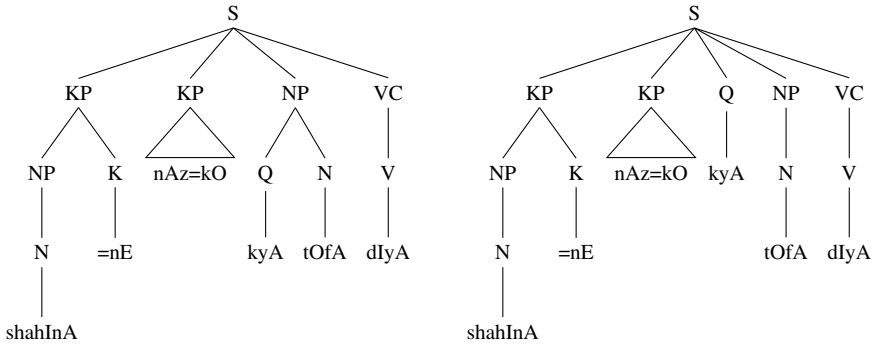
The two readings in (25) can be clearly disambiguated once prosody is taken into account. Recall from section 3.1 that *kya* is associated with different prosodic properties depending on its use. Like all other wh-words, the wh-constituent use

[4]This c-structure Q should not be confused with the Q operator assumed for LF-based approaches.

[5]We show the KP analysis in full at least once in the c-structures to show that the case clitic is analyzed as having an independent terminal node, following Butt and King (2004). A reviewer asks about how exactly the Q fits into the NP. The internal syntactic structure of the Urdu/Hindi NP is quite complex and well beyond the scope of this paper.

of *kya* is associated with a high tone (H*). In contrast, the polar *kya* is realized with a flat or falling tone. Thus, while the string in (24) is in principle ambiguous, prosodic information distinguishes between the two possibilities.

(25) **Wh-Question** **Polar kya**



We therefore need to integrate prosodic information into the analysis and we propose to do this via Bögel's (2015) conception and computational implementation of the prosody-syntax interface. The overall analysis in this section is based on Butt et al. (2020) and the reader is referred to Bögel (2015) and Butt et al. (2020) for full details of the prosodic disambiguation via the prosody-syntax interface.

Bögel regulates the exchange of information of syntactic vs. prosodic phrasing via the *Transfer of Structure*. As is well-known, while major prosodic and syntactic phrases tend to line up, the prosody-syntax interface is prone to mismatches. Information on such mismatches can be communicated via this interface, as well as information that can help determine syntactic and prosodic phrasing.

The so-called *Transfer of Vocabulary* in Bögel's system associates morphosyntactic and phonological information with lexical elements, rendering the multidimensional lexicon that is characteristic of LFG. This categorial, functional and phonological information is projected to the relevant representations within the LFG architecture (e.g., c-structure, f-structure and p-structure, respectively). Sample lexical entries under this architecture are shown in Table 2, with the familiar c- and f-structural information for the items *tofa* 'present' and *kya* 'what' shown in the left column (the syntactic information for *kya* is an abbreviation of the full lexical entry in (23)) and the phonological information in the right column.

The phonological information includes a p-form, which is an IPA rendering. It also provides information about what segments are involved and knowledge about the metrical properties of the item. For example, the noun *tofa* 'present' consists of two syllables which combine into a single prosodic word whose first syllable is stressed. In contrast, the *kya* 'what' underlyingly consists of a single unstressed syllable that does not necessarily form a prosodic word, but may cliticize onto another prosodic word. Initial evidence in our data points towards polar *kya* being able to cliticize. This is not possible for the wh-constituent version.

The prosodic information necessary for disambiguation is extracted from the

s(yntactic)-form			**p(honological)-form**	
tOfA N	(↑ PRED)	= 'tOfA'	P-FORM	[tofa]
	(↑ NUM)	= sg	SEGMENTS	/t o f a/
	(↑ GEND)	= masc	METR. FRAME	$('\sigma\sigma)_\omega$
kyA Q	{ (↑ QUESTION-TYPE)	= polar	P-FORM	[kja]
	| (↑ QUESTION-TYPE)	= constituent}	SEGMENTS	/k j a/
	...		METR. FRAME	σ

Table 2: Sample Lexical Entries in the Multidimensional Lexicon

speech signal, which Bögel represents as a series of vectors per syllable, as shown in Figure 2. Each syllable receives an index (for purposes of identification) and a "value" that is the IPA representation of the syllable.

PHRAS.	$(_\iota\ (_\varphi$	...	...	...	...	...	...	...	...	...	$)_\varphi)_\iota$	INTERPRETATION
TOBI	...	...	...	...	...	...	H*	...	...	...	...	↓
DUR.	0.25	0.17	0.18	0.14	0.31	0.13	0.24	0.15	0.19	0.14	0.16	SIGNAL
F_{0_mean}	193	200	222	241	198	231	248	224	193	174	205	↓
VALUE	[sha]	[hi]	[na]	[ne]	[naz]	[ko]	[kja]	[toh]	[fa]	[di]	[ja]	
INDEX	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	...

Figure 2: Syllable-wise analysis of the speech signal corresponding to (24)

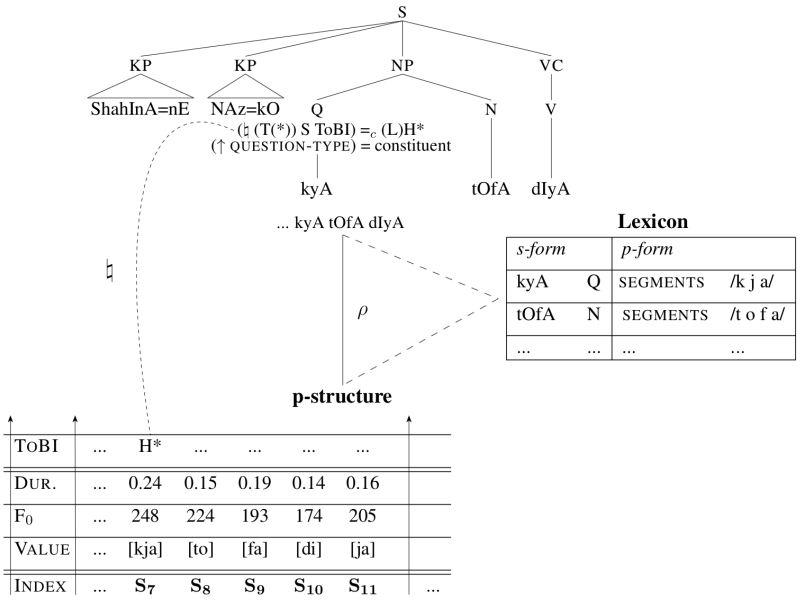
Relevant information for each syllable is extracted from the speech signal. This typically includes information about the F_0 value and the duration of the syllable. This "raw" phonetic information is used as the basis for the calculation of phonological, categorical information, such as whether a syllable can be considered to have a high tone (e.g., H* on the [kja], following the ToBI annotation (Silverman et al. 1992)), or the prosodic phrasing of the syllables. The vector associated with a syllable thus contains phonetic information as well as a phonological analysis.

We are now ready to integrate the prosodic information that is relevant for the disambiguation of wh-constituent *kya* vs. polar *kya*. This can be effected via annotations on the respective Q nodes which check (and ensure) that the instantiations of *kya* carry the right prosody. For wh-constituent *kya* this entails ensuring an (L)H*[6] as per the annotation in (26), which reads as follows: for all terminal nodes T under the current node * (in this case this is the Q node), for any syllable(s) S corresponding to this terminal node, the value for the attribute ToBI is constrained to be (L)H* (via the constraining equation with $=_c$). The $\natural$ represents the function

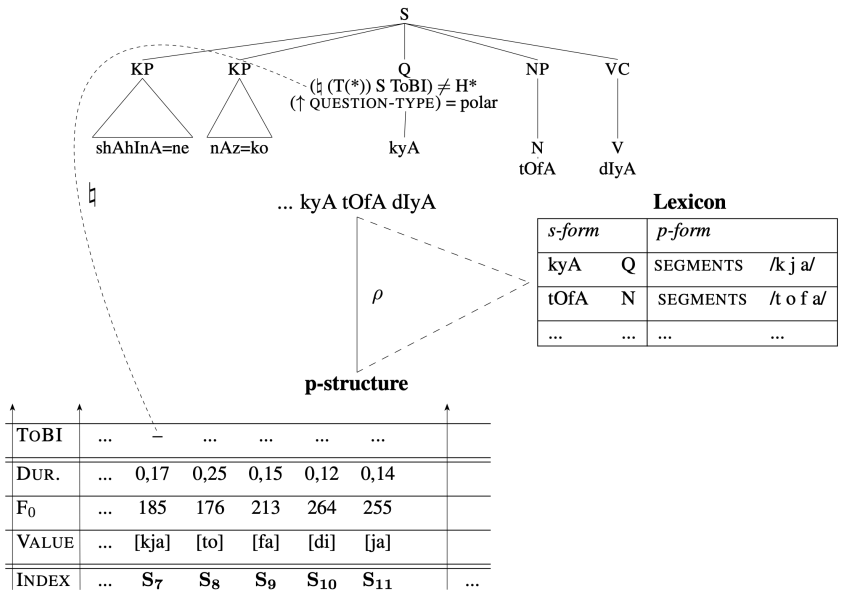
[6]Generally the wh-words carry an H*, for longer ones an LH* can also be observed, following the typical LH pattern of prosodic units in the languagae (Urooj et al. 2019).

between p-structure (the vector representation of the phonetic and phonological information) and c-structure, as illustrated in the full analysis in (26).

(26) Q
(ŧ (T(*)) S ToBI) =$_c$ (L)H*



(27) Q
(ŧ (T(*)) S ToBI) ≠ H*



Thus, if and only if there is an H* associated with the *kya* at p-structure can the *kya* be analyzed as a constituent question word. In contrast, the annotation for polar *kya* is as shown in (27). This equation says that for each terminal node T under the current node * (again, this is the Q node), for any syllable(s) S corresponding to this terminal node, the value for the attribute ToBI is required to be anything but H* (via the negative equation ≠). A polar *kya* reading is thus only possible in the absence of an H* on the *kya*, yielding the structure in (27).

4.3 Distribution and Scope of Polar *kya*

We are now ready to account for the clausal distribution and scope of polar *kya*. Recall from (12) that the *kya* can in principle occur anywhere in the clause, which is why we have analyzed it as a direct daughter of S that can scramble just like any other major constituents of a clause. Biezma et al. (2022) establish that the different positions of *kya* in the clause correlate with different interpretational possibilities. In the next subsections we go through these possibilities.

4.3.1 Clause Medial Polar *kya*

We begin with the situations in which polar *kya* appears anywhere within a clause, leaving aside the clause-initial and clause-final positions for the moment. As shown by Biezma et al. (2022), by default the polar *kya* associates with the constituents its right (e.g., 'book' in (28)), but other targets for the expression of uncertainty are also possible. In general, as a focus sensitive operator, polar *kya* can associate with any focused (generally prosodically prominent) item in the clause.

(28) ram=ne sita=ko **kya** kɪtab kɑl d-i
Ram.M=Erg Sita.F=Dat what book.F.Sg.Nom yesterday give-Perf.F.Sg
t^h-i
be.Past-F.Sg
'Had Ram given a/the book to Sita yesterday?

Thus possible targets for polar *kya* in (28) are: 1) 'book', 'yesterday' or the verb (by position); 2) Ram or Sita (by prominence marking). This is a fairly complex situation, but one that can be modeled via LFG's capability for indicating scope[7] via f-precedence and the use of a Metarulemacro[8] that identifies prosodically prominent items in a clause. In our analysis we use scopal relations at f-structure as an indication of what items the *kya* can target via focus association.

For the case in which *kya* takes "scope" via position, this is easily modeled via f-precedence in that we state that the *kya*, whose function is registered as an uncertainty operator at f-structure, can take scope (>s) over anything that it f-precedes. As XLE has not implemented full f-precedence, but a weaker version

[7] https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/notations.html#N4.2.10
[8] https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/notations.html#N3.5

of head precedence,[9] we work with this in our implementation. For example, if *kya* targets 'book', we have a situation in which the uncertainty operator h(ead)-precedes the OBJ, and this gives the *kya* "scope" (>s) over the OBJ. The two lines of annotation in (29) can be read as an if-then statement: if the uncertainty operator h-precedes the OBJ, then the uncertainty operator has scope over the OBJ.

(29) (↑ UNCERTAINTY-OP) <h (↑ OBJ)
(↑ UNCERTAINTY-OP) >s (↑ OBJ)

The annoations in (29) must, of course, be generalized across all possible grammatical functions (grammatical relations and adjuncts, henceforth abbreviated as GF-ADJ) and the scope relationship between *kya* and the other major constituents in the clause is determined for each of those items.

For example, the full analysis for an association of *kya* with 'book' is shown in Figure 3 and (30). The identification of *kya* as a marker of uncertainty rather than a wh-constituent is achieved through interfacing with prosodic information on the Q node, as discussed in section 4.2.[10] In addition to those annotations, the information about the potential scope of *kya* in (29) is added to the Q node. In the full implementation this is done via a template that is associated with clause medial *kya* and which records all the possible scope relationships at f-structure.[11]

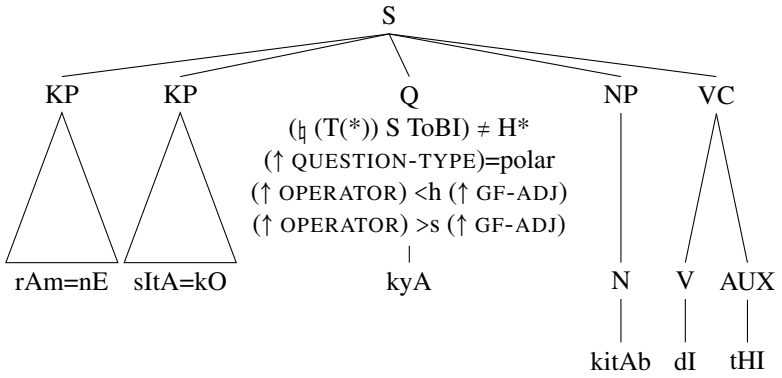


Figure 3: Clause medial *kya* targeting item to its right

[9]The definition of head precedence is that an f-structure f1 head precedes an f-structure f2 (f1 <h f2) if and only if f1 and f2 have heads and the head of f1 precedes the head of f2 in the c-structure, see https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/notations.html#N4.2.9

[10]The observant reader will note that the information that polar *kya* corresponds to a polar question-type is provided twice in the analysis. Once as part of the f-structure annotation on Q and once as part of the lexical entry of *kya*. Information in both places is necessary under Bögel's conception of the prosody-syntax interface, which allows for doing both comprehension (the basic case assumed here: there is a speech signal and it needs to be parsed) and production. From the production perspective, the annotation on Q is necessary as it must be matched to the right version of *kya* so that the correct prosodic realization can be produced.

[11]See Crouch et al. (2017), Dalrymple et al. (2004) and Asudeh et al. (2013) on templates in LFG.

(30)
$$\begin{bmatrix} \text{PRED} & \text{`dE}\langle\text{SUBJ, OBJ-GO, OBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED `Ram'}\end{bmatrix} \\ \text{OBJ-GO} & \begin{bmatrix}\text{PRED `Sita'}\end{bmatrix} \\ \text{OBJ} & \begin{bmatrix}\text{PRED `kitAb'}\end{bmatrix} \\ \text{UNCERTAINTY-OP} & \begin{bmatrix}\text{PRED `kyA'} \\ \text{>s [kitAB]}\end{bmatrix} \\ \text{CLAUSE-TYPE} & \text{interrogative} \\ \text{QUESTION-TYPE} & \text{polar} \end{bmatrix}$$

This results in the f-structure in (30), which is a standard f-structure except for the registration of an uncertainty operator and the information that this operator scopes over *kitAb* 'book'. The '[kitAb]' in (30) mirrors the actual XLE output and is short-hand in that it points to the f-structure of the OBJ. That is, 'book' has been identified as a target for focus association and thereby as the target of uncertainty.

We now turn to the situation in which *kya* associates with an item not to its immediate right, but which is prosodically prominent. Querying the necessary information about prosodic prominence from the p-structure can be achieved via a Metarulemacro. A Metarulemacro applies to every rule in a grammar. A classic example is same-category coordination: a Metarulemacro is defined which applies to all categories contained in the grammar and the grammar is automatically extended to allow coordination of each of these categories. That is, the possibility of allowing for coordination does not need to be expressed explicitly for all categories (e.g., N, NP, PP), but is applied automatically as needed across all rules of the grammar. The rule is thus stated at a "meta" level.

Similarly, we can define a Metarulemacro that checks for prosodic prominence. We say that for any c-structure category X in the grammar, we check (via the constraining equation with $=_c$) if there is a high tone (H) and if this high tone has the value 4, which is the highest value defined for H in Bögel's current prosody-syntax interface (Bögel and Raach 2020). If such a high tone is found to be associated with the XP, then the Metarulemacro registers a [PROM +] feature at f-structure (the interpretation of prosodic prominence varies according the discoursal context). Note that this is merely for convenience so that the prosody-syntax interface need not be continually checked. Once an H4 has been identified, the grammar can work with the PROM feature as part of its syntactic representation. The relevant information for the Metarulemacro is shown in (31). This checks whether any syllable S associated with a terminal node T has an H at p-structure and whether the level of prominence is 4. If so, the category is marked as [PROM +].

(31) $(\natural(T(*))\text{S}_{any}$ ToBI)$=_{\mathbf{c}}$ H
$(\natural(T(*))\text{S}_{any}$ PROMINENCE)$=_{\mathbf{c}}$ 4
$(\uparrow$ PROM) = +

We can now add annotations to the polar *kya* to check whether there is a promi-

nent item in the clause with which the *kya* could associate, as in (32), with the GF-ADJ abbreivating all grammatical relations and adjuncts.

(32) (↑GF-ADJ) = %F
(%F PROM) =c +
(↑ UNCERTAINTY-OP) >s (%F PRED)

The %F in (32) is a variable name that refers to an f-structure (e.g., ↓). We use a variable here to ensure that whichever f-structure is accessed by the first line of (32) (e.g., (↑ OBJ) or (↑ SUBJ) will end up being the same f-structure whose PRED the polar *kya* is marked as being able to associate with via the scope mechanism.[12] Figure 4 shows a sample analysis for a prominent item to the left of *kya*. In this case the prominent item has been identified as Ram.

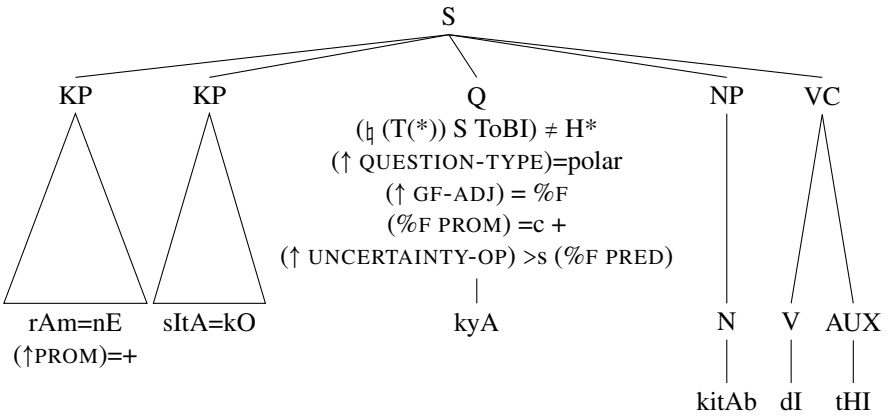


Figure 4: Clause medial *kya* associating with a prominent item to its left

(33)
$$\begin{bmatrix} \text{PRED} & \text{`dE}\langle\text{SUBJ, OBJ-GO, OBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED `Ram'}\end{bmatrix} \\ \text{OBJ-GO} & \begin{bmatrix}\text{PRED `Sita'}\end{bmatrix} \\ \text{OBJ} & \begin{bmatrix}\text{PRED `kitAb'}\end{bmatrix} \\ \text{UNCERTAINTY-OP} & \begin{bmatrix}\text{PRED `kyA'} \\ \text{>s [Ram]}\end{bmatrix} \\ \text{CLAUSE-TYPE} & \text{interrogative} \\ \text{QUESTION-TYPE} & \text{polar} \end{bmatrix}$$

4.3.2 Clause Initial Polar *kya*

In the clause initial position, polar *kya* precedes any combination of grammatical functions and adjuncts (NPs, KPs, AdvPs) as well as the verbal complex; see (34).

[12] In effect this works out to be a version of Roothian focus marking.

(34) **kya** ram=ne sita=ko kıtab d-i t^{h}-i
what Ram.M=Erg Sita.F=Dat book.F.Sg.Nom give-Perf.F.Sg be.Past-F.Sg
'Had Ram given a/the book to Sita?

Our data show that clause initial polar *kya* always associates with the verb, which is the most prominent item in a polar question under default prosody. That is, as a focus sensitive item *kya* associates with the item that is the default focus of a polar question: the verb. Given that the verb provides the PRED of the f-structure corresponding to the clause and therefore encompasses this entire f-structure for purposes of focus (see King 1997), the domain of focus may be extended beyond the verb to the other items licensed by the PRED and contained at the same level of f-structure as the PRED. Thus, *kya* may associate with the verb alone or with larger domains including the verb, e.g., the verb and the subject.

We again rely on scope marking for an association of polar *kya* with the verb. In this case, we can simply state that the scope of *kya* is over the PRED of the clause (the ↑ in (35) resolves to the f-structure associated with the S node).

(35) (↑ UNCERTAINTY-OP) >s (↑ PRED).

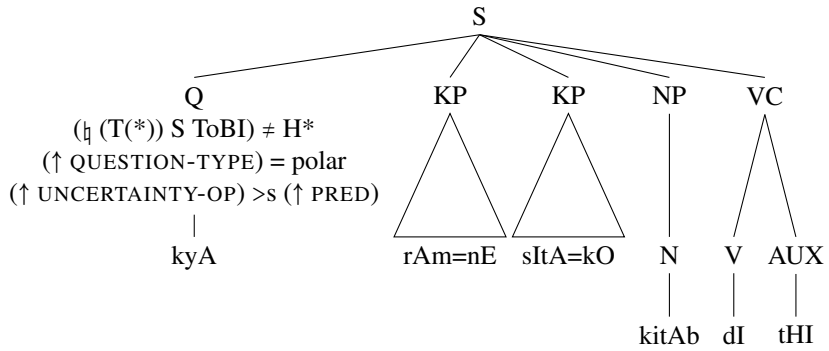


Figure 5: Clause initial *kya*

(36)
$$\begin{bmatrix} \text{PRED} & \text{'dE}\langle \text{SUBJ, OBJ-GO, OBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix}\text{PRED 'Ram'}\end{bmatrix} \\ \text{OBJ-GO} & \begin{bmatrix}\text{PRED 'Sita'}\end{bmatrix} \\ \text{OBJ} & \begin{bmatrix}\text{PRED 'kitAb'}\end{bmatrix} \\ \text{UNCERTAINTY-OP} & \begin{bmatrix}\text{PRED 'kyA'} \\ \text{>s [dE]}\end{bmatrix} \\ \text{CLAUSE-TYPE} & \text{interrogative} \\ \text{QUESTION-TYPE} & \text{polar} \end{bmatrix}$$

As before, the annotation on the Q node has the effect that the scope information

becomes part of the f-structural analysis and can be passed along to the semantic component for further analysis and interpretation. Concretely, the f-structure in (36) contains the information that the *kya* 'what' associates with the verb *de* 'give' and identifies that as the locus of uncertainty.

4.3.3 Clause Final Polar *kya*

Clause final polar *kya* is covered by the phrase structure rule in (37). We have exactly the same situation as with clause initial *kya* in that the other major constituents of the clause can appear in any order. This is modeled by the shuffle operator ",".[13]

(37) S $\longrightarrow$ {KP, NP, AdvP, VC} Q.

However, the clause final position of *kya* leads to a subtle difference in interpretation. As Biezma et al. (2022) show, in this position polar *kya* always pertains to the entire proposition. Associating with the entire proposition vs. only with the main predicate of the clause (=clause initial position) can work out to be semantically very similar since focus association with the main predicate means also being able to extend the focus domain to cover its dependents, in effect pretty much the entire proposition. However, there is a semantic difference and this becomes clear in the interaction of polar *kya* with alternative questions as in (38) and (39).

(38) **kya** candra=ne kofi ya cai p-i?
what Chandra.F=Erg coffee.F.Nom or tea.F.Nom drink-Perf.F.Sg
'Did Chandra drink tea or coffee?'
Alternative Question: Did Chandra drink tea or did she drink coffee?
Polar Question: Is it the case that Chandra drank either tea or coffee?'

(39) candra=ne kofi ya cai p-i **kya**?
Chandra.F=Erg coffee.F.Nom or tea.F.Nom drink-Perf.F.Sg what
'Did Chandra drink tea or coffee?'
*Alternative Question: Did Chandra drink tea or did she drink coffee?
Polar Question: Is it the case that Chandra drank either tea or coffee?'

Bhatt and Dayal (2020) note that clause initial polar *kya* allows for both a polar and an alternative question reading in (38), but that a clause final polar *kya* as in (39) only permits a polar question. Biezma et al. (2022) argue that the answer lies in the difference in the focus association possibilities of polar *kya* as correlated with position. Because clause initial *kya* focus associates with the verb, in (38) it is compatible with both 1) the reading in which there is one predicate in the clause and the disjunction is over the alternative NPs (polar); 2) the reading in which there are two predicate disjuncts ('drink tea' or 'drink coffee') in which *kya* can associate with one or the other of the predicates (alternative question). In contrast, clause final *kya* associates with the entire proposition, so it cannot target individual parts of the disjunction and therefore only the polar reading is available.

[13] https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/xle_toc.html#N1.14

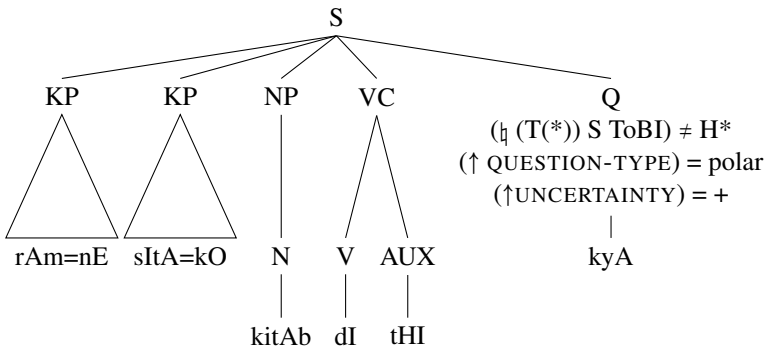


Figure 6: Clause final *kya*

(40)
$$\begin{bmatrix} \text{PRED} & \text{`dE}\langle\text{SUBJ, OBJ-GO, OBJ}\rangle\text{'} \\ \text{SUBJ} & \left[\text{PRED `Ram'}\right] \\ \text{OBJ-GO} & \left[\text{PRED `Sita'}\right] \\ \text{OBJ} & \left[\text{PRED `kitAb'}\right] \\ \text{UNCERTAINTY-OP} & \left[\text{PRED `kyA'}\right] \\ \text{UNCERTAINTY} & + \\ \text{CLAUSE-TYPE} & \text{interrogative} \\ \text{QUESTION-TYPE} & \text{polar} \end{bmatrix}$$

We model this as in Figure 6 and (40). In addition to the annoations for the prosody-syntax interface, the feature [UNCERTAINTY +] is introduced when *kya* is clause final. This marks the entire f-structure associated with the clause as uncertain. The [UNCERTAINTY +] is not a semantic analysis, it is a feature that needs to be passed on to the semantic component, for example via a glue semantic approach (Dalrymple 1999, Dalrymple et al. 2020, Asudeh 2022) or more generally via a resource based approach that treats the f-structure as containing the information necessary for a calculation of the clausal semantics (Bobrow et al. 2007).

All the information necessary for the semantic analysis sketched in (20) is present in (40). The fact that this is a question opening up alternatives (Q) can be deduced from QUESTION-TYPE. The fact that *kya* needs to focus associate with the entire proposition can be deduced from the entire f-structure being marked as [UNCERTAINTY +], whereby the Φ is can be instantiated to whichever f-structure contains [UNCERTAINTY +], in this case the full f-structure.

5 Summary and Conclusions

We have now presented a fairly complete account of polar *kya*. We sketched the prosodic and semantic insights arrived at in previous work (Butt et al. 2017, 2020,

Biezma et al. 2022) and developed a syntactic analysis. We adopted the flat, exocentric approach to Urdu/Hindi c-structure from the Urdu ParGram grammar and showed how the clause-internal distribution and correlated interpretational differences can be accounted for via registering scope at f-structure. Items that the *kya* is shown to take scope over at f-structure are those which are available for the focus association of *kya*. We follow Biezma et al. (2022) in analyzing polar *kya* as a focus sensitive item which associates with either a whole proposition or parts of a proposition to indicate uncertainty on the part of the speaker about that proposition or parts of that proposition.

Overall we have presented the groundwork for an end-to-end analysis with an architecture that extracts information directly from the speech signal and interfaces this phonological information with the morphosyntax, on the basis of which the interpretational possibilities can be calculated. The syntactic analysis is designed so that all information necessary for a semantic analysis is represented at f-structure. Essentially this represents a resource based approach to semantics, or also what is known as 'description by analysis', whereby the description of one structure is obtained by the analysis of another structure (Halvorsen 1983, Dalrymple 2001).

References

Asudeh, Ash. 2022. Glue Semantics. *Annual Review of Linguistics* 8, 321–341.

Asudeh, Ash, Dalrymple, Mary and Toivonen, Ida. 2013. Constructions with lexical integrity. *Journal of Language Modelling* 1(1), 1–54.

Bayer, Josef and Cheng, Lisa Lai-Shen. 2015. Wh-in-situ. In Martin Evaraert and Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, Oxford: Blackwell Publishing.

Belyaev, Oleg. 2022. Phrase structure and configurationality in Ossetic. In Miriam Butt, Jamie Y. Findlay and Ida Toivonen (eds.), *Proceedings of the LFG'22 Conference*.

Bhatt, Rajesh, Bögel, Tina, Butt, Miriam, Hautli, Annette and Sulger, Sebastian. 2011. Urdu-Hindi Modals. In M. Butt and T.H. King (eds.), *Proceedings of the LFG'11 Conference*, pages 47–67, Stanford, CA: CSLI Publications.

Bhatt, Rajesh and Dayal, Veneeta. 2020. Polar question particles: Hindi-Urdu *kya:*. *Natural Language and Linguistic Theory* 38, 1115–1144.

Biezma, María. 2020. Non-informative assertions: The case of non-optional *wh*-in-situ. *Semantics & Pragmatics* 13(18).

Biezma, María, Butt, Miriam, Jabeen, Farhat and Mumtaz, Benazir. 2022. Urdu/Hindi polar *kya* as an expression of uncertainty, Manuscript submitted for review, https://lingbuzz.net/lingbuzz/006997.

Biezma, Maria and Rawlins, Kyle. 2012. Responding to alternative and polar questions. *Linguistics and Philosophy* 35, 361–406.

Bobrow, Daniel G, Cheslow, Bob, Condoravdi, Cleo, Karttunen, Lauri, King, Tracy Holloway, Nairn, Rowan, de Paiva, Valeria, Price, Charlotte and Zaenen, Annie. 2007. PARC's bridge and question answering system. In *Proceedings of the GEAF 2007 Workshop. CSLI Studies in Computational Linguistics Online*.

Bögel, Tina. 2015. *The Syntax-Prosody Interface in Lexical Functional Grammar*. Ph. D. thesis, University of Konstanz.

Bögel, Tina and Raach, Lea. 2020. Swabian *ed* and *edda*: Negation at the interfaces. In

Miriam Butt and Ida Toivonen (eds.), *Proceedings of the LFG'20 Conference*, pages 47–67, Stanford, CA: CSLI Publications.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford, CA: CSLI Publications.

Butt, Miriam. 2014. Questions and information structure in Urdu/Hindi. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'14 Conference*, pages 158–178, Stanford, CA: CSLI Publictions.

Butt, Miriam, Bögel, Tina and Jabeen, Farhat. 2017. Polar *kya* and the Prosody-Syntax-Pragmatics Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'17 Conference*, pages 125–145, Stanford, CA: CSLI Publications.

Butt, Miriam, Jabeen, Farhat and Bögel, Tina. 2016. Verb cluster internal wh-phrases in Urdu: prosody, syntax and semantics/pragmatics. *Linguistic Analysis* 40(3–4).

Butt, Miriam, Jabeen, Farhat and Bögel, Tina. 2020. Ambiguity resolution via the syntax-prosody interface: the case of *kya* 'what' in Urdu/Hindi. In Gerrit Kentner and Joost Kremers (eds.), *Prosody in Syntactic Encoding*, pages 85–118, Berlin: Walter de Gruyter.

Butt, Miriam and King, Tracy H. 1996. Structural Topic and Focus without Movement. In M. Butt and T. H. King (eds.), *Proceedings of the First LFG Conference*, Stanford, CA: CSLI Publications.

Butt, Miriam and King, Tracy H. 2007. Urdu in a Parallel Grammar Development Environment. In T. Takenobu and C.-R. Huang (eds.), *Language Resources and Evaluation: Special Issue on Asian Language Processing: State of the Art Resources and Processing*, volume 41, pages 191–207.

Butt, Miriam and King, Tracy Holloway. 1997. Null Elements in Discourse Structure, written to be part of a volume that never materialized, http://ling.uni-konstanz.de/pages/home/butt/.

Butt, Miriam and King, Tracy Holloway. 2004. The Status of Case. In Veneeta Dayal and Anoop Mahajan (eds.), *Clause Structure in South Asian Languages*, pages 153–198, Berlin: Kluwer Academic Publishers.

Butt, Miriam, King, Tracy Holloway, Niño, María-Eugenia and Segond, Frédérique. 1999. *A Grammar Writer's Cookbook*. Stanford, CA: CSLI Publications.

Crouch, Dick, Dalrymple, Mary, Kaplan, Ronald M., King, Tracy Holloway, Maxwell III, John T. and Newman, Paula. 2017. *XLE Documentation*. Palo Alto Research Center.

Dalrymple, Mary (ed.). 1999. *Semantics and Syntax in Lexical Functional Grammar: The Resource Logic Approach*. Cambridge, MA: The MIT Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. New York: Academic Press.

Dalrymple, Mary, Kaplan, Ronald M. and King, Tracy Holloway. 2004. Linguistic generalizations over descriptions. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'04 Conference*, pages 199–208, Stanford, CA: CSLI Publications.

Dalrymple, Mary, Patejuk, Agnieszka and Zymla, Mark-Matthias. 2020. XLE+Glue – A new tool for integrating semantic analysis in XLE. In Miriam Butt and Ida Toivonen (eds.), *Proceedings of the LFG'20 Conference*, pages 89–108, Stanford, CA: CSLI Publications.

Dayal, Veneeta. 1996. *Locality in WH Quantification*. Dordrecht: Kluwer Academic Publishers.

Dayal, Veneeta. 2000. Scope Marking: Cross-linguistic Variation in Indirect Dependency. In Uli Lutz, Gereon Müller and Arnim von Stechow (eds.), *Wh-Scope Marking*, pages 157–193, Amsterdam: John Benjamins, volume 37 of *Linguistics Today*.

Dayal, Veneeta. 2017. Does Hindi-Urdu have feature driven wh movement to Spec vP? *Linguistic Inquiry* 48(1), 159–172.

Féry, Caroline, Pandey, Pramod and Kentner, Gerrit. 2016. The prosody of focus and givenness in Hindi and Indian English. *Studies in Language* 40(2), 302–339.

Gambhir, Vijay. 1981. *Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi*. Ph. D. thesis, University of Pennsylvania, Philadelphia.

Glassman, Eugene H. 1977. *Spoken Urdu*. Lahore: Nirali Kitaben.

Halvorsen, Per-Kristian. 1983. Semantics for Lexical-Functional Grammar. *Linguistic Inquiry* 14, 567–615.

Han, Chung-Hye and Romero, Maribel. 2004. The Syntax of WHETHER/Q ... OR Questions: Ellipsis combined with movement. *Natural Language and Linguistic Theory* 22, 527–564.

Kidwai, Ayesha. 2000. *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. Oxford: Oxford University Press.

King, Tracy Holloway. 1997. Focus Domains and Information Structure. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'97 Conference*, Stanford, CA: CSLI Publications.

Mahajan, Anoop. 1997. Rightward Scrambling. In Dorothee Beerman, David LeBlanc and Henk van Riemsdijk (eds.), *Righward Movement*, pages 185–213, Amsterdam: John Benjamins.

Malik, Muhammad.K., Ahmed, Tafseer, Sulger, Sebastian, Bögel, Tina, Gulzar, Atif, Butt, Miriam and Hussain, Sarmad. 2010. Transliterating Urdu for a Broad-Coverage Urdu/Hindi LFG Grammar. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC2010)*, Malta.

Manetta, Emily. 2012. Reconsidering Rightward Scrambling: Postverbal Constituents in Hindi-Urdu. *Linguistic Inquiry* 43(1), 43–74.

Manetta, Emily and Gribanova, Vera. 2016. Ellipsis in wh-in-situ languages: deriving apparent sluicing in Hindi-Urdu and Uzbek. *Linguistic Inquiry* 47(4), 631–668.

Masica, Colin. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

Montaut, Annie. 2004. *Hindi Grammar*. München: Lincom-Europa.

Platts, John T. 1884. *A dictionary of Urdu, classical Hindi, and English*. W. H. Allen and Co.

Rooth, Mats. 1985. *Association With Focus*. Ph. D. thesis, University of Massachusetts, Amherst.

Rooth, Mats. 2016. Alternative semantics. In Caroline Féry and Shinichiro Ishihara (eds.), *Oxford Handbook of Information Structure*, Oxford: Oxford University Press.

Silverman, Kim, Beckman, Mary, Pitrelli, John, Ostendorf, Mari, Wightman, Colin, Price, Patti, Pierrehumbert, Janet and Hirschberg, Julia. 1992. ToBI: A standard for labeling English prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, Banff.

Slade, Benjamin. 2011. *Formal and Philological Inquiries into the Nature of Interrogatives, Indefinites, Disjunction, and Focus in Sinhala and other Languages*. Ph. D. thesis, University of Illinois at Urbana-Champaign.

Syed, Saurov and Dash, Bhamati. 2017. A unified account of the yes/no particle in Hindi, Bangla, and Odia. In Michael Yoshitaka Erlewine (ed.), *Proceedings of Generative Linguistics in the Old World (GLOW) in Asia XI (volumes 1 and 2)*, MIT Working Papers in Linguistics, The MIT Press.

Urooj, Saba, Mumtaz, Benazir and Hussain, Sarmad. 2019. Urdu intonation. *Journal of South Asian Linguistics* 10, 3–22.