DISCOURSE FUNCTIONS OF QUESTION WORDS

Louise Mycock
University of Oxford

Proceedings of the LFG13 Conference

Miriam Butt and Tracy Holloway King
(Editors)

2013

CSLI Publications

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

Abstract

Butt (2012) presents an analysis of echo questions in Urdu/Hindi which makes crucial reference to the information structure status of the 'wh' question word involved. Such a question word occupies the post-verbal position reserved for Background Information in Urdu/Hindi, which is characterized as [−NEW, −PROM(INENT)] according to the information feature system proposed by Butt & King (1996). This contrasts with the information feature values most often associated with a question word, namely those identified with Focus: [+NEW, +PROM]. One consequence of Butt's (2012) analysis is that it opens up an intriguing possibility: if a question word can belong to either the information structure category Focus or the information structure category Background Information, can it also belong to the other two major categories in Butt & King's (1996) system, i.e. Topic and Completive Information? And if so, how is this to be captured? I argue that question words can indeed belong to these two information structure categories. I propose to capture the relevant generalizations by having question words fully populate the information feature space, a proposal which relies on introducing an interface feature **Q** that is potentially relevant at multiple levels of the grammar, in line with Dalrymple & Mycock's (2011) approach to interface phenomena within the LFG framework, subsequently revised in Mycock & Lowe (2013). An initial review of a small sample of cross-linguistic data reveals support for a proposed distinction between [+PROM] and [−PROM] question words. By integrating question words fully into Butt & King's (1996) analysis of discourse functions and adding **Q** to the inventory of semantic/pragmatic information to be included in interface structures, this work further increases our understanding of interface phenomena and their analysis within the LFG framework.

1   Introduction[1]

In order to account for facts about word order in Urdu and Turkish, specifically the relations between particular syntactic positions and the information structure status of the constituents which may occupy them, Butt & King (1996) define four discourse functions based on binary values for two features, [±NEW] and [±PROM(INENT)], as shown in Figure 1. Butt & King's (1996) approach is based on Choi's (1996) analysis of information

structure in LFG, which itself is an extension of Vallduví's (1992) trinomial partition of a sentence on the basis of its information structure.

	[+NEW]	[−NEW]
[+PROM]	FOCUS	TOPIC
[−PROM]	COMPLETIVE INFORMATION	BACKGROUND INFORMATION

Figure 1. Information features and discourse functions (Butt & King 1996)

I take the information feature [±NEW] used in Butt & King's (1996) system to refer to relational rather than referential givenness–newness because, as Lambrecht (1994: 489) points out, "the conveying of [new] information is in principle independent of the previous mention or non-mention of the designata of the different constituents in a sentence". For instance, given an appropriate context a sentence consisting solely of anaphoric expressions such as *She did it* serves to establish relations between the various elements of the proposition, thus conveying new information and changing the addressee's representation of the world. This relational definition of the information feature [±NEW] is key to accounting for the discourse functions which question words may have. As for [±PROM], Butt & King (1996) define this feature in terms of whether the information is "of primary importance to the information structure of the discourse at hand" or not. Butt & King (1996) show that the four discourse functions defined in terms of these two binary-valued information features are associated with specific c-structure positions in Urdu and Turkish.

While Butt & King's influential work on discourse functions has been referred to in studies of constituent ('wh') questions within the LFG framework, previous analyses have not explored the possibility that question words can fully populate the information feature space defined by Butt & King (1996). Mycock (2006) proposed that all non-echo question words inherently have Focus status, thus [+NEW, +PROM] in Butt & King's system; cf. Horvath's (1986: 118) Focus Constraint. Gazdik (2011) meanwhile analyzed question words as having the feature value [+PROM]. Both approaches are inconsistent with Butt's (2012) analysis of Urdu/Hindi echo questions as having Background Information status, i.e. information feature values [−NEW, −PROM].

In this paper, I seek to explore the possibility that question words can have the same information feature values, and hence the same discourse functions, as non-interrogatives. In order to account for discourse functions of question words, I propose to introduce the feature **Q**, which is an interface feature similar to those first proposed in Dalrymple & Mycock (2011) and

extended to discourse functions such as Focus in Mycock & Lowe (2013). Question words will fully populate the information feature space as a result of the incorporation of the feature **Q**.

In §2, I present a summary of the claims presented in Butt (2012) relating to the discourse functions of question words in Urdu/Hindi as signalled by the syntactic positions that they may occupy. I outline my proposal concerning the introduction of the feature **Q** (§3), before defining each of the discourse functions included in the expanded information feature space and exemplifying them using English data (§4). The results of a preliminary survey of the syntax of constituent questions cross-linguistically are presented and discussed in §5, and a possible correlation between relative prominence and how prominence is encoded is considered. Finally, in §6, I provide the conclusion and outline some wider issues which remain to be addressed.

2   Constituent questions in Urdu/Hindi (Butt 2012)

Butt (2012), drawing on previous work including Manetta (2012), presents data from Urdu/Hindi, which show that question words occupy different positions depending on the type of question asked. Butt (2012) links these differences to the distinct discourse functions of the question word.

The default word order in Urdu/Hindi is SOV, but certain syntactic positions in this language are associated with specific discourse functions and therefore it is best understood as being to some extent a discourse-configurational language; see Butt & King (1996) for details. Key to Butt's (2012) analysis of question formation in Urdu/Hindi are the immediately preverbal Focus position and the immediately post-verbal, clause-final position associated with Background Information. In a regular Urdu/Hindi question like (1), the question phrase by default occupies the immediately pre-verbal Focus position.

(1)   sita=ne   d^{h}yan=se   **[kɪs=ko]**$_{Focus}$ dek^h-a
        Sita.F=ERG   carefully   who.OBL=ACC   see-PERF.M.SG
        t^h-a?
        be.PAST-M.SG
        'Whom had Sita looked at carefully?'   (Butt 2012)

In an echo question, by contrast, the question phrase occupies the position reserved for Background Information immediately following the verbal complex.

(2)   sita=ne   d^{h}yan=se   dek^h-a   t^h-a
            Sita.F=ERG   carefully   see-PERF.M.SG   be.PAST-M.SG
            [kɪs=ko]Background Information?
            who.OBL=ACC
            'Sita had looked carefully at who?'   (Butt 2012)

Butt (2012) states that this is not surprising, but rather reflects how information structure interacts with question semantics to give different pragmatic interpretations: a constituent question containing a post-verbal question word "cannot be interpreted as a typical information-seeking question, but receives an echo question interpretation", implying that this is as a result of the question word's status as Background Information. This is particularly clearly illustrated by examples from Urdu/Hindi given the transparent relationship between syntactic position and the relevant discourse function (Background Information) in this language.

3   The feature Q

The relevant generalizations as reported in Butt (2012) concerning question words in a typical information-seeking question and an echo question in Urdu/Hindi could be incorporated into the information feature space given in Figure 1 by simply permitting the possibility that question words can have Focus or Background Information status just as non-interrogatives can. However, this cannot be the whole story because it is not the case that interrogatives and non-interrogatives will have identical properties when they apparently have the same information structure status as proposed above. For instance, in English a question word with the discourse function Focus occupies SpecCP (appears 'ex situ'), as illustrated by A's utterance in (3), but there is no such restriction on non-interrogative Focus, as the location of the 'answer constituent' *flowers*, which has the discourse function Focus in B's response in (3), demonstrates.

(3)   A: <u>What</u> did Lily buy at the market?
          B: She bought <u>flowers</u>.

I therefore propose to augment Butt & King's (1996) system by introducing a feature **Q**. **Q** will be part of any question word's lexical entry and will be included in the 'interface structures' associated with the relevant string units, as first proposed in Dalrymple & Mycock (2011), and revised and extended in Mycock & Lowe (2013). With the incorporation of **Q**, the feature space becomes more complex, as shown in Figure 2. I argue that this is justified and, indeed, necessary if we are to account for the discourse functions of question words whose properties are not identical to their non-interrogative counterparts.

	[+NEW]	[−NEW]
[+PROM]	FOCUS	TOPIC
	Q: QUESTIONING FOCUS	**Q**: SORTING KEY
[−PROM]	COMPLETIVE INFORMATION	BACKGROUND INFORMATION
	Q: NON-SORTING KEY	**Q**: ECHO QUESTION

Figure 2. Expanded version of Butt & King's (1996) information feature space including the interface feature **Q**

 Q's status as an interface feature is justified because it represents information that may be relevant at multiple levels of the grammar, for instance for the purposes of clause typing or for determining which intonational pattern can appropriately be used with a clause. Note that **Q** is not a feature which is introduced only by particular lexical items. It could also be associated with a specific intonational contour, as in *She BOUGHT flowers?* where capitals indicate that the verb bears the sentence's main stress and the result is questioning of the event type. This would be another way to introduce **Q** as a feature in the interface structures associated with string units.[2] For an LFG analysis of declarative questions, see Dalrymple & Mycock (2011); for an LFG approach to the prosodic encoding of interface features, see Mycock & Lowe (2013).

4 Discourse functions of question words

In this section I define and exemplify using English data the discourse functions Questioning Focus (§4.1), Echo Question (§4.2), Sorting Key (§4.3) and Non-Sorting Key (§4.4), and formulate a hypothesis relating to the syntactic encoding of relative prominence, before going on to consider cross-linguistic data and its implications in relation to the information feature [±PROM] in §5. Differences in the information structure of questions containing single versus multiple question words lead me to deal with them separately. Constituent questions containing a single question word are

---

[2] Falk (2012) proposes a function Q, but under his analysis Q is an i-structure attribute. He defines Q as "the function of question words" and proposes that a question word bears both Q, which he suggests is classified as [+PROM], and another discourse function. The justification for treating Q as a discourse function/i-structure attribute comparable to Topic or Focus, however, remains unclear.

covered in §4.1; constituent questions containing multiple question words are discussed in §4.3 and §4.4.

4.1 (New Information) Focus and Questioning Focus

With respect to question–answer pairs, the information structure category Focus, as exemplified in (4), has received the most attention in the literature. This kind of exchange involves a constituent question whose form and function are regularly treated as being prototypical: the sentence is a request for the addressee to supply information to fill a gap in the questioner's knowledge; the information gap is denoted by a clause-initial question word, either on its own or in a phrase.

(4) [+NEW, +PROM]
    A: <u>What</u> did Lily buy at the market?
         Q: QUESTIONING FOCUS
    B: She bought <u>flowers</u> at the market.
         NEW INFORMATION FOCUS

The classification of the underlined constituents in (4) as examples of Focus is widely accepted and their [+NEW, +PROM] feature specification is straightforward given the definitions of the two information features presented in §1. In both cases, the information is of primary importance, i.e. [+PROM]: in A's sentence the question word indicates the gap in the speaker's knowledge that s/he seeks to be filled, while in his/her response B provides the information that has been requested. Both the question word *what* and the answer constituent *flowers* in (4) are [+NEW] in the relational sense because the information they express changes the respective addressee's representation of the world (in the case of the question word, it expresses a gap in the speaker's knowledge) and establishes new pragmatic relations.

    As discussed in §3, in English the difference between interrogative and non-interrogative foci is reflected in their syntax: *what* appears clause initially in the 'ex situ' SpecCP position in (4), while the neutral position of its non-interrogative counterpart (*flowers*) is 'in situ'.[3] Syntactic differences between the two types of foci are found in other languages too. For example, in a neutral constituent question in Hungarian, multiple question words/phrases (interrogative foci) appear together immediately before the verb, whereas only a single non-interrogative Focus constituent may appear in immediately preverbal position. I follow Dik (1997), who groups these two

---

[3] I adopt the terms 'in situ' and 'ex situ' for the sake of convenience. Within a non-derivational framework like LFG, there is of course no assumption that movement is involved.

types of Focus together as a major sub-type of Focus, viz. Information Gap Focus, but also distinguishes between the two using the terms Questioning Focus and New Information Focus respectively for the types of interrogative and non-interrogative Focus illustrated in (4).[4]

4.2 Background Information and Echo Question

I follow Butt (2012) in analysing a question word in an echo question as having the same information feature specification as Background Information, i.e. [−NEW, −PROM]. Expressions that have this information structure status do not change the addressee's representation of the world nor establish new pragmatic relations; they do not encode information of primary importance. For example, in A's initial utterance in (5), flowers were mentioned as having been bought by Lily, thus establishing the relation between the flowers – referred to in B's sentence as *them* − and the other elements of the proposition. Under the relational definition of newness therefore, *them* in (5) is classified as [−NEW]. In this particular context, *them* (referring to the flowers) does not represent information of primary importance, i.e. [−PROM]; cf. *her mother* (Focus) and *she*, i.e. Lily, (Topic; see §4.3). This means that *them* in (5) is classified according to Butt & King's (1996) system as Background Information.

(5)   [−NEW, −PROM]
      A: Who did Lily buy flowers for?
      B: She bought <u>them</u> for her mother.
          BACKGROUND INFORMATION

Butt (2012) proposes that a question word in an echo question similarly has the information feature classification [−NEW, −PROM], i.e. the opposite of Questioning Focus (§4.1). With respect to [±NEW], in an echo question such as (6) it is not the relational status of the questioned element that is at issue because this is, after all, expressed in the preceding utterance.[5]

---

[4] Dik's (1997) other main sub-type of Focus is Contrast. I will have nothing to say about contrast in this article. While Choi's (1996, 1999) original characterization incorporates contrast, it is not part of Butt & King's (1996) version of the information feature space.

[5] I confine myself in this paper to discussion of echo questions that request repetition of information already provided, as illustrated by (6). Such questions are also referred in the literature to as a type of reprise question, e.g. by Ginzburg & Sag (2000) following Bolinger (1978). I do not discuss another type of question with which such echo questions are often grouped, namely one used to express shock or disbelief that includes a prosodically prominent in-situ question word in English. Such questions, while similar in terms of their secondary nature and the in-situ position of a question word, are functionally distinct – they do not express a true gap

(6)   [−NEW, −PROM]
        A: Lily bought flowers yesterday.
        B: She bought <u>WHAT</u> yesterday?
             Q: ECHO QUESTION
        A: Flowers.

Utterance of an echo question does not establish new pragmatic relations; it is simply a request for repetition of information that the addressee did not catch the first time – s/he is asking for reiteration of an unheard linguistic expression. This supports Butt's [−NEW] classification, which reflects the 'secondary' nature of an echo question. An echo question refers back to an immediately preceding utterance and cannot be uttered 'out of the blue', in contrast to a constituent question containing a question word whose discourse function is Questioning Focus, as in (4), which can.

While an echo question signals a gap in speaker knowledge, the question word is not classified as information of primary importance according to Butt (2012), so it is [−PROM]. It seems reasonable to posit that this is also related to the secondary nature of this type of question. Furthermore, a classification of [−NEW, +PROM] would not be appropriate: the echo question word is not a Topic, nor does it determine what will be the Topic(s) in the addressee's response; cf. the Sorting Key question word in a multiple constituent question discussed in §4.3.

Finally, it is interesting to compare the syntax of the echo question in (6) with its non-echo counterpart in (4). The question word *what* appears in clause-initial position (SpecCP) when it has the discourse function Questioning Focus, but it appears in the same position as a non-interrogative equivalent when it has the discourse function Echo Question. In terms of the syntactic position in English of these question words with distinct discourse functions therefore, we see a contrast between a question word classified as [−PROM] (in situ) and a question word classified as [+PROM] (ex situ). This suggests a correlation between relative prominence, i.e. the value of the information feature [±PROM], and long-distance dependencies, an issue which will be considered further in §5.

4.3   Topic and Sorting Key

A Topic shares one information feature specification with Background Information and one with Focus, being classified as [−NEW, +PROM]. For example, the Topic *she* in (7) does not represent new information: the relation between Lily and the rest of the proposition was established in A's

---

in the speaker's knowledge. I leave the issue of the relationship between such questions, the echo questions I discuss here, and other types of constituent question for future work.

sentence. As a Topic though, *she* is prominent in informational structural terms. A Topic is crucial in determining the basic division of information into Topic and Comment in a categorical utterance, a distinction which may be encoded in syntactic structure (see, e.g., Sasse 1987).

(7)   [−NEW, +PROM]
         A: What did Lily do?
         B: <u>She</u> bought flowers at the market.
              TOPIC

    The more Topic-like status of one question phrase in a multiple constituent question such as the one in (8) has long been observed, see e.g. Bolinger (1978), Erteschik-Shir (1986) and Kuno & Takami (1993). Bolinger (1978) maintains that the leftmost question word in a multiple constituent question can be seen as belonging to the Topic part of the Topic–Comment organization of the sentence. He claims that the clause-initial position of a question word "topicalizes it as a first assumption" (Bolinger 1978: 133). This question word is effectively adopted as a Topic and any other question words are predicated about it. I follow Bolinger (1978) in assuming that the discourse functions of question words in multiple constituent questions are not identical. On this basis, I assume that the information structure of non-echo single and multiple constituent questions differ in certain fundamental respects, meaning that an analysis of the feature specification of all non-echo question words as [+NEW, +PROM] obscures crucial distinctions relating to discourse functions.

    Kuno & Takami (1993: 112) refer to the Topic-like question word in a multiple constituent question as the Sorting Key. It is this question word which communicates how the questioner expects information to be organized in the addressee's response, and therefore is crucial in determining the (in)felicity of an answer, as the relative felicity of (9b) and (9b') in response to (9a) shows. I use Kuno & Takami's (1993) term Sorting Key to refer to the discourse function of such a question word. A Sorting Key shares the feature specification [+PROM] with Questioning Focus, the discourse function of the only question word in the type of neutral constituent question discussed in §4.1.

(8)   [−NEW, +PROM]
         A: <u>Who</u> bought what?
              Q: SORTING KEY
         B: Lily bought flowers, Charlie bought cakes and Fiona bought balloons.

(9)   a.   <u>Which of these climatic conditions</u> occurs in which countries?
       b.   <u>Typhoons</u> occur in Japan, Korea and China; <u>hurricanes</u> occur in
             …
       b'.   #<u>In Japan</u>, typhoons and early summer rain spells occur; <u>in
             Thailand</u>, they have monsoons and tornadoes; …

<div align="right">(Kuno 1982: 144)</div>

I refer to the Sorting Key as the more Topic-like question word advisedly. I do not seek, as others may have done, to propose that this question word *is* a Topic; there is good evidence that this is not the case. For example, in Hungarian question words do not appear in the Topic field (Puskás 2000). In this way a Sorting Key is not the same as its non-interrogative counterpart, with which it shares the feature specification [−NEW, +PROM]. Such differences provide further justification for distinguishing Sorting Key from Topic, and thus for introducing the interface feature **Q** as a way of capturing the relevant distinction.

Sorting Key as a discourse function distinct from Topic also has a bearing on the more general issue of the value of information features and the discourse functions which they have been used to define. Butt & King (1996), in common with Choi (1996, 1999), propose binary values for the information features [±NEW] and [±PROM]. However, the differences between Topic and Sorting Key, which I hypothesise are due to the Sorting Key being the most Topic-like of all the question words in a multiple constituent question without actually being a Topic, indicate that binary conceptions of information features alone are too simplistic when it comes to accounting for the discourse functions of constituents within a sentence. This is because the information structure status of any constituent depends on the information structure status of other elements in the same sentence. Rather than binary values then, the relevant information features may be best understood in terms of a spectrum. Dalrymple & Nikolaeva (2011: 66) in their work on differential object marking make the same point regarding the feature [±PROM] in relation to primary and secondary Topics. A more refined analysis of discourse functions in general awaits an approach to information features which will enable relative levels of prominence, newness, etc. to be modelled.

4.4   Completive Information and Non-Sorting Key

The fourth and final discourse function defined in Butt & King (1996) is Completive Information, which is classified as [+NEW, −PROM]. To give an example, in (10) *flowers* were not mentioned in the preceding question so the information that Lily has just bought these items at the market is new to the addressee A, but that information is not of primary importance in this

context; it simply represents unsolicited additional information which gives A a fuller picture of the event under discussion.

(10) [+NEW, −PROM]
      A: Where has Lily been shopping?
      B: She's just bought <u>flowers</u> at the market.
          COMPLETIVE INFORMATION

I propose that the interrogative constituent with the information feature specification [+NEW, −PROM] is any question word in a multiple constituent question which does not have Sorting Key status, for instance *what* in (11). I use the term Non-Sorting Key to refer to the discourse function of such question words. This term and its definition reflect the dependent relationship, in terms of information structure status, which Bolinger (1978) and Kuno & Takami (1993), inter alia, assume exists between question words in a multiple constituent question.[6] Note that Completive Information, by its very nature, displays a directly comparable lack of independence in terms of information structure status.

(11) [+NEW, −PROM]
      A: Who bought <u>what</u>?
          Q: NON-SORTING KEY
      B: Lily bought flowers, Charlie bought cakes and Fiona bought
          balloons.

A Non-Sorting Key is [+NEW], a feature shared with Questioning Focus, because it establishes new pragmatic relations, thus changing the addressee's representation of the world. It is [−PROM] because it represents information that is *relatively* less important than other information – most notably the Sorting Key – in the given context.

Once more, the issue of relative values for an information feature like [±PROM] has arisen. It is interesting in this respect to consider multiple constituent question formation possibilities in a language with word order that is freer than that of English, at least in terms of the relations which exist between grammatical functions and syntactic positions. As the examples in (12) show, when there is more than one question phrase in a Hungarian constituent question, different orderings of the immediately preverbal question phrases are possible.

---

[6] The nature of the relationship between Sorting Key and Non-Sorting Key question words is also reflected to some extent in their diametrically opposed feature specifications, each of which shares one feature with that of Questioning Focus (see §4.1), i.e. the Sorting Key is [+PROM] while a Non-Sorting Key is [+NEW].

(12) a. Ki   ki-t   ki-nek   mutat-ott   be?
          who.NOM   who-ACC   who-DAT   introduce-PAST.3SG   VM
      b. Ki-t   ki-nek   ki   mutat-ott   be?
          who-ACC   who-DAT   who.NOM   introduce-PAST.3SG   VM
      c. Ki-nek   ki-t   ki   mutat-ott   be?
          who-DAT   who-ACC   who.NOM   introduce-PAST.3SG   VM
          'Who introduced who to who?

I contend that these different ordering possibilities in Hungarian will be best understood in relative terms with respect to the question words' information structure status. As with the notion of the Sorting Key discussed in §4.3, an approach to discourse functions which incorporates relative levels of prominence, newness, etc. is required before a full analysis of such phenomena can be provided. This goes beyond the scope of the work presented in this paper, but I wish to highlight this as a matter of significant importance for future research; see also §5.2.

Finally, an important comparison remains to be made in relation to the English data presented in §4.3 and §4.4. This involves the Sorting Key and Non-Sorting Key question words, which differ in terms of their value for the information feature [±PROM]. Note that a [+PROM] Sorting Key question word appears clause initially in English. With the possible exception of a question phrase which has the grammatical function subject (whose syntactic position has been the topic of debate in the literature and which I set aside here), the Sorting Key question phrase occupies SpecCP and thus appears ex situ. By contrast, a [−PROM] Non-Sorting Key question word appears in situ in English. The same correlation was observed with respect to Questioning Focus versus Echo Question in §4.2.

4.5   Hypothesis: syntactic encoding of relative prominence

When the syntax of question words with the discourse functions Questioning Focus and Echo Question (§4.2) and Sorting Key and Non-Sorting Key (§4.4) are compared, we see a correlation between relative prominence, i.e. the values for [±PROM], and syntactic marking or 'highlighting' of this prominence involving long-distance dependencies. Consequently, a more general point of comparison presents itself with respect to the encoding of question words' discourse functions, which in turn provides the basis for the formulation of a hypothesis. We have seen that question words that have either of the two types of discourse functions that are classified as [+PROM] (Questioning Focus and Sorting Key) appear ex situ and involve long-distance dependencies in English. By contrast, question words that have either of the two types of discourse functions that are classified as [−PROM] (Echo Question and Non-Sorting Key) appear in situ; they are not syntactically highlighted or marked in any way. It seems that in English

constituent questions, greater relative prominence is signalled syntactically by long-distance dependencies. Given this correlation, in §5 I seek to test a hypothesis concerning prominence and how it is encoded against data from a preliminary survey of cross-linguistic data.

(13) PRINCIPLE OF RELATIVE PROMINENCE ENCODING
     A [−PROM] question word will only be syntactically 'highlighted' in a language (i.e. appear ex situ, as the filler element in a long-distance dependency) if its [+PROM] question word counterpart is also by default syntactically highlighted.

5 [±PROM] and the syntax of constituent questions: a preliminary survey

The preliminary survey of cross-linguistic data presented in this section, against which the hypothesis given in §4.5 is tested, is necessarily small, comprising English and 14 other languages. The criteria for selecting a language for inclusion were that data were available on the syntax of all of the following: (i) constituent questions containing a single question word,[7] (ii) multiple constituent questions and (iii) echo questions.

The primary reason for the relatively small sample size is the fact that investigation of echo question formation cross-linguistically is made difficult because this type of question is often overlooked in descriptive work. An important exception, and the mainstay of the survey presented in this paper, is some of the research based on Comrie & Smith's (1977) questionnaire. These works, however, rarely describe the prosody of echo (or sometimes any) questions in detail. For this reason, it was unfortunately not possible to consider how prosody interacts with syntax in relation to the discourse functions of question words in this preliminary survey.[8]

With the exception of English, the data referred to in the rest of this section come from the 14 descriptive grammars listed in Table 1, which were

---

[7] For the purposes of this preliminary work, I confined the survey to languages that were reported in the relevant descriptive grammar as having one main strategy for forming constituent questions. This meant that Japanese (Hinds 1986), for instance, was included, whereas Maori (Bauer 1993) was not. In Japanese, the main constituent question formation strategy is 'wh in situ', but in Maori the strategy employed varies depending on the grammatical function of the question word. Mycock (2006: 382–383) provides a brief discussion concerning the analysis of Maori constituent question formation within LFG.

[8] On the prosodic marking of question words in a 'wh-in-situ' language, see Mycock (2006), who provides an LFG analysis of non-echo constituent questions in Japanese. Further research is required to determine if such an analysis is also appropriate for other wh-in-situ languages.

written by authors whose research was guided by Comrie & Smith's (1977) questionnaire.

Language	Classification (Lewis et al 2013)	Reference
Cairene Egyptian Colloquial Arabic	Afro-Asiatic, Semitic	Olmstead Gary & Gamal-Eldin (1982)
Catalan	Indo-European, Italic	Hualde (1992)
Gulf Arabic	Afro-Asiatic, Semitic	Holes (1990)
Hungarian	Uralic	Kenesei et al (1998)
Japanese	Japonic	Hinds (1986)
Kannada	Dravidian	Sridhar (1990)
Kashmiri	Indo-European, Indo-Iranian	Wail & Koul (1997)
Kobon	Trans-New Guinea, Madang	Davies (1981)
Korean	Language isolate	Sohn (1994)
Koromfe	Niger-Congo	Rennison (1997)
Modern Greek	Indo-European, Greek, Attic	Joseph & Philippaki-Warburton (1987)
Persian	Indo-European, Indo-Iranian	Mahootian (1997)
Punjabi	Indo-European, Indo-Iranian	Bhatia (1993)
Turkish	Altaic, Turkic	Kornfilt (1997)

Table 1. Details of the 14 languages included in the preliminary survey (along with English) and the descriptive grammars from which the data on constituent questions come

5.1   The syntax of Questioning Focus and Echo Question

Table 2 summarizes the findings for Questioning Focus and Echo Question in the 15-language sample. One key issue was considered: Did question words appear ex situ or in situ? These data reveal a correlation between relative prominence and syntactic 'highlighting' of a question word.

The sample shows that question words with either discourse function – Questioning Focus or Echo Question – may appear in situ or ex situ, although one particular combination was not found. We find languages in which question words with either of these discourse functions appear in situ (e.g. Japanese) or ex situ (e.g. Hungarian), i.e. regardless of their value for the information feature [±PROM]. There are also languages, like English, in which only [+PROM] question words, i.e. those with the discourse function Questioning Focus, can appear ex situ; in such languages a [−PROM] question

word (i.e. with Echo Question status) appears in situ. In these languages therefore, we find a syntactic correlate of the proposed difference in the relative information structure prominence of question words with these discourse functions. What is most striking though is the lack of a language in which a [−PROM] question word (Echo Question) appears ex situ while a [+PROM] question word (Questioning Focus) appears in situ. This is consistent with the Principle of Relative Prominence Encoding presented as (13) in §4.5.

QUESTIONING FOCUS [+NEW, **+PROM**]	ECHO QUESTION [−NEW, **−PROM**]	Languages
QW in situ	QW in situ	Cairene Egyptian Colloquial Arabic, Gulf Arabic, Japanese, Kannada, Kobon, Korean, Persian, Punjabi
QW ex situ	QW ex situ	Hungarian, Kashmiri, Turkish
QW ex situ	QW in situ	Catalan, English, Koromfe, Modern Greek
QW in situ	QW ex situ	

Table 2. A survey of the syntax of question words (QWs) with the discourse functions Questioning Focus and Echo Question in a sample of 15 languages

5.2 The syntax of Sorting Key and Non-Sorting Key

The remaining two discourse functions, Sorting Key and Non-Sorting Key, also differ in terms of their specified relative prominence, and are particularly interesting because, by definition, they co-occur in a multiple constituent question. Table 3 shows whether a question word with the relevant discourse function appears by default in situ or ex situ in the 15-language sample. The three groups of languages shown in Table 3 represent the three main types of question formation strategy that are regularly identified in the literature, viz. wh in situ (both Sorting Key and Non-Sorting Key question words appear in situ), 'multiple fronting' (both Sorting Key and Non-Sorting Key question words appear ex situ) and the English-type strategy (Sorting Key appears ex situ, any Non-Sorting Key appears in situ).[9]

Table 2 and Table 3 are for all intents and purposes identical: we see the same alignment between relative prominence and syntactic encoding, consistent with the Principle of Relative Prominence Encoding that was

---

[9] For an LFG analysis of the typology of constituent questions, see Mycock (2006).

presented as (13) in §4.5. The syntax does not differ according to the discourse function of the question word in the case of the first two groups of languages (both in situ, both ex situ), while in the third group a question word is only ex situ – and a long-distance dependency is involved – when its information feature specification includes [+PROM] (i.e. its discourse function is Sorting Key). Directly comparable with the findings summarized in Table 2 that relate to the syntax of Questioning Focus and Echo Question in the sample, no language in this survey had a [−PROM] Non-Sorting Key question word that appeared ex situ with a [+PROM] Sorting Key question word in situ. Once again, we see a correlation between relative prominence and how it is encoded.

SORTING KEY [−NEW, **+PROM**]	NON-SORTING KEY [+NEW, **−PROM**]	Languages
QW in situ	QW in situ	Cairene Egyptian Colloquial Arabic, Gulf Arabic, Japanese, Kannada, Kobon, Korean, Persian[10], Punjabi
QW ex situ	QW ex situ	Hungarian, Kashmiri, Turkish
QW ex situ	QW in situ	Catalan, English, Koromfe, Modern Greek
QW in situ	QW ex situ	

Table 3. A survey of the syntax of question words (QWs) with the discourse functions Sorting Key and Non-Sorting Key in a sample of 15 languages

It is interesting to see that even in ex-situ languages, there may be a distinction in the syntax corresponding to the distinction in information structure status between a Sorting Key and Non-Sorting Key question words. For instance, Rudin (1988) shows that in certain ex-situ Slavic languages, the first question word can be separated from the rest (see also Mycock 2006: 376–378). It must also be remembered that when multiple ex-situ question

---

[10] The neutral constituent question formation strategy in Persian presented in Mahootian (1997) is wh in situ. Other sources present a slightly different picture. According to Raghibdoust (1994) and Lotfi (2003), for instance, all three constituent question formation strategies – wh in situ, multiple fronting and the English type – are possible in Persian. However, it should be noted that questions formed according to the three different strategies "differ from each other with respect to the degree of acceptability" (Raghibdoust 1994: 36), hinting that context may play a role which remains to be fully explored. For this reason, Mahootian's (1997) Persian data have been included in this preliminary survey.

words appear to target the same syntactic position, it does not follow that they must all have the same discourse function. Recall the Hungarian examples in (12), in which all question words appear immediately before the verb, in what is often referred to in the literature as 'Focus position'.[11] Clearly, some principle(s) must determine the order in which the question words appear. I would argue that key to understanding such data are the relative values of the relevant information features for each question word (see §4.3), possibly amongst other factors. Simply designating every constituent in preverbal position as, by definition, having the discourse function 'Focus' does not capture the complexities involved. In this respect, Simpson's (2007) discussion and analysis of the pragmatic constraints on word order in Warlpiri offer a potentially illuminating perspective. Simpson (2007: 421) argues that in Warlpiri "prominence and newness apply to spans within a clause, rather than to a single phrase structure position". A similar approach may be appropriate in the analysis of multiple preverbal question words and their discourse functions in Hungarian.

5.3   Findings and future research

Taken together, the findings of the preliminary survey reported in §5.1 and §5.2 support the hypothesis that a [−PROM] question word will only be syntactically 'highlighted' in a language (i.e. appear ex situ, as the filler element in a long-distance dependency) if its [+PROM] question word counterpart is also by default syntactically highlighted. Of course, further research is required to determine whether the empty cell in Table 2 or Table 3 can be filled, and whether any in-situ/ex-situ difference identified relates to the relative prominence of the question words concerned or to some other factor or factors.

6   Conclusion and related issues

I have argued that question words can have the same values as non-interrogatives for the information features [±NEW] and [±PROM] proposed by Butt & King (1996). However, data show that it is still important to make a distinction between interrogative and non-interrogative constituents. In order to account for the distinct properties of question words as well as the information structural features that they share with their non-interrogative counterparts, I proposed to augment Butt & King's (1996) system by introducing the interface feature **Q**, which will be part of a question word's lexical entry. An analysis incorporating **Q** means that question words can

---

[11] Forst et al (2010) and Mycock (2006), working within the LFG framework, analyse the relevant position in Hungarian as being SpecVP.

fully populate the feature space but are not conflated with non-interrogatives that have identical values for the two relevant information features. This approach better covers the relevant data than two previous LFG treatments of the discourse functions of question words (Mycock 2006, Gazdik 2011), both of which restricted the discourse function that question words can bear in some way, and both of which were, as a result, inconsistent with Butt's (2012) [−NEW, −PROM] analysis of question words in Urdu/Hindi echo questions. A preliminary survey of data from a small sample of 15 languages supports the hypothesis that there is a correlation between relative prominence and the means by which prominence may be encoded. Both [+PROM] and [−PROM] question words may appear ex situ and be involved in a long-distance dependency, but the latter can only be syntactically highlighted in this way if the former are too, i.e. such syntactic highlighting is not used to mark [−PROM] alone.

The matter of whether long-distance dependencies should be regarded as somehow greater in terms of the degree of highlighting that they involve compared to non-syntactic marking of relative values for, e.g., prominence, warrants further consideration. In English, a [+PROM] question word appears ex situ in clause-initial position; by contrast, a [−PROM] question word appears in situ but is prosodically prominent to some degree. It appears that in some sense syntactic highlighting is itself 'more prominent' than other non-syntactic alternatives. This seems a reasonable hypothesis given that long-distance dependencies are likely to be more costly in processing terms.

The analysis of discourse functions of question words presented in this paper represents a starting point. Its inadequacies, particularly with respect to capturing a category like Sorting Key as outlined in §4.3, highlight an important issue, also raised by others such as Dalrymple & Nikolaeva (2011: 66) and Simpson (2007), in relation to the definition of discourse functions in terms of binary-valued information features such as [±NEW] and [±PROM]. It is clear that such an approach to information structure will ultimately prove to be inadequate. It is too simplistic to classify elements as being either new or not, prominent or not. An analysis purely in terms of binary features obscures the fine-grained distinctions that need to be made between elements *relative to one another* with regard to their information structural status. Progress in this area will mean that future treatments of discourse functions are able to capture more accurately the information structure of sentences, including constituent questions.

References

Bauer, Winifred with William Parker & Te Kareongawai Evans (1993). *Maori*. London: Routledge.
Bhatia, Tek K. (1993). *Punjabi*. Routledge: London.
Bolinger, Dwight (1978). 'Asking More Than One Thing at a Time'. In

Henry Hiz (ed.), *Questions*. Dordrecht: Reidel. 107–150.

Butt, Miriam (2012). *Questions in Urdu/Hindi: Moving Beyond Movement*. Presented at the LFG2012 Conference.

Butt, Miriam & Tracy Holloway King (1996). 'Structural Topic and Focus without Movement'. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG96 Conference*. Stanford, CA: CSLI Publications.

Choi, Hye-Won (1999). *Optimizing Structure in Context: scrambling and information structure*. Stanford, CA: CSLI Publications.

Choi, Hye-Won (1996). *Scrambling: Optimality-theoretic Interaction between Syntax and Discourse*. Ph.D. thesis, Stanford University.

Comrie, Bernard & Norval Smith (1977). 'Lingua Descriptive Studies: Questionnaire'. *Lingua* 42: 1–72.

Dalrymple, Mary & Louise Mycock (2011). 'The Prosody-Semantics Interface'. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*. Stanford, CA: CSLI Publications.

Dalrymple, Mary & Irina Nikolaeva (2011). *Objects and Information Structure*. Cambridge, UK: Cambridge University Press.

Davies, John (1981). *Kobon*. North-Holland: Amsterdam.

Dik, Simon C. (1997). *The Theory of Functional Grammar. Part 1* (ed. Kees Hengeveld). Berlin: Mouton de Gruyter.

Erteschik-Shir, Nomi (1986). 'Wh-questions and focus'. *Linguistics and Philosophy* 9: 117–149.

Falk, Yehuda (2012). *Superiority Effects*. ms, The Hebrew University of Jerusalem.

Forst, Martin, Tracy Holloway King & Tibor Laczkó (2010). 'Particle verbs in computational LFGs: issues from English, German and Hungarian'. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG10 Conference*. Stanford, CA: CSLI Publications.

Gazdik, Anna (2011). *Multiple Questions in French and Hungarian*. Ph.D. thesis, Université Paris Diderot/ELTE.

Ginzburg, Jonathan & Ivan A. Sag (2000). *Interrogative Investigations: the form, meaning, and use of English interrogatives*. Stanford, CA: CSLI Publications.

Hinds, John (1986). *Japanese*. Croom-Helm: London.

Holes, Clive (1990). *Gulf Arabic*. Routledge: London.

Horvath, Julia (1986). *FOCUS in the Theory of Grammar and the Syntax of Hungarian*. Dordrecht: Foris.

Hualde, José Ignacio (1992). *Catalan*. Routledge: London.

Joseph, Brian D. & Irene Philippaki-Warburton (1987). *Modern Greek*. Croom-Helm: London.

Kenesei, István, Robert M. Vago & Anna Fenyvesi (1998). *Hungarian*. Routledge: London.

Kornfilt, Jaklin (1997). *Turkish*. Routledge: London.

Kuno, Susumu (1982). 'The Focus of the Question and the Focus of the Answer'. *Papers from the Parasession on Nondeclaratives*. Chicago: Chicago Linguistics Society. 134–157.

Kuno, Susumu & Ken-Ichi Takami (1993). *Grammar and Discourse Principles*. Chicago: University of Chicago Press.

Lambrecht, Knud (1997). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press.

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.) (2013). *Ethnologue: Languages of the World, 17th edition.* Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Lotfi, Ahmad R. (2003). 'Persian Wh-Riddles'. In Cedric Boeckx & Kleanthes K. Grohmann (eds.), *Multiple Wh-Fronting*. Amsterdam: John Benjamins. 161–186.

Mahootian, Shahrzad (1997). *Persian*. Routledge: London.

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

Mycock, Louise (2006). *The typology of constituent questions: a Lexical-Functional Grammar analysis of 'wh'-questions*. Ph.D. thesis, University of Manchester.

Mycock, Louise & John L. Lowe (2013). 'The Prosodic Marking of Discourse Functions'. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*. Stanford, CA: CSLI Publications.

Olmstead Gary, Judith & Saad Gamal-Eldin (1982). *Cairene Egyptian Colloquial Arabic*. North-Holland: Amsterdam.

Puskás, Genoveva (2000). *Word Order in Hungarian: The Syntax of Ā-Positions*. Amsterdam: John Benjamins.

Raghibdoust, Shahla (1994). 'Multiple *Wh*-Fronting in Persian'. *Cahiers linguistiques d'Ottawa* 21: 27–58.

Rennison, John R. (1997). *Koromfe*. Routledge: London.

Rudin, Catherine (1988). 'On Multiple Questions and Multiple Fronting'. *Natural Language and Linguistic Theory* 6: 445–501.

Sasse, Hans-Jürgen (1987). 'The thetic/categorical distinction revisited'. *Linguistics* 25: 511–580.

Simpson, Jane (2007). 'Expressing Pragmatic Constraints on Word Order in Warlpiri'. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Chris Manning (eds.), *Architecture, Rules, and Preferences: Variations on Themes by Joan Bresnan*. Stanford, CA: CSLI Publications. 403–427.

Sohn, Ho-Min (1994). *Korean*. Routledge: London.

Sridhar, Shikaripur N. (1990). *Kannada*. Routledge: London.

Vallduví, Enric (1992). *The Informational Component.* London: Garland.

Wali, Kashi & Omkar N. Koul (1997). *Kashmiri*. Routledge: London.