FUNCTIONAL CONSTRAINTS ON WH-QUESTION FORMATION[1]

Sergio de Moura Menuzzi (UFRGS \ CNPq)

Gabriel de Ávila Othero (PUCRS \ CNPq)

smenuzzi@yahoo.com.br

gabriel_othero@terra.com.br

ABSTRACT: In this article, we seek to demonstrate the descriptive and explanatory power of two central ideas in Optimality Theory: (i) the idea that there can be conflict among grammatical constraints; and (ii) that these conflicts are solved by the ranking of the constraints. We illustrate these points with an analysis of the sentence patterns in question-formation, as shown in Ackema & Neeleman (1998). However, our analysis is different from A&N (1998), as we propose functional constraints acting in the formation of interrogative sentence patterns. We actually propose three constraints: Economy, Morphological Marking and Focalization. As we understand, a grammar is a system of optimization in two senses: (i) its conditions are codifications of functional demands over the form of the sentences; and (ii) the interaction among these demands must include a way to solve eventual conflicts among them.
KEYWORDS: Optimality Theory; Syntax; Wh-questions.

1. THE DIVERSITY OF WH-QUESTIONS

Languages present different syntactic mechanisms to express wh-questions. Some languages, for instance, **demand** that the interrogative element appear in front of the sentence as well as the inversion of the (auxiliary) verb and the subject (such as English and German):

(1)

a.   John **will** meet **Mary** tomorrow.

b.   *John **will** meet **who** tomorrow?[2]

c.   **Who will** [John ___ meet ___ tomorrow]?

---

[1] This is a preliminary version of the text. It intends to be an introduction to OT syntax and also to develop some of our ideas on the interaction of functionally motivated constraints in syntactic structure. Please feel free to send us comments by e-mail. Thanks to Les Zsoldos for comments on the first draft of this paper.

[2] This is a good sentence in English if it is used as an echo-question. Echo-questions are used in restricted contexts and we will not discuss them here (cf. Parker & Pickeral 1985, Sobin 1990, Pires & Taylor 2007).

Other languages **allow** these structural changes, but do not demand them (such as Portuguese and Spanish):

(2)

a. [O Paulo **diria a verdade**] em uma situação como esta.

[Paul **would say the truth**] in a situation like this.

b. [O Paulo **diria o quê**] em uma situação como esta?

[Paulo **would say what**] in a situation like this?

c. **O que** [o Paulo **diria** ___] em uma situação como esta?

What [Paul **would say** ___] in a situation like this?

d. **O que diria** [o Paulo ____ ___] em uma situação como esta?

What would say [Paulo ___ ___] in a situation like this?

In other languages, these changes **are not allowed**, under any circumstances (such as Chinese and Japanese):

(3)

a. [Yuehan **da-le shei**]?

Yuehan **hit who**?

b. ****Da-le** [Yuehan ___ **shei**]?

Hit Yuehan **who**?

c. ****Shei** [Yuehan **da-le** ___]?

Who Yuehan **hit**?

d. ****Shei da-le** [Yuehan ___ ___]?

Who hit Yuehan?[3]

Finally, there is a fourth group of languages which demands that **all** interrogative elements (in case that there is more than one) move to the front of the sentence (such as Bulgarian and Czech). This strategy is avoided in some languages that allow certain movements, such as English:

---

[3] Meaning 'Who has Yuehan hit'?

(4)

a. ***Kakvo** [Ivan **kupuva koj** ___]?

What [Ivan **bought for** whom ___]?

b. ***Kakvo kupuva** [Ivan ___ **koj** ___]?

What bought [Ivan ___ for whom ___]?

c. **Koj kakvo kupuva** [Ivan ___ ___ ___]?

For whom what bought [Ivan ___ ___ ___]?

(5)

a. **What** did John buy **for whom**?

b. ***For whom what** did John buy?

Superficially, it seems that each different language has a distinct mechanism to form wh-questions. This makes one think about these two questions: (i) Are these changes in the basic patterns of the sentences accidental? (ii) Why do languages apparently vary so much among each other when it comes to wh-question formation?

In the next sections, we will pursue these questions and try to show that the apparently random changes in syntactic patterns of wh-question formation in natural languages follow some universal principles which we will state as **constraints**. We intend to show what mechanisms act in wh-question formation. We believe that there are some **functional constraints** competing in the language systems, trying to **optimize** some important aspect for each specific language. These constraints are related to **Morphological Marking**, **Focalization** and **Economy**.

2. CONSTRAINTS INVOLVED IN WH-QUESTION FORMATION

2.1 MORPHOLOGICAL MARKING

It is not news that languages tend to use morphemes in the beginning of a string of words, marking it with a 'morphological mark', and signalizing it as a specific kind of

string. For example, it is common to begin a nominal phrase with an article. And the article can tell if the nominal phrase is definite or indefinite, as we can see in (6):

(6)
a. [+Definite]: **The** book on Linguistics
b. [-Definite]: **A** book on Linguistics

In some languages, such as Portuguese and Spanish, the article also determines if the nominal phrase is singular or plural:

(7)
a. [+Definite, +Singular]: **a** [garota bonita]

                        the [beautiful girl]

b. [+Definite, -Singular]: **a+s** [garota bonita][4]

                        the+plural-mark [beautiful girl]

In many languages we can see a similar phenomenon when it comes to question formation: certain specific morphemes can 'morphologically mark' a sentence as being a question. In other words, they mark the sentence with the feature **[+question]**:

(8)
a. **Czy** [zamykacie okna]? (Polish)
    Are you closing the windows?
b. **Aya** [Ali ketab darad]? (Persian)
    Does Ali have any book?
c. **Walay** [sarai khaza khuwakhae]? (Pashto[5])
    Does the man like the woman?

---

[4] This is not acceptable in standard formal Portuguese (Brazilian or European Portuguese).
[5] Indo-European language (Indo-Iranian language family) spoken in Afghanistan and western Pakistan.

In the languages that exhibit verb-subject inversion (such as English), the function of the verbal dislocation to the beginning of the sentence is precisely to mark the sentence with the feature [+question]:

(9)

a. **Can** [I __ call John for you]?
   ✓Morphological Marking: [+Question]

b. *[I **can** call John for you]?
   * Morphological Marking: [+Question][6]

In some languages, the sentences that are not marked with a specific morpheme (or in the languages that do not have this morpheme, but dislocate the verb to the initial position) violate the principle of Morphological Marking. In a similar way, we assume that the dislocation of any interrogative element (such as a wh-element) has a similar function, namely, to morphologically mark the sentence with the feature [+question]. That is, some languages, in order to morphologically mark a sentence as a question, **move some element to the sentence initial position**.

But we believe that there is another important detail in this mechanism: wh-questions do not demand a simple yes-or-no answer, as in (9). They are asking for some specific information, i.e., they have a specific **informational focus**. And some languages move precisely the wh-element to the sentence initial position in order to mark the sentence as interrogative. We will call this condition Morphological Marking with focus. Let's see some examples:

(10)

a. ***Can** [I __ call **who** for you]?
   ✓ Morphological Marking: [+Question]
   * Morphological Marking: [+WithFocus]

---

[6] We will mark with a ***** every violated condition and with the symbol ✔ every satisfied condition.

b. **Who can** [I __ call ___ for you]?
  - ✓ Morphological Marking: [+Question]
  - ✓ Morphological Marking: [+WithFocus]

In (10a), the sentence satisfies the constraint Morphological Marking [+Question], once there is a moved element (the auxiliary *can*) at the beginning of the sentence. However, this sentence violates the constraint Morphological Marking [+WithFocus], once the element that is the informational focus (the wh-word *who*) does not move, i.e., it remains *in situ*. We can conclude, at a first approximation, that the syntactic mechanisms involved in forming wh-questions in English have to deal with two constraints: Morphological Marking [+Question] and Morphological Marking [+WithFocus].

In (11), we give the definition of these two constraints:

(11) **Morphological Marking [MM]**:

a. **MM[+Q]**: an interrogative sentence must be structurally different than an affirmative sentence; it must begin with a specific marker (such as an auxiliary).

b. **MM[+FOC]**: an interrogative sentence with informational focus must begin with a specific element marking the focus (such as a wh-word).

The constraint **[MM]** is one general constraint that can act in many particular cases – every time some morphosyntactic information needs to be 'marked' in a constituent. In the case of wh-questions, there are **two** pieces of information to be pointed out; therefore, [MM] acts as it were **two** constraints: one demands the interrogative sentence to be marked MM[+Q]; and the other, [MM+ FOC].

2.2 FOCALIZATION

When an element is an informational focus, it is also the focus of attention of a sentence and it demands the most prominent position in the sentence – the initial position (cf. Dik 1989, 1997, Givón 1993, among others). We can see this in the question-answer pairs in (12) and (13). The question is such that it demands that the

information focus of the answer be **[Mary]** not **[John]** in (12) and **[John]** not **[Mary]** in (13).

(12)

A: **TO WHOM** did you see John giving flowers ___ ?

B: **TO MARY** I saw John giving flowers ___ .

　# **JOHN** I saw ___ giving flowers to Mary.[7]

(13)

A: **WHO** did you see ___ giving flowers to Mary?

B: **JOHN** I saw ___ giving flowers to Mary.

　# **TO MARY** I saw John giving flowers ___ .

The principle that the elements that are informational focus should occupy a prominent position in the sentence is also present when it comes to interrogative elements, such as (**TO**) **WHOM** or **WHO** in the sentences above. After all, an interrogative element defines the informational\attention focus in a sentence. In question-answer pairs with wh-elements, the focalized constituent will be the one fit to answer the question asked by the wh-element, as we saw in (12) and (13). In short, the principle of **Focalization** – focused elements must occupy the initial position in the sentence – also applies to interrogative elements (cf. Dik 1978, Costa 1998). The definition of the constraint **Focalization** is the following:

(14) **Focalization (FOC)**: elements that are the focus of attention must occupy the initial position in a sentence.

It is important to notice that, even though the constraints **MM[+FOC]** and **FOC** are related and cover some of the same cases, they **do not have the same effects**: MM[+**FOC**] is satisfied if **one** wh-element is moved to the sentence initial position; FOC can only be truly satisfied if **all** focused elements are moved to sentence initial position. Let us see the analysis of an English wh-question with these two constraints in mind:

---

[7] We use # to mark a contextually inadequate sentence.

(15)

a. *John gave **what present** to Mary?

   *MM[+FOC]

   * FOC

b. **What present** did John give __ to Mary?

   ✓MM[+FOC]

   ✓ FOC

c. *John gave **what present to whom**?

   * MM[+FOC]

   ** FOC

d. **What present** did John give __ to whom?

   ✓MM[+FOC]

   ✓* FOC

e. ***What present to whom** did John give __ __?

   ✓MM[+FOC]

   ✓✓ FOC

Sentence (15a) is ungrammatical, and it violates both constraints. Sentence (14b) is grammatical, and it satisfies both constraints, since there is an element moved to the sentence initial position (satisfying MM[+FOC]) and this element is the wh-constituent, informational focus (satisfying FOC). Sentence (15c) is ungrammatical, it violates MM[+FOC] once and FOC twice, since there are two wh-constituents that contain new information, and none of them have been moved to the prominent position of the sentence. Sentence (15d) is grammatical, but it only satisfies FOC once, since there is just **one** wh-constituent that moves to the sentence initial position. Finally, sentence (15e) is ungrammatical in English, even though it optimally satisfies both constraints. Sentences like that are well formed in languages like Bulgarian, as we saw in (4). This suggests that there may be another constraint acting in the formation of interrogative sentences. Let us see what constraint we are talking about.

2.3 ECONOMY

As we know, natural languages tend to present specific versions of the economic principle of 'the least-effort', which states something as follows: 'do not do anything unless you really have to'. When it comes to syntax, to be economic is to **avoid any changes in the basic sentence pattern of the language, unless you really have a good reason for that**. In the languages where subject-verb inversion takes place, it is easy to see this principle acting. In English, for example, the basic pattern of the declarative sentence – which is the unmarked pattern – is SVC. This pattern cannot be changed unless this change has a reflex in the propositional meaning or in the illocutionary force of the sentence. If the sentence is a simple declarative, any changes in the basic order of the elements are forbidden. After all, this change will violate the principle of economy for no reason.

(16)

a. [I can call John for you].

   Declarative sentence: ✓ **ECONOMY**

b. ***Can** [I __ call John for you].

   Declarative sentence: ***ECONOMY**

We can define **ECONOMY** as follows:

(17) **ECONOMY**: keep the basic pattern of the sentence.[8]

In some situations, **ECONOMY** can be violated, allowing the use of a different syntactic pattern. That can occur when **some other principle is demanding a change in the basic order of the sentence and is conflicting with Economy**. That happens, for example, with the interrogative questions: as we have seen, there is a universal principle that says 'interrogative sentences should be me morphologically marked as interrogatives,' and that means, in many languages, to move the verb to the sentence

---

[8] This constraint is based on the definition of other similar economy constraints, such as those proposed by Grimshaw (1997) and Legendre, Smolensky & Wilson (1998).

initial position. In doing so, the language is violating **ECONOMY**, in order to satisfy the constraint **MM[+Q]**, as we see in (18):

(18)
a. *[I can call John for you]?
   ✓**ECONOMY**
   *MM[+Q]

b. **Can** [I __ call John for you]?
   * **ECONOMY**
   ✓MM[+Q]

In (18a), the interrogative sentence satisfies **ECONOMY**, but violates **MM[+Q]**, and the sentence is ungrammatical. In (18b), we have the opposite: **ECONOMY** is violated in order to satisfy **MM[+Q]**, and we have a grammatical sentence in the language. So (18) show us that **ECONOMY** and Morphological Marking are conflicting conditions in the process of question formation in English. Whereas **ECONOMY** requires that the basic order of the sentence be maintained, **MM** requires that the sentence start with the verb. We can conclude that, in English, the constraint **MM+[Q]** is **more important** than **ECONOMY**. In other words, the language can violate **ECONOMY** in order to satisfy **MM[+Q]**.

(19) In English: **MM[+Q] >> ECONOMY**

3. Tableaux, hierarchies and the diversity of wh-question formation processes

The basic idea of Optimality Theory is that different grammars are the result of the different ways the languages adopt to resolve conflicts among universal conditions (such as Morphological Marking, Focusing, and **ECONOMY**). To solve the conflicts among universal constraints, it is necessary to establish some **hierarchy** concerning these constraints. Let's see how this works when it comes to wh-question formation.

Let's check the different hierarchies we find when we analyze the data we presented so far.

a) **MM >> ECONOMY >> FOC**: English, German

	Output candidate	MM[+Q]	MM[+FOC]	ECONOMY	FOC
a.	*Mary **would** buy **what for whom**?	*!	*		**
b.	***Would** Mary __ buy **what for whom**?		*!	*	**
c.	***What** Mary **would** buy __ for whom?	*!		*	*
d.	☞ **What would** Mary __ buy __ **for whom**?			**	*
e.	***For whom what would** Mary __ buy __ __ ?			***!	

Tableau 1: wh-questions in English and German[9]

In tableau 1, sentences (a) and (c) are excluded because they violate the leftmost constraint, the most important constraint in English: **MM[+Q]**. These violations are **fatal** (marked with a **!**). Sentence (b) is excluded because it violates the second most important constraint, **MM[+FOC]**. There are, then, only two possibilities left: (d) and (e). Both violate **ECONOMY**, the next constraint in the hierarchy. Sentence (e), however, violates **ECONOMY** three times, whereas sentence (d) violates it only twice. As there is no sentence left in our analysis, sentence (d) is the 'optimal candidate, i.e., the one that is grammatical (pointed out by ☞). Even though candidate (d) is not 'perfect' (it violates **ECONOMY** twice and **FOC** once), it is 'the best' candidate in the analyses, for it does not violate any of the most important constraints (**MM**) and it violates **ECONOMY** less than candidate (e).

---

[9] To 'read' the tables (called *tableaux*), one should start from left to right: the higher (more important) a constraint is, the leftmost position it occupies. The shadowed squares are irrelevant for the optimal candidate, since the conflicts were already solved by some higher constraint.

b) **MM >> FOC >> ECONOMY**: Bulgarian

	Output candidate	MM[+Q]	MM[+FOC]	FOC	ECONOMY
a.	*Ivan **kupuva koj kakvo**?	*!	*	**	
b.	***Kupuva** Ivan __ **koj kakvo**?		*!	**	*
c.	***Kakvo** Ivan **kupuva koj** __?	*!		*	*
d.	***Kakvo kupuva** Ivan __ **koj** __?			*!	**
e.	☞ **Koj kakvo kupuva** Ivan __ __ __?				***

Tableau 2: wh-questions in Bulgarian

In the grammar of Bulgarian the most important condition is **MM**, since, in this language, it is necessary that all interrogative sentences be marked **and** that all wh-elements move to sentence initial position. Therefore, the optimal candidate is (e), even though it violates **ECONOMY** (a low-ranked constraint in the language) three times, in order to satisfy the most important conditions **MM** and **FOC**.

c) **ECONOMY >> MM >> FOC**: Chinese, Japanese

	Output candidate	ECONOMY	MM[+Q]	MM[+FOC]	FOC
a.	☞ Yuehan **wheisheme da-le shei**?		*	*	**
b.	***Da-le** Yuehan **wheisheme** __ **shei**?	*!		*	**
c.	***Shei** Yuehan **wheisheme da-le** __ ?	*!	*		*
d.	***Shei da-le** Yuehan **wheisheme** __ __ ?	**!			*
e.	***Wheisheme shei da-le** Yuehan __ __ __ ?	***!			

Tableau 3: wh-questions in Chinese and Japanese

Chinese is opposite to Bulgarian in relation to **ECONOMY**. In Chinese **ECONOMY** cannot be violated to satisfy another constraint, under any circumstance. It is the most important constraint to be obeyed in the language. Therefore, the optimal candidate is (a), which presents two wh-elements in situ.

d) **MM << >> ECONOMY >> FOC**: Portuguese, Spanish

In Portuguese (at least in Brazilian Portuguese), there are four possibilities to form a wh-question, concerning the constraints we have studied so far:

(20)

a. O João **tem** dado **o quê pra quem**?

   John **has** given **what to whom**?

b. **Tem** o João ___ dado **o quê pra quem**?[10]

   Has John __ given **what to whom**?

c. **Pra quem** o João **tem** dado **o quê** ___?

   To whom John **has** given **what** ____?

d. **Pra quem tem** o João ___ dado **o quê** ___?

   To whom has John __ given **what** ___?

e. ***O que pra quem tem** o João dado ___ ___ ___?

   What to whom has John given ___ ___ ___?

We formalize these four possibilities in the following tableau[11]:

	Output candidate	MM[+FOC]	MM[+Q]	ECONOMY	FOC
a.	☞ O João *teria* dado *o quê pra quem*?	*	*		**
b.	☞ *Teria* o João __ dado *o quê pra quem*?	*		*	**
c.	☞ *Pra quem* o João *teria* dado *o quê* __?		*	*	*
d.	☞ *Pra quem teria* o João __ dado *o quê* __?			**	*
e.	**O que pra quem teria* o João dado __ __ __?			***!	

Tableau 4: wh-questions in Brazilian Portuguese

Sentences (a) to (d) are grammatical in Portuguese. It seems that **MM** and **ECONOMY** are in the same level of importance in the hierarchy for Portuguese, and both are more important (and thus are ranked higher) than **FOC**. It seems we have some kind

---

[10] Only in very formal discourse.
[11] The dotted lines indicate that the constraints are equal in the ranking of the language.

of **optionality** among the four grammatical candidates. The only clear ungrammatical candidate is (e), which violates ECONOMY in order to satisfy MM and FOC[12].

5. SUMMARY AND CONCLUSION

Even though wh-questions apparently differ throughout the languages, we tried to show here that they are the result of a grammar that is trying to resolve the conflicts of four general and universal principles, namely **MM[+Q], MM[+FOC]**, **Focalization** and ECONOMY. The diversity among languages exists because different languages use different ways to solve the conflicts among the constraints of universal grammar. Each language has its own organization of the hierarchy of the same basic and universal principles in the grammar.

REFERENCES

ACKEMA, P.; NEELEMAN, A. Whot?. In: BARBOSA, P. et al. (eds.), **Is the best good enough?** Optimality and competition in syntax. Cambridge: MIT Press, 1998.

COSTA, J. **Word order variation.** A constraint-based approach. Holland Academic Graphics, Haia, 1998. *LOT Dissertations.*

DIK, S. **Functional Grammar**. Amsterdan: North Holland, 1978.

DIK, S. **The theory of Functional Grammar**. Dordrecht: Foris, 1989.

DIK, S. **The theory of Functional Grammar:** Part II. Berlin, New York: Mouton de Gruyter, 1997.

GRIMSHAW, G. Projections, heads, and optimality. **Linguistics Inquiry** 28(3), 1997.

GIVÓN, T. **On understanding grammar**. New York: Academic Press, 1979.

GIVÓN, T. **English grammar – a function-based introduction**. Vol. II. Amsterdam/Philadelphia: John Benjamins, 1993.

GIVÓN, T. **Syntax**. Amsterdam/Philadelphia: John Benjamins, 2001.

---

[12] This poses a problem for the theory, since OT is a theory of **one optimal candidate**. How should we explain the optionality among more than one candidate? We will not seek the analysis for Portuguese here, neither will we discuss the issue of optionality (cf. Keer & Bakovic 1997, Costa 1998, and Pesetsky 1998 for a discussion on the problem of optionality within the OT framework).

KEER, E.; BAKOVIC, E. Have faith in syntax. In: CURTIS, E.; LYLE, J.; WEBSTER, G. (eds). **Proceedings of the Sixteenth West Coast Conference on Formal Linguistics**. CSLI, Stanford, 1997.

LEGENDRE, G.; SMOLENSKY, P.; WILSON, C. When is less more? Faithfulness and minimal links in *wh*-chains. In: BARBOSA, P. et al. (eds.), **Is the best good enough?** Optimality and competition in syntax. Cambridge: MIT Press, 1998.

PARKER, F.; PICKERAL, J. Echo Questions in English. **American Speech**, vol. 60, n. 4, 1985.

PESETSKY, D. Principles of sentence pronunciation. In: BARBOSA, P. et al. (eds.), **Is the best good enough?** Optimality and competition in syntax. Cambridge: MIT Press, 1998.

PIRES, A.; TAYLOR, H. L. The syntax of wh-in-situ and common ground. In: MASULLO, P. (ed.) **Romance Languages: structure, interfaces, and microparametric variation**. Amsterdam: John Benjamins, 2007.

SOBIN, N. On the syntax of English echo questions. **Lingua** 81:2-3, 1990.