JOURNAL OF ENGLISH STUDIES – vol. xx (2026): xx-xx. http://doi.org/10.18172/jes.6612

A CORPUS-ASSISTED ANALYSIS OF CONJOINED WH-QUESTIONS IN ENGLISH

MARIAN ALVES-CASTRO
Universidad del País Vasco/Euskal Herriko Unibertsitatea
marian.alves@ehu.eus

AZLER GARCIA-PALOMINO
Universidad del País Vasco/Euskal Herriko Unibertsitatea
azler.garcia@ehu.eus

ABSTRACT. This corpus study investigates conjoined *wh*-questions (CWHs), where two or more coordinated *wh*-phrases appear at the beginning of a sentence (e.g., *When and where did they first meet?*). Previous research suggests that English speakers readily accept adjunct coordination but generally reject argument coordination, while mixed patterns (involving an adjunct and an argument) are subject to interspeaker variation. Using data from the *Corpus of Contemporary American English* (COCA), this study confirms that English strongly favors adjunct coordination (over 90% of CWHs), while mixed coordination is uncommon (under 8%), and argument coordination is virtually absent. Notably, all coordinated arguments in the corpus were optional, suggesting structural constraints on CWHs. Finally, the data support a biclausal analysis of CWHs, reinforcing the idea that coordinated *wh*-phrases originate in separate clauses.

Keywords: corpus linguistics, syntax, coordination, mixed languages, *wh*-questions, optional arguments.

ANÁLISIS ASISTIDO POR CORPUS DE ORACIONES INTERROGATIVAS CON COORDINACIÓN EN INGLÉS

RESUMEN. Este estudio de corpus investiga oraciones interrogativas que comienzan con dos o más sintagmas interrogativos coordinados (por ejemplo, *When and where did they first meet?*). Investigaciones previas indican que en inglés se acepta la coordinación de adjuntos, mientras que la de argumentos es sistemáticamente rechazada; la coordinación mixta (adjunto y argumento) muestra mayor variabilidad. Con datos del *Corpus of Contemporary American English* se confirma que el inglés favorece abrumadoramente la coordinación de adjuntos (más del 90% de los casos), mientras que la coordinación mixta es poco frecuente (menos del 8%) y la de argumentos casi nula. Todos los argumentos coordinados que aparecen en los datos son opcionales, lo que sugiere restricciones estructurales en estas interrogativas. Finalmente, los datos respaldan un análisis bi-oracional, reforzando la idea de que los sintagmas interrogativos coordinados se originan en cláusulas separadas.

Palabras clave: lingüística de corpus, sintaxis, coordinación, lenguas mixtas, interrogativas parciales, argumentos opcionales.

Received: 28/02/2025

Revised version accepted: 29/10/2025

1. INTRODUCTION

Questions featuring two or more coordinated *wh*-phrases in a left-peripheral position, as illustrated in (1), appear across a wide variety of typologically diverse languages. We will borrow the term "conjoined *wh*-question" (CWH) from Gračanin-Yuksek (2) to refer to this construction ("Conjoined").

(1) **When and where** did they first meet?

In CWHs, the *wh*-phrases in sentence-initial position have different grammatical functions and, when both are adjuncts (i.e., optional modifiers), they contribute different types of information to the description of the event (time, place, manner, etc.).[1] Thus, CWHs must be distinguished from sentences like (2), where the two *wh*-phrases have the same syntactic function:

[1] The distinction between arguments (or complements) and adjuncts goes back to Tesnière's concepts of *actants* (105-125) and *circonstants* (125-129). Arguments are elements required by the verb's valency or, at the very least, shaped by it. Adjuncts, on the other hand, are more flexible–they are not determined by the verb's valency and can be added or left out without affecting grammaticality. Typically, they function as sentence modifiers (see Allerton 40, 58); Faulhaber (4-8), among others). However, the line between arguments and adjuncts is not always clear. Many tests have been proposed to distinguish the two, but the issue remains unresolved (cf. Helbig 78–87; Herbst and Schüler 113–116; see also the discussion in Hole. We will use the term "adjunct" to refer to constituents which are not part of the thematic

(2) **What and who** do you miss?

These interrogatives will not be considered in this study because they do not display the same syntactic behavior as questions like (1), suggesting that they instantiate a different phenomenon. Thus, the two *wh*-phrases can only be conjoined when they share the same grammatical function, as illustrated by contrast in (3):

(3) a. Who and what do you miss?

b. *What and to whom did you give?

This restriction is also reflected in the corresponding declaratives, as in (4):

(4) a. I miss my friends and all my old books.

b. *I gave a book and to Mary.

Similarly, although both (1) and (2) favor single-pair interpretations, potential answers to these questions can only be coordinated when the *wh*-phrases have identical syntactic functions, as shown in the ungrammaticality of (5a) and the grammaticality of (5b), fragment answers to (1) and (2) respectively.

(5) a. *last summer and in the local pub

b. my friends and all my old books

We therefore treat sentences such as (1) and (2) as instances of two distinct phenomena. Following standard practice in the literature on CWHs, our analysis focuses on cases in which the conjoined *wh*-phrases either have distinct grammatical functions or, if both are adjuncts, contribute different meanings.

It must be noted that, in English, it has been observed that coordination of adjuncts, as in (1), is generally accepted, whereas coordination of arguments, as in (3b) above, is rejected. Some interspeaker variation (indicated by the symbol "%") arises in mixed coordination (adjunct and argument), as in (6):

(6) %**Where and what** did you eat?

The aim of the present study is to see if these observations are confirmed by a quantitative analysis of the different patterns of coordination present in the *Corpus of Contemporary American English* (COCA). The working assumption is that a higher frequency in the corpus will be indicative of a higher frequency of use and, consequently, of a higher degree of acceptance among speakers.

structure of the predicate, i.e., syntactic objects whose presence is not required to build a grammatical sentence. Arguments, for their part, are the constituents required by the predicate (roughly, direct and indirect object, and subject).

By analyzing the different coordination patterns, we aim to identify the constraints governing this construction in English, as understanding these constraints may also shed light on its correct analysis. In this respect, mixed CWHs and sentences involving the coordination of two *wh*-arguments are particularly relevant to this study.

The paper is organized as follows. Section 2 sets the scene by providing a brief literature review on the construction in English. Section 3 presents the details of the corpus study and section 4 closes the paper with the conclusions.

2. CONJOINED WH-QUESTIONS IN ENGLISH: A BRIEF LITERATURE REVIEW

Conjoined wh-questions do not appear to have held a predominant position on the research agenda during the early decades of generative grammar. Consequently, studies on CWHs before the 1990s are scarce, with only a few exceptions such as the articles by Browne, Grimshaw, and Rudin. However, in the 1990s and the first two decades of the twenty-first century, researchers turned to the construction with renewed interest and, as a result, a number of relevant papers were published.

Not all languages are equally represented in the discussion, however. Thus, while Eastern European languages such as Hungarian, Serbo-Croatian, Polish, Russian or Romanian feature quite prominently in the literature, other languages like Spanish, Greek, French or Italian are very clearly underrepresented. Concerning English, the discussion of CWHs appears frequently in papers focusing on other languages, but there are also some articles whose primary object of study is English, as will be seen below.

The discussion of CWHs usually revolves around two topics: (i) the characterization of the construction, with special attention paid to the (im)possible combinations of *wh*-phrases; and (ii) the proposal of an analysis. These two axes provide a helpful framework for organizing this literature review.

2.1. Patterns of coordination

A pervasive observation reported in the literature is that not all languages allow for the same combinations of *wh*-phrases. Languages have been shown to fall into one of the four groups in (7), a typology proposed by Lipták (154) based on the study of the 13 languages mentioned here. The author points out that this classification is based on CWHs which have the same prosodic contour as simple questions. Those with the typical prosodic signature of parentheticals are excluded from her study.

(7) *Lipták's typology of languages*

a. free languages (argument and argument, argument and adjunct, adjunct and adjunct): Hungarian, Polish, Croatian, Romanian, Bulgarian and Russian

b. mixed languages (optional argument and adjunct, adjunct and adjunct): English and German

c. adjunct languages (adjunct and adjunct): English, Italian, Spanish and Dutch

d. non-coordinating languages (Ø): Chinese and Japanese[2]

Leaving aside languages claimed not to allow the construction at all (such as Chinese and Japanese) and focusing on those where CWHs are attested, different levels of restrictiveness emerge. The most restrictive languages, exemplified in (7c), allow only the coordination of *wh*-adjuncts (which are optional by definition). In contrast, the least restrictive languages, shown in (7a), permit both arguments (whether optional or obligatory) and adjuncts in the construction. Additionally, there is an intermediate type: so-called mixed languages, which permit adjuncts and optional arguments.

Across languages, the main divide appears between single fronting and multiple fronting languages, with the latter being much more permissive with the combinations of *wh*-phrases allowed, (7a) v. (7b, c). As can be seen in this typology, English (a single-fronting language) is classified as both a mixed language and an adjunct language. According to Lipták (154), some of her informants appeared to use a more restrictive variant of the language, allowing only the coordination of adjuncts, while others accepted the inclusion of optional arguments.

The consensus observed regarding the acceptability of sentences involving the coordination of two *wh*-adjuncts would justify the classification of English as an adjunct language. Browne already claimed that "none of the question words can be a 'bound' part of the sentence; all must be 'free' parts" (223), where "bound" roughly means obligatory argument of the verb. However, interspeaker variation is commonly reported in the case of English. As mentioned in the introduction, acceptability judgements become more unstable in the case of the mixed pattern, see (6) above.[3] The fact that mixed coordination is not accepted by all speakers has

[2] However, Kasai convincingly argues that Japanese not only has CWHs but also permits all combinations of *wh*-phrases. He advocates a biclausal-plus-ellipsis analysis of Japanese CWHs. Similarly, Zhang provides examples of CWHs in Chinese and argues for a monoclausal analysis (2136). See section 2.2 below for details on the different analyses.

[3] When discussing acceptable and unacceptable patterns, we primarily rely on claims found in the literature. It should be noted that most research on CWHs focuses on theoretical issues and often relies on a small number of informants, sometimes without specifying exact numbers (e.g., Lipták). Exceptions include Whitman ("What and How"), who collected grammaticality judgments on mixed patterns from 18 native speakers, and Lewis et al., who

awoken the interest of researchers, who try to answer the questions of why such interspeaker variation should arise and how the grammar should tackle it. Sentences like those in (8), involving the coordination of an adjunct (typically, *where*) with an object (*what*) have been discussed in many articles on CWHs (see Gračanin-Yuksek's "What and Why" (6-12), "Conjoined" (4-7); Lewis et al.; Whitman's *Category* (73-85), "What and How"; among others).

(8) **Where and what** did you eat? v. %**What and where** did you eat?

Sentences headed by optionally transitive predicates like *eat* seem to represent the type of context in which mixed coordinations are most likely to occur. This context has been explored in greater detail in the literature. However, questions regarding the (im)possibility of having subjects in CWHs are generally overlooked, with the notable exceptions of Gračanin-Yuksek ("What and Why" (12) and *About* (42)), who briefly mentions that subjects are not possible, and Citko (303-304), who offers a more detailed discussion. Similarly, the coordination of two arguments is often dismissed as impossible and left unexplored. We will come back to this in sections 3.4 and 3.5 below.

Interspeaker variation can complicate the determination of which combinations of *wh*-phrases are permitted and which restrictions apply to the construction. When judgments from native speakers alone may lead to unreliable conclusions, corpora can provide crucial support. The frequency with which certain patterns appear in a corpus can offer a more accurate representation of native speaker behavior, leading to a clearer understanding of actual usage.

2.2. Analyses

The divide between single and multiple fronting languages mentioned above translates into different analyses for these two groups of languages. For the former, a monoclausal derivation, represented schematically in (9a), is usually assumed; for the latter, a biclausal one, as shown in (9b).

(9) a. [$_{CP}$ [$_{\&P}$ wh_1 and wh_2] [$_{C'}$ [$_{TP}$ … t_1…t_2…]]]

b. [$_{\&P}$ [$_{CP}$ wh_1 [$_{TP}$ … t_1 …]] and [$_{CP}$ wh_2 [$_{TP}$ … t_2…]]][4]

studied processing of mixed CWHs with 42 participants. Clearly, a deeper understanding of CWHs would benefit from more extensive experimental studies, which is a matter for future research. For the purposes of this paper, we will use "(un)acceptable" and related terms in a broader, more informal sense rather than the strict experimental linguistics sense.

[4] The labels in the bracketed structures follow the standard conventions of the X-bar theory's structural schema, as proposed in Chomsky's *Lectures* and *The Minimalist*. List of abbreviations: CP stands for Complementizer Phrase; &P for Coordination Phrase; TP for Tense Phrase; t_1 and t_2 are the traces left by the moved *wh*-phrases (wh_1 and wh_2); finally, C'

2.2.1. Monoclausal analyses

The monoclausal or "small coordination" accounts assume the projection of only one CP. Consequently, all the coordinated *wh*-phrases have their base position in this CP and move to its Specifier (Spec).[5] As multiple *wh*-fronting is a required step in this derivation, monoclausal accounts are assumed to be restricted to multiple fronting languages, as just mentioned. The different variants of the analysis depart from each other basically with respect to the strategy used to insert the coordinating particle. Although the norm is that the authors remain vague in this respect, some have proposed sideward movement (as in Nunes) of the *wh*-phrases as the mechanism of insertion of the coordinating conjunction. The *wh*-phrases are copied and merged with the coordinating head in a different derivational space, (10b). The &P thus assembled will be later merged into SpecCP, as in (10c):

(10) a. [$_{C'}$ did [$_{TP}$ you meet Peter where when]]

b. [$_{\&P}$ when [and where]] → sideward movement

c. [$_{CP}$ [$_{\&P}$ when [and where]] [$_{C'}$ did [$_{TP}$ you meet Peter t_{where} t_{when}]]

Among the proponents of small coordination are Kazenin and Gribanova for Russian, Haida and Repp for all free languages, and Zhang and Potter and Frazier for English. Advocates of sideward movement are, for instance, Haida and Repp, and Citko and Gračanin-Yuksek ("Towards"). However, assuming a monoclausal analysis is not the most common option for single fronting languages like English, which are more generally assumed to be underlyingly biclausal.

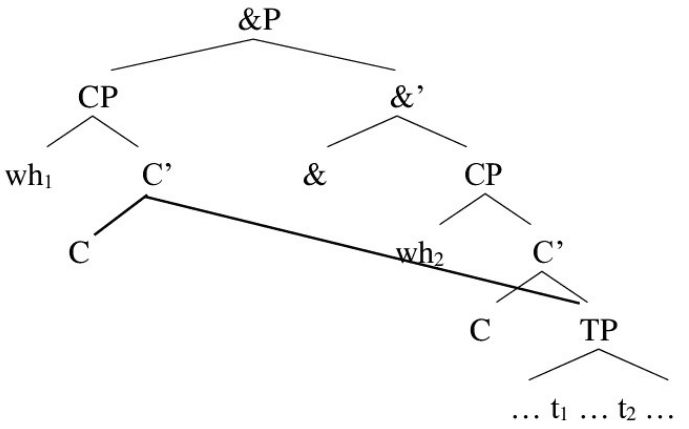
2.2.2. Biclausal analyses

Biclausal analyses have been proposed within both multi-dominance and single-dominance frameworks. Barbara Citko and Martina Gračanin-Yuksek, working within the multi-dominance framework, have addressed this construction in several papers (Gračanin-Yuksek, *About*; "What and Why"; "Conjoined"; Citko; Citko and Gračanin-Yuksek, "Towards"; "Multiple"), and presented some alternative accounts, with the most prominent being the two in (11) and (12), referred to as "bulk sharing" and "non-bulk sharing," respectively. See also Raţiu.

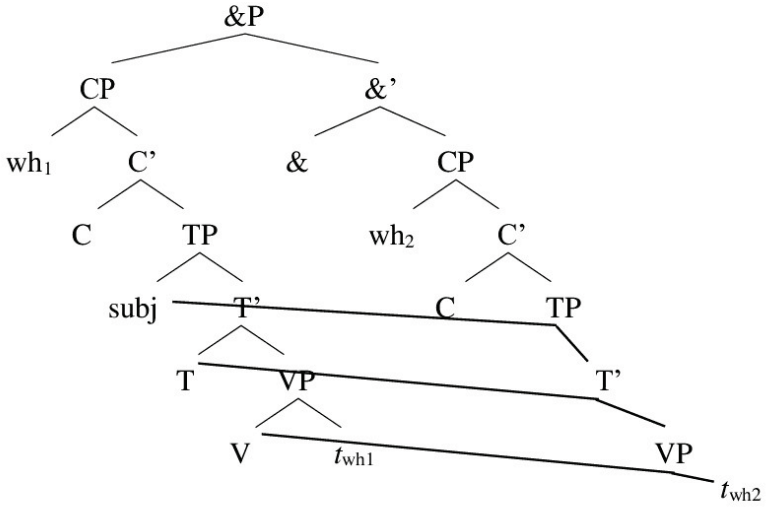
is an intermediate category in the X-bar schema, as it is projected between the complementizer head C and the maximal category CP.

[5] Not all authors endorse an analysis in terms of movement of all the *wh*-phrases to the left. Comorovsky and Skrabalova, for instance, argue that the *wh*-phrases are base-generated in their left-peripheral position (qtd. in Gračanin-Yuksek, "Conjoined" 15).

(11) bulk sharing



(12) non-bulk sharing



In this type of analysis, there is actual *sharing* of one or more constituents, i.e., the same constituent (or group of constituents) is part of two different CPs. Notice also that, in the case of bulk sharing in (11), the two *wh*-phrases have their thematic position in the same TP, whereas, in the case of non-bulk sharing, two different TPs are projected.

In the mainstream single-dominance framework, two biclausal analyses are usually proposed, one in terms of Across the Board (ATB) rightward movement, the other one in terms of ellipsis. Both approaches start with the assumption that each coordinated *wh*-phrase occupies the Spec position of an independent CP, (13b), but differ in the strategy used to reduce the structure of the first conjunct.

(13) a. When and where did you meet Peter?

b. [$_{CP}$ when did you meet Peter t_{when}] and [$_{CP}$ where did you meet Peter t_{where}]

Proponents of the ATB movement account assume that the constituent in italics in (14) below moves to the right to derive the CWH in (13a). Variants of this analysis have been proposed, for example, by Park and Haida and Repp for English-type languages. For a different view, see Giannakidou and Merchant, who explicitly argue against this analysis and in favor of ellipsis.

(14) [$_{CP}$ when [*did you meet Peter* t_{when}]] and [$_{CP}$ where [*did you meet Peter* t_{where}]]

→ [$_{CP}$ when t_i and where t_i][did you meet Peter t_{wh}]$_i$?

In the ellipsis analysis, each *wh*-phrase moves to SpecCP in its own clause in the usual manner. The biclausal structure is then reduced by eliding everything in the first conjoined CP except for the *wh*-phrase. Ellipsis takes place under identity with the corresponding constituent in the second conjunct (reverse sluicing). This type of analysis is proposed for English by Browne, Giannakidou and Merchant, Lipták, Tomaszewicz ("Against" and "Wh&Wh"), and Whitman (*Category*), among others. A schematic representation is provided in (15), where strikethrough represents ellipsis.

(15) [$_{CP}$ when did you meet Peter t_{when}] and [$_{CP}$ where did you meet Peter t_{where}]

→ [$_{CP}$ when ~~did you meet Peter t_{when}~~] and [$_{CP}$ where did you meet Peter t_{where}]

3. THE CORPUS STUDY

3.1. About COCA

The *Corpus of Contemporary American English* (COCA), created by Mark Davies of Brigham Young University, provides a comprehensive representation of modern American English. This corpus contains over 1 billion words, making it one of the largest resources of its type. Spanning from 1990 to 2019, COCA includes more than 25 million words each year from a variety of genres: spoken language, fiction, popular magazines, newspapers, academic texts, and, with an update in March 2020, subtitles from TV shows and movies, blogs, and other web content. This diversity ensures a broad representation of both written and spoken American English, ranging from formal to informal registers.

As will be explained in the following section, the search was restricted to the subcorpus for the period 2015-2019, which contains, according to the corpus website, around 100 million words, offering a substantial sample for analyzing language patterns and trends during this period while preserving the rich diversity and broad scope of the overall corpus.

3.2. Methodology

The search tool available on COCA was used to search for concordances with the strings in the following sets in the two possible linear orders.

(16) a. *wh*-adjunct & *wh*-adjunct:

{when, where, how, why} and {when, where, how, why}

b. mixed coordination:

{when, where, how, why} and {what, who(m), to who(m)}

c. *wh*-argument & *wh*-argument:

{what, who(m), to who(m)} and {what, who(m), to who(m)}

Since the number of concordance lines for some "adjunct and adjunct" patterns in the entire corpus is too large for analysis, we decided to limit the search to the subcorpus from 2015 to 2019. For this, we applied the restriction option available in the corpus search tool.

The lists of concordances obtained were pruned manually in order to discard those sentences that are irrelevant to the present study. The discarded sentences can be grouped together in the following categories:

1) Sentences that are not interrogative, as the following:

(17) a. Arnold cuts **where and how** he wants. The bi-lateral incision ends up crooked, wonky-looking … (FIC 2019)

b. More and more, users are turning to their phones and tablets to view content **where and when** they want. (MAG 2016)

2) Sentences in which the conjoined *wh*-phrases have the same grammatical function, as in (18). These are especially abundant in the argument and argument patterns.

(18) a. The footprints of the fallen towers are now a haunting memorial to **what and who** was lost here. (TV 2016)

b. I know **who and what** I am and do not misrepresent myself or my information. (MAG 2018)

3) Sentences in which the two *wh*-phrases belong in two different questions. The first *wh*-phrase, in these cases, is *in situ* in the first sentence and the second in the left periphery of the second sentence. This category also includes questions that are typically analyzed as involving ellipsis in the second conjunct, as in (19c).

(19) a. … and researchers have a lot of questions about what kind of animals are living **where and how** they interact with humans and with one another. (NEWS 2015)

b. There was strength in knowing what was **what and who** was who. (FIC 2017)

c. I want a full chronology, who called **who and when**, all right? (MOV 2016)

4) Sentences in which the *wh*-phrases are nominalized.

(20) a. … and then worry about the **where**, **what and how** of a new stadium (NEWS 2015)

b. … were integrally linked to differences in the **what and how** of assessing student learning. (ACAD 2015)

c. … you aren't wondering what will happen, the true weight of the **who and why** comes to the forefront. (MAG 2017)

5) Sentences where sluicing appears to have applied across the board, leaving just the coordinated *wh*-phrases, as in (21).

(21) a. How often is the song played? **When and why and where**? (MAG 2015)

b. It's Payne. It's Mason. We need to talk. **When and where**? (MOV 2019)

c. Yes, Dina, I do. Question is … **What and where**? I'm gonna find out both. (TV 2018)

6) Finally, a number of sentences were discarded because they did not strictly involve the coordination of the *wh*-phrases in (16) above, either because the *wh*-phrase was preceded by a preposition (*since when, with whom, about what,* etc.) or because it was followed by a noun (22c), an adjective, an adverb (*how long, how often, how far*), or a quantifier (*how many, how much*). Some examples follow.

(22) a. … a computer monitor that measures **where and how long** a person gazes at the screen (MAG 2017)

b. … she will now be able to control not only her suffering but **where, with whom and when** she dies. (SPOK 2016)

c. One of the biggest concerns is that it would give organizations way too much latitude to determine **when and what privacy protections** a consumer should receive. (NEWS 2015)

Although the sentences in categories (1) to (6) above are acceptable and interesting in their own right, they are excluded to maintain coherence with the definition of the object of study, which is restricted to the single-word *wh*-pronominals and adverbials in (16) above. Discarding these sentences will also help to keep the corpus search within bounds.

Once the process of pruning is finished, the sample of sentences to be analyzed is made up of main and embedded full questions with two or more coordinated *wh*-phrases in the left periphery. These sentences are well-formed and the *wh*-phrases, all members of the inventory presented in (16) above, have different grammatical functions in the sentence.

3.3. Analysis

For the statistical analysis, the list of concordances was grouped into three datasets, namely, one that included all concordances (786 observations), one that included two coordinated adjuncts (727 observations), and one that included mixed coordination of an adjunct and a *what* or *who* argument (55 observations).[6] We then fitted a Bayesian multinomial logistic regression model to each dataset using the R (R Core Team) package *brms* (Bürkner). *brms* provides a user-friendly interface to the probabilistic programing language *Stan* (Stan Development Team). The inferential statistics presented here is intended to answer our research questions through specific model selection (Winter 109) and to shift our focus from terminology such as statistical significance onto estimation and the uncertainty therein (Wakefield 22-23).

The model with all concordances is an intercept-only model with no group-level adjustments, i.e., random effects. The response variable pattern includes all possible combinations of CWH types found in the subcorpus: "adjunct & adjunct" and "adjunct & argument". We treated the response variable as categorical instead of binomial despite there being only two levels in order to estimate the baseline probabilities for each coordination pattern. The model formula is defined as follows: brms::bf(pattern ~ 1). In addition, we specified moderately normally distributed and moderately informed priors for the mean probability of "adjunct & argument" patterns to be centred around 8%, with relative uncertainty around the mean: *Normal*(-2.5, 1). Prior selection was informed by theoretical claims that English more

[6] Observations of *whom* and *to whom* were excluded from this dataset due to their low number of occurrences (n = 4) and because of their grammatical restrictions to function as subjects or complements of the verb *to be*, a constraint that will be made relevant to the statistical analysis below.

frequently shows adjunct combinations than mixed "adjunct & argument" combinations (see the overview in Gračanin-Yuksek, "Conjoined", among others).

The model including "adjunct & adjunct" combinations only is also an intercept-only model with no group-level adjustments aiming to estimate the probability of each 11 "adjunct & adjunct" combinations from the data (e.g. *where and how*, *when and why*, etc.) in any order. The model formula is defined as follows: brms::bf(combo ~ 1). Flat, uninformative priors were set for this model due to the absence of exact quantification regarding probabilities of "adjunct & adjunct" combinations in the literature. As a consequence, very few constraints were placed on the data.

The model including "adjunct & argument" combinations only was built to estimate the probability of each grammatical function of the coordinated argument (i.e., subject, object, or complement of the verb *to be*) as a function of whether the argument *wh*-phrase was either *what* or *who*. The specific adjunct item was entered as a group-level adjustment to control for the variability derived from whatever *wh*-phrase was coordinated with the argument. Similarly to the "adjunct & adjunct" model, flat, uninformative priors were set for this model due to the absence of exact quantification regarding probabilities of "adjunct & argument" combinations in the literature.

3.4. Results

3.4.1. Coordination of two adjuncts

The raw frequency of the "adjunct and adjunct" pattern for the period under study after pruning is of 727 sentences, distributed as shown in figure 1. The figures beside each bar represent total token numbers for each *wh*-adjunct combination.

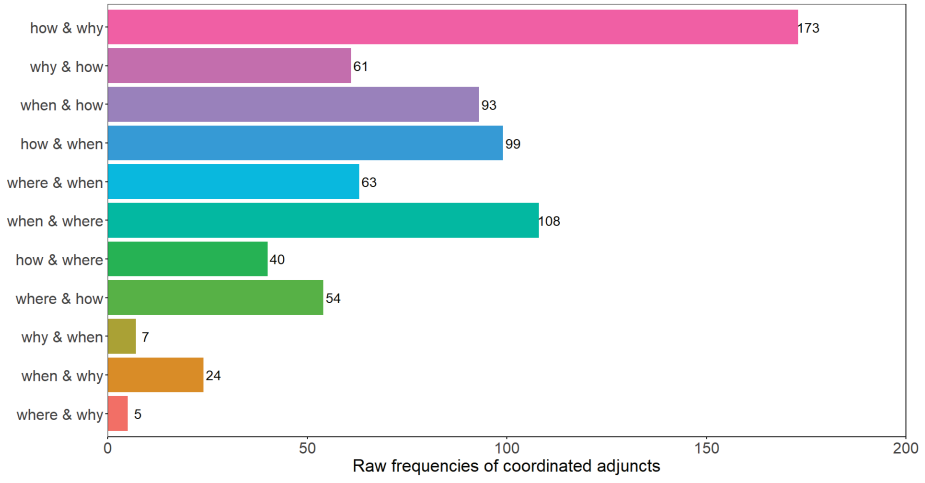


Figure 1. Raw frequencies of the different "adjunct & adjunct" patterns.

The first observation worth mentioning is that not all combinations of *wh*-adjuncts are equally represented in the sample. Moreover, there are substantial differences among coordination patterns and even between the two linearities involving the same conjoined phrases. Thus, the most frequent combinations are *how and why* (173 sentences, which amount to around 23.8%) followed by *when and where* (with 108 sentences, i.e., around 14.9% of all cases). *How and when* and *when and how* follow closely with 99 and 93 sentences, respectively. This is the only case in which there seems to be no clear preference for one linear order over the other. On the opposite side are the combinations of *why* with *when* and *where*. As can be seen, there are no concordances for *why and where* in the period under study; for *where and why* the raw frequency is 5, for *why and when* 7, and for *when and why* 24. Some examples of the "adjunct and adjunct" pattern follow in (23) below:

(23) a. Throughout the course of the Mueller investigation, there are repeated questions about **when and why** various members of the Trump campaign and White House met with people connected to the Russian government. (SPOK 2019)

b. Scholars have been operating within the framework of the liberal and republican traditions to determine **where and how** speech protections ought to apply. (ACAD 2017)

c. **Where and when** will we see the first drone deliveries? (MAG 2015)

d. Figuring out **when and how** to reveal what you know is a game like the one played by the TV detective Columbo. (FIC 2019)

As for the statistical model, model estimates are outlined in figure 2 below, with median estimates and 95% credible intervals shown alongside probability distributions. Please note that a summary of the model coefficients is included in the Appendix. Model estimates confirm the patterns shown by visual inspection of raw frequencies such that combinations of *how and why* are the most frequent at 23.8% [20.7%, 26.9%]. Combinations of *where and where*, *how and when*, and *when and how* appear to cluster together at relatively high frequencies; and the same applies to *where and when*, *why and how*, and *where and how*, although at somewhat lower frequencies. *how and where* seems to occur at even lower rates, different from all combinations mentioned above (5.5% [4%, 7.3%]). The remaining combinations occur only marginally.

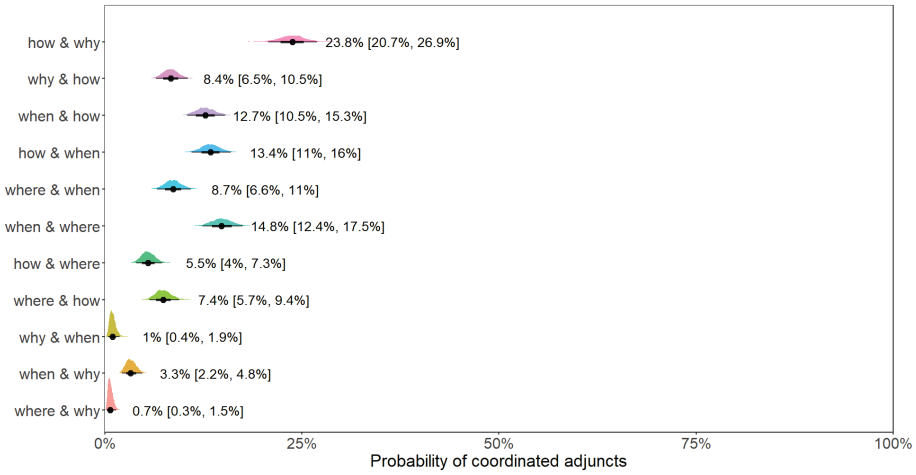


Figure 2. Model estimates for "adjunct & adjunct" combinations.

3.4.2. Mixed coordination

The presentation of the results of mixed patterns is organized in three sub-sections: in the first, the argument coordinated with an adjunct will be *what*; in the second, *who(m);* and in the third *to who(m)*. It is worth noting that, although the results regarding adjuncts coordinated with argument *what* and adjuncts coordinated with argument *who* will be presented separately below, the analysis was run conjointly (see Appendix for model coefficients).

3.4.2.1. What

After pruning the list of concordances, only 43 sentences are found to be relevant to the present study. The raw frequency of these mixed patterns is clearly lower than most of the patterns of adjunct coordination presented in the previous section. The distribution of these 43 examples according to the grammatical function of *what* (object, subject or complement of verb *to be*) is provided in table 1.

Complementary to table 1 is figure 3 below, which shows that the sentences where *what* is an object clearly outnumber the others: while object *what* is estimated to occur 81.7% [14.4%, 98.7%] of the time, the rates of subject and complement *what* are considerably lower, 8.8% [0.2%, 78.1%] and 6.1% [0.3%, 35.8%], respectively. Table 1 reveals that coordination with *how* is the most frequent. It can also be seen that there are no results for many of the combinations and, in those which are represented in the sample, no clear preference is observed for one of the two possible linearities.

Table 1. Raw frequency of the "*what* & adjunct" patterns.

WHAT	Grammatical function of *what*			
	SU	OB	COMP	Total
what and where	0	1	2	7
where and what	0	3	1	
what and how	2	14	0	30
how and what	2	12	0	
what and when	0	1	0	4
when and what	0	2	1	
what and why	1	1	0	2
why and what	0	0	0	
Total	5	34	4	43

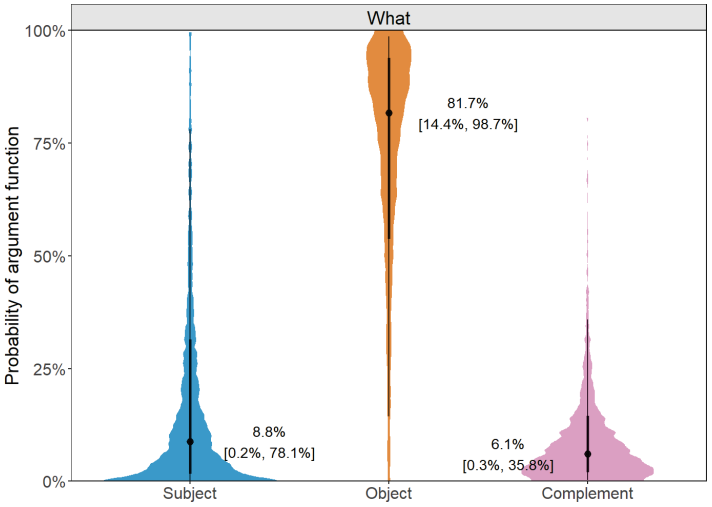


Figure 3. Model estimates for "*what* & adjunct" patterns.

In (24) to (26) some examples are provided of *what* in the three functions just mentioned.

(24) *what* = object

a. … , the difference in value between programs has also grown, making the choice of **what and where** to study more complex. (MAG 2017)

b. **What and how** do students learn? (ACAD 2019)

c. I tell him it is his to figure out **who and what and how** he loves, and what this family will be. (FIC 2018)

(25) *what* = subject

a. So all these lands are extremely important to understand **how and what** took place in the biblical text. (MOV 2017)

b. Getting specific about **how and what** influenced what is always difficult. (MAG 2018)

(26) *what* in copula sentence

What and where are environmental values? (ACAD 2018)

3.4.2.2. Who(m)

In principle, *who* can be a subject or an object. As in the case of *what*, the sentences headed by verb *to be* are counted separately. The results are shown in table 2 below and some examples are presented in (27) to (29).

(27) *who* in copula sentence

a. If we can't find out **who and where** they are, we can't stop all this. (TV 2017)

b. Things that I look back at now and, and frankly, I wonder **where and who** that person was because I don't know that person. (SPOK 2018)

(28) *who* = subject

a. We are back now with just a very small portion of the women who are survivors of team USA Gymnastics doctor, Larry Nassar, who was provided with unfettered access to young girls for thirty years by enablers around every corner. And the question is **how and who** will be held accountable? (SPOK 2018)

b. This accelerated growth rate of new on-demand services will ultimately lead to a Battle Royale to fulfill the local logistics of these apps. **How and who** is fulfilling the on-demand app services? (MAG 2016)

c. But the fact of the matter is, I'm awfully sophisticated about **why, how and who** built this country. (SPOK 2019)

(29) *who* = object

a. This is a family. This is **who and how** we love. (FIC 2018)

b. No matter **who and how** the council selects, observers say Saudi Arabia is in for a prolonged "period of succession" as the torch is passed from one generation to the next - one that couldn't come at a worse time. (NEWS 2015)

Table 2. Raw frequency of the pattern "*who* & adjunct".

WHO	Grammatical function of *who*			
	SU	OB	COMP	
who and where	0	0	**6**	7
where and who	0	0	**1**	
who and how	0	**2**	0	5
how and who	**3**	0	0	
who and when	0	0	0	0
when and who	0	0	0	
who and why	0	0	0	0
why and who	0	0	0	
Total	**3**	**2**	**7**	**12**

With only 12 sentences, the raw frequency of the patterns in which *who* is coordinated with an adjunct is sparse in the period under study.[7] 7 of those 12 sentences (58%) are headed by the copula. In the 3 cases where *who* is a subject, it is the second *wh*-phrase in the coordination. In the whole corpus, however, there are some examples of *who* as first conjunct. In the two cases, where *who* is an object, shown in (29) above, it is the first *wh*-phrase. The lack of clear-cut patterns is evidenced by model estimates too, shown in figure 4. Even though there seems to be a tendency for argument *who* to be a complement of the verb *to be* when coordinated with an adjunct, the statistical results suggest that this trend is not compatible with arguing that *who* favors complement functions over subject or object functions.

[7] Things do not improve significantly when the whole corpus is considered. Of the 180 sentences found, 66 are combinations of *who and where*. Attending to the grammatical function of *who*, in 88 sentences (i.e., 49% of the total) *who* is in a sentence headed by the copula; in 52 cases (29%), it is a subject; and in 40 cases (22.2%), it is an object. As can be seen, the distribution is similar to that in the sample. Similarly, in the whole corpus the lowest frequency is found in the combinations of *who* with *when* and *why*, where the numbers are particularly low (*who and when*: 13; *when and who*: 9, *who and why*: 8; *why and who*: 5).

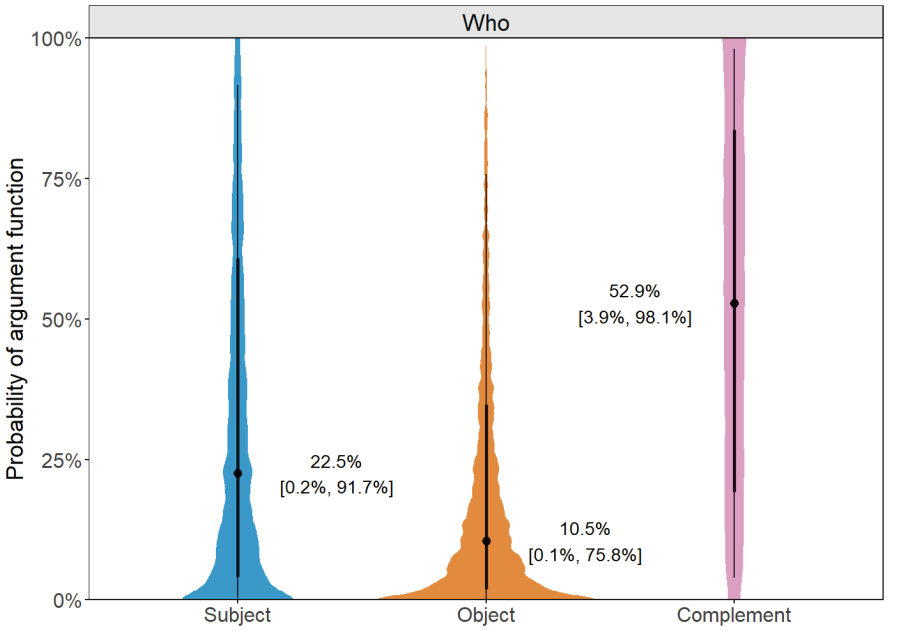


Figure 4. Model estimates for "*who* & adjunct" patterns

In the whole corpus there are also some cases in which object *who* is the second coordinated *wh*-phrase. We will come back to these ordering issues in the discussion of the results below.

The corpus was also searched for the combination of *whom* with the different adjuncts. There were only 3 relevant sentences, shown in (30):[8]

(30) a. But Claire was an indoor-woman, born in Boston, not Vermont, only here because she'd made an indoor-woman mistake in **whom and how** she loved, biting off so much more than she could chew that … (FIC 2019)

b. **How and whom** should I forgive? (ACAD 2015)

c. … with the artificially intelligent brain of Watson, a Watson who can choose **when and whom** to kill. (MOV 2016)

[8] In the whole corpus there are only 10 sentences, distributed as follows: 2 *whom and where*, 1 *where and whom*, 1 wh*om and how*, 3 *how and whom*, 2 *when and whom*, and 1 *why and whom*.

3.4.2.3. To who(m)

Occurrences of coordination of the prepositional phrase *to who(m)* with an adjunct are also very rare, with only one sentence in the period under study, (31) below, and 36 in the whole corpus.[9]

(31) The next crown prince will be a passing of the torch to the next generation – **to whom and how** that torch is passed is the real controversy. (NEWS 2015)

Table 3 recapitulates the results of the search of mixed coordination patterns.

Table 3. Mixed coordination ("*what*, *who*, *whom*, *to whom* & adjunct").

RECAP	what		who		whom		to whom		Total
object & adjunct	34	+	2	+	3	+	1	=	40
subject & adjunct	5	+	3	+	--	+	--	=	8
comp & adjunct	4	+	7	+	--	+	--	=	11
Total	43	+	12	+	3	+	1	=	**59**

3.4.3. Coordination of two arguments

The pronominals conjoined in the argument-only pattern are *who* (nominative, subject), *who(m)* (accusative/dative, direct object/indirect object), *what* (nominative/accusative, subject/direct object), and *to who(m)* (indirect object). Basically, the patterns of interest would involve the coordination of a subject with an object (direct or indirect; *who and what*, *who and who(m)*) and the coordination of two objects (direct and indirect object *what/who(m) and (to) who(m)*).

The search for the string *who and what* threw 57 hits for the period under study. However, none of these concordance lines is relevant because the two *wh*-phrases either (i) have the same function, (32); (ii) belong in two different sentences, (33); (iii) are nominalized, (34); or (iv) are the remnants of sluicing, as in (35):

[9] Of those 36, 17 (almost 50%) are of the string *how and to whom*. The rest are distributed as follows: 3 *to whom and how*; 8 *when and to whom*; 7 *where and to whom*; 1 *to whom and where*. In the corpus, there is also a clear preference for *to whom* to appear as the second *wh*-phrase. The predicates heading the relevant sentences can be divided into two groups: (i) predicates expressing direction (*to be headed, direct,* and *lead*) and (ii) ditransitive predicates (*sell, loan, give (affection)* and *offer*). In the former case, both *where* and *to whom* are arguments that express the goal of the action, as in (i):
(i) […] we lose our best chance of finding out **where and to whom** this stuff is headed. (TV 2018)
They are, therefore, not considered in the final count of mixed patterns. It must be noted that all the sentences in which *where* is an argument (in any of the patterns under study) have been excluded from the final count.

(32) We've been talking a great deal in recent years about memory and memorials – about **who and what** deserves to be remembered. (SPOK 2018)

(33) Let's see who's **who and what**'s what like we did back in the day. (MOV 2018)

(34) Here's the **who and what** of the SAVOR festival, and don't forget that napkin. (NEWS 2019)

(35) Well, you can't always get what you want. – **Who and what**, Cece) – Lester Bangs returned your call and David Geffen screamed. (TV 2016)

Something similar happens with *what and who*, with 15 concordances, *whom and what* with 5 and *what and whom* with 1. None of these concordances is relevant for the reasons just mentioned. For the remaining combinations of *wh*-phrases (*what and to who(m), to who(m) and what, whom and who, whom and who*), there are 0 concordances for the period under study.

In the whole corpus there are four sentences that seem to involve the coordination of two arguments. They are shown in (36) below.

(36) a. Through the expansion process, dance educators may find ways to express that derived meaning in relation to **whom and what** they teach. (ACAD 1996)

b. For me, this type of change needs to be part of a browser, with as suggested before with a recognisable cookie database so people know **what and to whom** is being disclosed verified by a third party, not what the site says. (WEB 2012)

c. ??Rumors were widespread, however, regarding **what and to whom** promises had been made in order to create the Bell Helicopter deal in the first place. (ACAD 1998)

d. Check bill: A good way to keep tabs on what your child is doing with their phone and to ensure that they are following your rules, is to check the cell phone bill each month. You are able to see what time, **to whom and what** they bought on the bill. (BLOG 2012)

We will come back to these sentences in the following section.

3.5. Discussion

Table 4 displays a summary of the raw frequency of the different patterns of CWHs found in the corpus for the period 2015-2019.

Table 4. Summary of raw frequencies across patterns (2015-2019).

PATTERN	raw frequency	%
adjunct & adjunct	727	92.49
mixed	59	7.51
argument & argument	0	0
Total	786	100

The raw results are in line with model estimates, shown in figure 5 (also see Appendix for model coefficients), in that "adjunct & adjunct" combinations are remarkably more frequent (92.5% [90.6%, 94.1%]) than mixed adjunct and argument combinations (7.5% [5.9%, 9.4%]). This shows clearly why English is described as an adjunct language. The argument patterns are completely absent in the period under study. However, as seen in the previous section, the latter patterns are not inexistent in the corpus, although the numbers appear to be anecdotal.

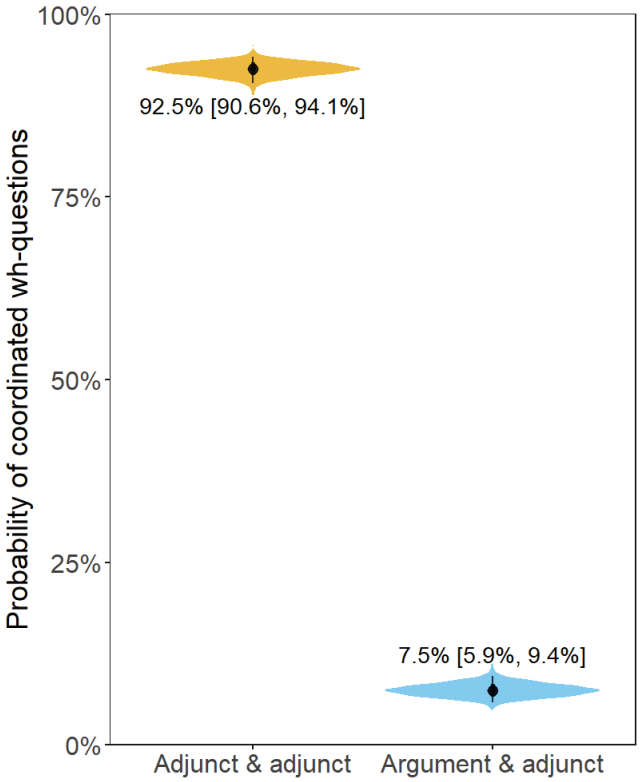


Figure 5. Model estimates for the whole dataset.

3.5.1. Adjunct and adjunct

Even though English is classified as an adjunct or mixed language (see section 2.1 above) which is supposed to freely allow for the coordination of two adjuncts, the corpus search reveals important differences in the raw frequencies of the different combinations of *wh*-adjuncts. In some cases (notably the coordination of *why* with *where*), they are lower than the frequencies of some mixed patterns. This may be due to pragmatic reasons. It may simply be less common to ask for the reason why and the place where something happens than to ask for the moment and the place where it happens, for instance.

Another aspect which is worth mentioning is the apparent preference for one of the linearities over the other in all the combinations of *wh*-adjuncts. In principle, it seems dubious that the reasons for this preference will be found in the grammar. It seems more plausible that they are related to the communicative purpose in the context of use.

Given that the adjunct pattern is widely accepted by native speakers, we will not dwell on it, but rather turn to the more controversial mixed and argument patterns.

3.5.2. Mixed coordination

The mixed patterns of coordination have been shown to be by far less frequent than the adjunct patterns. One of the reasons for this may be that, in the case of the latter, there are 6 possible combinations of *wh*-phrases (with 2 linearities each) as opposed to the 4 existing in the case of each mixed coordination. Another reason might be in the fact that many mixed patterns require the sentence to be headed by a specific type of predicate (optionally transitive in the case of the pattern "object and adjunct"), as will be seen directly. It is likely that these two factors are not enough to explain the huge difference in the frequencies of the two patterns (727 v. 59). The reluctance of some speakers to accept these patterns is a more plausible explanation for the scarcity of mixed coordination in the corpus. However, the question remains as to where this reluctance originates.

3.5.2.1. *What* and adjunct

As seen in section 3.4.2.1 above, when *what* appears in a mixed CWH, in 79% of the cases (34 out of 43) it is an object. In these cases, the transitive predicates that appear in the CWHs are *eat, teach, analyze, measure, learn, tell, love, improve, study, buy, cut, prepare, write, see, think, prune,* and *investigate*. What all these predicates have in common is that they are optionally transitive (i.e., they can appear

with or without a direct object).[10] Consider the sentences in (37) to (39) below, headed by some of these predicates:

(37) a. The value of a college degree varies greatly based on **where and what** you study, but … (MAG 2017)

b. [where you study] and [what you study]

(38) a. **What and how** do students learn? (ACAD 2019)

b. [what do students learn] and [how do students learn]

(39) a. …, and I'm always watching to see **when and what** he prunes and when he plants what and where. (FIC 2018)

b. [when he prunes] and [what he prunes]

Notice that, in all the sentences above, the two questions being asked, shown in (b), are well-formed and morpho-syntactically parallel, though not identical, as one of the sentences lacks a direct object. As discussed in Gračanin-Yuksek, the two *wh*-questions also differ in terms of interpretation ("What and Why" 13). The questions with the *wh*-adjunct have what has been called an *at-all* reading, as opposed to the *it*-reading of the question with *what*. Thus, in a sentence like (37a), the first conjoined question would be interpreted as asking *where you study at all*, rather than *where you study what you study*. In other words, the answer to the *what* question (say, linguistics) is not interpreted as the direct object of *study* in the *where*-clause (*where you study linguistics*). Importantly, the fact that this interpretation cannot arise already points towards a biclausal analysis. In other words, if the two *wh*-phrases had their base position in one and the same sentence, this is precisely the interpretation that would be expected to arise.

The question, then, is why sentences like these are not uniformly accepted by all speakers. The answer may lie in the lack of morpho-syntactic and interpretive symmetry between the two conjoined clauses. The clauses differ, at least, at the VP level: one is headed by a transitive predicate, while the other is headed by its intransitive counterpart. Notably, this kind of asymmetry does not occur in CWHs

[10] We will use the term *optionally transitive* to refer to predicates whose objects can be left unexpressed without resulting in ungrammaticality, and *obligatorily transitive* for those that require an explicit object. *Optionally transitive* predicates have been the focus of intense debate across various grammatical frameworks. Key discussions center on how these predicates should be classified and represented syntactically, particularly when the object is not expressed (see Allerton (59); Faulhaber (8-10) within valency theory; Bhatt and Pancheva for an overview from a generative perspective; see also Landau). In the context of this paper, the primary focus is not on how these predicates should be analyzed or classified within theoretical frameworks. Rather, what matters most is whether the absence of an explicit object leads to ungrammaticality. This distinction–between optionality and obligatoriness–will be the main criterion for analysis.

following the "adjunct and adjunct" pattern, which speakers tend to accept more consistently.

The analyses proposed so far would face challenges in deriving mixed questions precisely because they rely heavily on constituent identity. This issue is particularly acute for ATB movement-based analyses, which assume strict syntactic parallelism across conjuncts. Ellipsis-based analyses may offer a more flexible alternative, as long as they can relax the identity condition–traditionally understood as requiring syntactic isomorphism, but which some have argued can be satisfied under looser, semantic, or featural correspondences. See, for instance, the discussion in Merchant (108-146). Under such approaches, mismatches that would be problematic for ATB movement may be more easily accommodated within ellipsis.

In the sample, there are instances of CWHs where the first conjoined clause is not well-formed. This happens in two types of situations: (i) when *what* functions as a subject, as seen in examples (40) and (41) below; and (ii) when *what* functions as the object of an obligatorily transitive predicate, as illustrated in examples (44) and (45).

In the first type, the ungrammaticality arises because the first conjunct lacks a subject, as in (40b) and (41b):

(40) a. So all these lands are extremely important to understand **how and what** took place in the biblical text. (MOV 2017)

b. *[how took place in the biblical text] and [what took place in the biblical text]

(41) a. Getting specific about **how and what** influenced what is always difficult. (MAG 2018)

b. *[how influenced what] and [what influenced what]

What may be surprising is that, given the ungrammaticality of one of the conjuncts, these CWHs are not uniformly rejected as ill-formed. A possible mitigating factor is the fact that the obligatory *wh*-phrase is positioned at the end of the sequence of interrogative constituents, ensuring its adjacency to the verb in the second conjunct. This observation is reinforced by the fact that reversing the order of the *wh*-phrases leads to outright deviant sentences, as seen in (42):

(42) a. *So all these lands are extremely important to understand **what and how** took place in the biblical text.

b. *Getting specific about **what and how** influenced what is always difficult.

Interestingly, some sentences in the corpus reveal what appears to be a strategy to guarantee the grammaticality of the second conjunct which consists of supplying an overt subject, as in (43):

(43) a. There should be selection criteria which help to justify **what and why** *a collection* is being digitalized. (ACAD 2018)

b. [what is being digitalized] and [why *(a collection) is being digitalized]

Here, the indefinite *a collection* is necessary to satisfy the subject requirement in English, further supporting a biclausal analysis of CWHs. We will come back to subjects in CWHs in section 3.5.2.2 below.

In the second type of situation, where *what* serves as the object of an obligatorily transitive predicate, the question headed by the *wh*-adjunct lacks a required object, which leads to its ill-formedness, as in (44) and (45):

(44) a. Take some time and explore your "why" as it might give you clear direction on **how and what** to do next.

b. *[how to do next] and [what to do next]

(45) a. We have not yet obtained the SF-86s. So we're really not in a position to evaluate **what and when** he disclosed different things. (SPOK 2017)

b. [what he disclosed (*different things)] and [when he disclosed *(different things)]

When *what* is the first *wh*-phrase in the sequence, as in (45), speakers often insert a nominal in the second conjunct to guarantee its grammaticality, typically an indefinite or generic NP, such as *different things* in the example. It is also possible to find specific NPs, in which case *what* refers to a subset of the entity denoted by the supplied NP, *the contents of current courses* in (46):

(46) a. Concept map assessment can be used to provide information about **what and how** to improve the contents of current courses within PETE programs. (ACAD 2015)

b. [what to improve] and [how to improve *(the contents of current courses …)]

In the examples discussed, there is clear tension between grammaticality and morphosyntactic identity or parallelism, which may underlie the hesitation or outright rejection by some speakers.

Finally, in a small number of cases (4 sentences in the sample), *what* appears as the complement of the verb *to be*, as shown in (47). In these instances, both conjuncts are well-formed and exhibit fully parallel syntactic structure, which appear to be factors that may enhance acceptability. For this reason, the relatively small number of sentences containing *what* with the verb *to be* may seem unexpected. As will be shown later, the number of instances of *who* in the same type of context is slightly higher, although still not substantial. This pattern would be worth testing

with native speakers, in order to determine whether the small number of such sentences is accidental or reflects a genuine restriction.

(47) a. … none of which assuaged her real hunger as she asked everyone who entered her room **when and what** the next meal would be. (FIC 2017)

b. [when the next meal would be] and [what the next meal would be]

All in all, the relative rarity of cases like those discussed above highlights the role of structural symmetry in determining acceptability. Sentences that exhibit ungrammaticality in either conjunct are notably infrequent, suggesting that they are generally rejected, even by speakers who may tolerate other types of mixed patterns.

3.5.2.2. *Who(m)* and adjunct

The low raw frequency of all patterns of coordination of *who* with an adjunct may be indicative of the reluctance of the speakers to use this construction. Of the 12 sentences found in the period under study, only in 2 is *who* an object. As in the case of *what,* the two sentences are headed by optionally transitive predicates: *love* and *select*, which results in the acceptability of the two questions being asked, as can be seen in (48) and (49):

(48) a. This is a family. This is **who and how** we love. (FIC 2018)

b. [who we love] and [how we love]

(49) a. No matter **who and how** the council selects, … (NEWS 2015)

b. [who the council selects] and [how the council selects]

To these sentences, 3 additional examples must be added where the object is *whom*. These sentences, example (30) above, are also headed by optionally transitive predicates (*love, forgive* and *kill*). They are repeated here for convenience as (50).

(50) a. But Claire was an indoor-woman, born in Boston, not Vermont, only here because she'd made an indoor-woman mistake in **whom and how** she loved, biting off so much more than she could chew that … (FIC 2019)

b. **How and whom** should I forgive? (ACAD 2015)

c. … with the artificially intelligent brain of Watson, a Watson who can choose **when and whom** to kill. (MOV 2016)

In all these sentences, both questions are grammatical, and the same structural asymmetries and interpretive differences (*at-all* reading in the clause with no object v. *it* reading in the clause with the object) between the two coordinated clauses are

observed, as previously illustrated with the sentences in (37) to (39) above, where the object was *what*.

Who is a subject only in the 3 sentences, as in (51):

(51) a. But the fact of the matter is, I'm awfully sophisticated about **why, how and** [**who** built this country]. (SPOK 2019)

b. And the question is **how and** [**who** will be held accountable]? (SPOK 2018)

c. **How and** [**who** is fulfilling the on-demand app services]? (MAG 2016)

These sentences are parallel to those in (40) and (41) discussed above, where the subject is *what*. As in those sentences, here the *wh*-subject is the last *wh*-phrase in the sequence, which results in the acceptability of the question in brackets and the CWH as a whole, in spite of the ungrammaticality of the other underlying questions (cf. **How built this country, or* **How will be held accountable*). This deviance may contribute to the scarcity of such sentences in the corpus, likely reflecting speakers' reluctance to use them.

The presence of subjects in mixed CWHs in English has not been seriously addressed in the literature. Browne provides the sentence in (52) to illustrate the impossibility of having arguments in CWHs (224).

(52) *I don't know **who and with what** broke the window.

More recently, Gračanin-Yuksek, based on the ungrammaticality of the sentences in (53), discards the possibility of having *wh*-subjects in CWHs in English (*About* 42).

(53) a. *Tell me **who and when** sang.

b. *Tell me **who and why** ate.

c. *Tell me **who and how** fixed the sink.

d. *Tell me **who and where** gave a talk.

Citko, however, notices that many native speakers find sentences like those in (54), drawn from the Internet, acceptable (304):

(54) a. **Who and where** are Today's Great Military Thinkers?

b. **Where and what** is your niche?

c. **Where and who** are the world's illiterates?

d. **Who (and when)** discovered that the earth's axis is on a 23 degree tilt?

e. **What and where** went wrong with Rudy's campaign?

She reports that one reviewer found a contrast between sentences (54a-c), on the one hand, and (54d-e), on the other. This reviewer finds the latter degraded but the former quite acceptable. Given the observations made previously in this section, this contrast is to be expected. The questions in (54a-c) are headed by verb *to be*, which will make the two questions being asked well-formed and morpho-syntactically symmetric, two important factors in the acceptability of CWHs. Recall that, in the corpus, there are also 7 sentences headed by verb *to be*, which are similar to (54a-c). (55) shows the two examples shown in (27) above.

(55) a. If we can't find out **who and where** they are, we can't stop all this. (TV 2017)

b. Things that I look back at now and, and frankly, I wonder **where and who** that person was because I don't know that person. (SPOK 2018)

The situation is rather different in (54d-e), where the question headed by the adjunct is ungrammatical, as in (56), which leads to the degradation perceived by the reviewer of Citko's article. Once again, it seems that the position of the *wh*-subject relative to the other *wh*-phrases has an impact on the perception of grammaticality.

(56) a. *When discovered that the earth's axis is on a 23 degree tilt?

b. *Where went wrong with Rudy's campaign?

In the corpus, although not within the period under study, the following sentences feature *who* as the first conjoined *wh*-phrase.

(57) a. **Who and** [**how** can save Detroit from this decay]? ((BLOG 2012)

b. … who pays his salary, what is the job description, and **who and** [**how** is he held accountable]? (BLOG 2012)

c. You're first link explains both **who and** [**how** they were picked]. (WEB 2012)

In (57a), similar to Citko's examples, the question in brackets is ungrammatical due to the absence of a subject. In this sentence (and in the previous examples from Citko) the conjunction and the second *wh*-phrase may be intended as parenthetical elements (see the parentheses around *and when* in (54d)). However, even with a prosodic pause, it seems difficult to separate the interrogative adverb from the verb clearly enough for the listener to interpret *who* as the subject. As a result, the construction may be perceived as ungrammatical. A prosodic analysis would be necessary to evaluate this possibility.

In contrast, in (57b) and (57c), the issue arising from the lack of a subject is resolved by the addition of a pronoun (*he* and *they*, respectively), which correlates with *who*. This approach aligns with the strategy described for obligatorily transitive

predicates and subject *what* in section 3.5.2.1 above. However, the insertion of this pronoun disrupts the morpho-syntactic identity of the two questions being asked, potentially resulting in the rejection of these sentences by some speakers. It may be worth measuring speakers' tolerance for this type of disruption to parallelism through specifically designed questionnaires.

3.5.2.3. *To who(m)*

In the years under study, only one sentence was found–(31), repeated here as (58)–where this prepositional argument is coordinated with a *wh*-adjunct.

(58) The next crown prince will be a passing of the torch to the next generation – **to whom and how** that torch is passed is the real controversy. (NEWS, 2015)

The prepositional phrase, which expresses the goal of the act of passing, is optional in the context of this sentence, as witnessed in the fact that the two questions being asked are well-formed, as in (59) below, even though this argument appears only in the first conjunct. This example confirms that the optionality of the argument is a key factor in determining whether it is permitted in CWHs.

(59)[to whom that torch is passed] and [how that torch is passed]

3.5.3. Coordination of two arguments

The coordination of two arguments in English (and other so-called mixed languages) is not really addressed in the literature. Examples like those in (60) are usually discarded as ungrammatical due to the ungrammaticality of the two questions being asked, and the issue is not pursued any further. See, for instance, Lipták (155).

(60) a. ***What and to whom** did they give?

b. *[What did they give] and *[to whom did they give]

As mentioned in section 3.4.3, no cases of coordination of two arguments have been found in the period under study. Moreover, expanding the search to the entire corpus yields only one sentence, (61), where a direct object is coordinated with an indirect object.

(61) Through the expansion process, dance educators may find ways to express that derived meaning in relation to **whom and what** they teach. (ACAD 1996)

It must be noted that the coordination of two arguments would basically involve the coordination of a subject with an object (direct or indirect) or of two objects (as in the example above). If this is taken together with the observation that optionality plays a key role in determining the acceptability of the different patterns of coordination, two explanations for the rarity of occurrences emerge: (i) the obligatoriness of the subject in English, and (ii) the fact that objects can only appear in the construction when the sentence is headed by an optionally transitive predicate, which represents a limited subset of all verbs in the language.

In the case of coordination of two objects, the predicate must be ditransitive (a relatively small subset of all predicates) and the two complements must be optional in the specific context of use of the construction. The sentence found in the corpus is headed by predicate *teach*, which can be used transitively or intransitively. Moreover, the indirect object is also optional, as can be seen in the fact that the two conjuncts in (62) are well-formed.

(62) … in relation to [whom they teach] and [what they teach]

A single occurrence in the corpus is hardly a solid basis to assert that CWHs with two conjoined internal arguments are permitted in English. These findings clearly show that speakers do not favor the coordination of two arguments. However, the question remains as to whether this pattern is genuinely prohibited by the grammar or whether its virtual absence from the corpus is due to the factors mentioned earlier. Clearly, further insights into this specific pattern could be gained through acceptability judgment experiments, something we leave for future research.

The following sentence, drawn from Windeatt (61), provides evidence that discourse context contributes to the felicitousness of CWHs. In (63), the arguments *what* and *to whom*–typically obligatory with the predicate *give*–appear in coordination with several *wh*-adjuncts. In the context of this sentence, however, both arguments are optional, as shown by the fact that the first instance of the predicate (underlined) occurs without any explicit internal argument.

(63) But always when she gave, she arranged her gifts wisely, and certainly knew **what, when, how, to whom and why** she gave.

It is plausible that, when presented with an isolated sentence containing a predicate like *give* (as in (60) above), speakers interpret its internal arguments as obligatory, leading them to judge the corresponding CWH as ungrammatical. However, when the same sentence is embedded in a discourse context where arguments are optional, higher rates of acceptance might be expected. This issue should be explicitly tested through experimental research.

4. CONCLUSION

The results of this corpus study support the characterization of English as (primarily) an adjunct language, as the "adjunct & adjunct" patterns significantly outnumber the mixed and argument patterns. In the corpus data from 2015-2019, 92.5% of the examples were estimated to involve the coordination of adverbial *wh*-phrases, while 7.5% were instances of mixed coordination, and no cases of argument coordination were found. Further examination of the entire corpus confirmed the consistency of these findings. The lack of argument coordination in both the 2015-2019 data (0 cases) and the broader corpus (1 case) supports the claim that this type of coordination is typically rejected.

In the analysis of the data, it was observed that all the arguments found in the different patterns of coordination, whether mixed or argument coordination, were consistently optional. This restriction suggests that the coordination of adjuncts with obligatory arguments (i.e., subjects and the objects of certain predicates) is systematically avoided, reinforcing the idea that English and other mixed languages impose specific constraints on how adjuncts and arguments can combine within this construction.

Similarly, the required optionality of the argument is the most probable reason behind the absence of argument patterns from the corpus. These patterns depend on highly specific contexts of use, making them less common in natural language and, consequently, more difficult to capture in the corpus.

Concerning the analysis of CWHs, the data analyzed in this study point at a biclausal underlier for this construction in English. Thus, some of the examples discussed involving the coordination of at least one argument show clearly that the two conjoined *wh*-phrases cannot originate in the same clause. The very reluctance to accept sentences with obligatory arguments can be taken to point in the same direction. However, no compelling evidence was found to support any of the specific biclausal analyses suggested in previous research.

Beyond simply confirming previous observations, our findings provide new empirical precision regarding the rarity of argument coordination and the consistent optionality of arguments when they appear. Given the dependency on context observed, future experimental acceptability studies should base their analyses on contextually grounded data, rather than on isolated sentences, to ensure ecological validity. Such studies could complement the corpus findings, particularly by probing constructions that may be theoretically possible but rarely attested, and by examining the degree of parallelism between conjuncts, a factor that may ultimately bear on the tenability of current analyses. Future research could also extend this corpus-based approach to other varieties of English in order to test the generality of the restrictions observed here.

ACKNOWLEDGEMENTS

The preliminary results of this study were presented at the 10th Biennial International Conference on the Linguistics of Contemporary English (BICLCE), held at the University of Alicante (Spain) from 26–28 September 2024. We would like to thank the audience for their valuable comments. This research was partially supported by Gobierno Vasco/Eusko Jaurlaritza (IT1537-22 (HiTT / Hizkuntzalaritza Teorikorako Taldea)) and Ministerio de Ciencia e Innovación del Gobierno de España (PID2021-128511NB-I00).

REFERENCES

Allerton, David J. *Valency and the English Verb*. Academic Press, 1982.

Bhatt, Rajesh, and Roumyana Pancheva. "Implicit Arguments." In *The Blackwell Companion to Syntax*, edited by Martin Everaert and Henk van Riemsdijk, vol. 2, pp. 558-588. Blackwell Publishing, 2006.

Browne, Wayles. "Conjoined Questions and the Limitation on English Surface Structure." *Linguistic Inquiry*, vol. 3, no. 2, 1972, pp. 223-226. *JSTOR*, http://www.jstor.org/stable/4177703.

Bürkner, Paul-Christian. "*brms: An R Package for Bayesian Multilevel Models Using Stan*." *Journal of Statistical Software*, vol. 80, no. 1, 2017, pp. 1-28. https://doi.org/10.18637/jss.v080.i01.

Chomsky, Noam. *Lectures on Government and Binding: The Pisa Lectures*. 7th ed., Foris Publications, 1981.

Chomsky, Noam. *The Minimalist Program*. MIT Press, 1995.

Citko, Barbara. "The Puzzles of Wh-Questions with Coordinated Wh-Pronouns." *Challenges to Linearization*, edited by Theresa Biberauer and Ian Roberts, Mouton de Gruyter, 2013, pp. 295-330, https://doi.org/10.1515/97816 14512431.295

Citko, Barbara, and Martina Gračanin-Yuksek. "Multiple (Coordinated) (Free) Relatives." *Natural Language and Linguistic Theory*, vol. 34, 2016, pp. 393-427, https://doi.org/10.1007/s11049-015-9306-8

Citko, Barbara, and Martina Gračanin-Yuksek. "Towards a New Typology of Coordinated Wh-Questions." *Journal of Linguistics*, vol. 49, no. 1, 2013, pp. 1-32, https://doi.org/10.1017/S0022226712000175

Comorovski, Ileana. *Interrogative Phrases and the Syntax–Semantics Interface*. Dordrecht: Kluwer, 1996.

Davies, Mark. *The Corpus of Contemporary American English (COCA)*. 2008-present, www.english-corpora.org/coca/

Faulhaber, Susen. *Verb Valency Patterns: A Challenge for Semantics-Based Accounts*. De Gruyter Mouton, 2011. https://doi.org/10.1515/9783110240788

Giannakidou, Anastasia, and Jason Merchant. "Reverse Sluicing in English and Greek." *The Linguistic Review*, vol. 15, no. 2-3, 1998, pp. 233-256, https://doi.org/10.1515/tlir.1998.15.2-3.233

Gračanin-Yuksek, Martina. *About Sharing*. 2007. Massachusetts Institute of Technology, PhD thesis. http://hdl.handle.net/1721.1/39343

Gračanin-Yuksek, Martina. "What and Why Can't Be Shared." April 2009. Middle East Technical University, http://users.metu.edu.tr/martina/what_and_why_Apr_2009_single_spaced_submitted_revised_FINAL_web.pdf

Gračanin-Yuksek, Martina. "Conjoined Wh-questions." *The Wiley Blackwell Companion to Syntax*, edited by Martin Everaert and Henk van Riemsdijk, 2nd ed., Hoboken: Wiley-Blackwell, 2017, 1127-1161, https://doi.org/10.1002/9781118358733.wbsyncom015

Gribanova, Vera. "Structural Adjacency and the Typology of Interrogative Interpretations." *Linguistic Inquiry*, vol. 40, no. 1, 2009, pp. 133-154, https://doi.org/10.1162/ling.2009.40.1.133

Grimshaw, Jane. "On the Syntax of Conjoined Wh-Words in English Questions." *University of Massachusetts Occasional Papers in Linguistics*, vol. 3, no. 1, 1978, https://scholarworks.umass.edu/umop/vol3/iss1/2

Haida, Andreas, and Sophie Repp. "Monoclausal Question Word Coordinations Across Languages." *Proceedings of the North East Linguistics Society (NELS)*, vol. 39, edited by Suzi Lima, Kevin Mullin, and Brian Smith, Amherst, MA: GLSA, 2011, pp. 373-386. https://api.semanticscholar.org/CorpusID:221804053

Helbig, Gerhard. *Probleme der Valenz- und Kasustheorie*. 3rd ed., Niemeyer, 1992.

Herbst, Thomas, and Susen Schüller. *Introduction to Syntactic Analysis. A Valency Approach*. Tübingen: Gunter Narr Verlag, 2008.

Hole, Daniel. "Arguments and Adjuncts." *Syntax - Theory and Analysis: An International Handbook*, vol. 2, edited by Tibor Kiss and Artemis Alexiadou, De Gruyter Mouton, 2015, pp. 1284-1320. http://doi.org/10.1515/9783110363708

Kasai, Hironobu. "Coordinated Wh-questions in Japanese." *Lingua*, vol. 183, 2016, pp. 126-148. https://doi.org/10.1016/j.lingua.2016.05.009

Kazenin, Konstantin I. "On Coordination of Wh-phrases in Russian." Tübingen University and Moscow State University, 2002.

Landau, Idan. "The Explicit Syntax of Implicit Arguments." *Linguistic Inquiry*, vol. 41, no. 3, 2010, pp. 357-388.

Lewis, Shevaun, Dave Kush, and Bradley Larson. "What and When Can You Fill a Gap With Something?" *Paper presented at the 25th CUNY Human Sentence Processing Conference*, CUNY, 2012.

Lipták, Anikó. "Strategies of *Wh*-Coordination." *Linguistic Variation*, vol. 11, no. 2, 2012, pp. 149-188. *Linguistic Variation*, https://hdl.handle.net/1887/61419

Merchant, Jason. *The Syntax of Silence*. Oxford University Press, 2001.

Nunes, Jairo. "Sideward Movement." *Linguistic Inquiry*, vol. 32, no. 2, 2001, pp. 303-344. https://doi.org/10.1162/00243890152001780

Park, Myung-Kwan. "The Syntax of Conjoined Wh-Elements in English: A Preliminary Study." *Studies in Generative Grammar*, vol. 16, no. 3, 2006, pp. 555-565.

Potter, David, and Michael Frazier. "English Wh&Wh Constructions: Conjoin and Move." *Proceedings of the 38th West Coast Conference on Formal Linguistics*, edited by Rachel Soo, Una Y. Chow, and Sander Nederveen, Cascadilla Proceedings Project, 2021, pp. 355-365. https://www.lingref.com/cpp/wccfl/38/paper3580.pdf

Raţiu, Dafina. "A Multidominance Account for Conjoined Questions in Romanian." *Romance Linguistics 2010*, edited by Julia Herschensohn, Amsterdam: John Benjamins, 2011, pp. 257-270, https://doi.org/10.1075/cilt.318.16rat

R Core Team. *R: A Language and Environment for Statistical Computing*. Manual, R Foundation for Statistical Computing, 2025.

Rudin, Catherine. "On Multiple Questions and Multiple Wh-Fronting." *Natural Language and Linguistic Theory*, vol. 6, 1988, pp. 445-501. https://doi.org/10.1007/BF00134489

Skrabalova, Hana. "Parataxe Apparente et Coordination des Interrogatifs en Tchèque" [Apparent Parataxis and Coordination of Interrogatives in Czech]. *Coordination et Subordination: Typologie et Modélisation* [*Coordination and Subordination: Typology and Modelling*], edited by Isabelle Bril and Georges Rebuschi, Éditions Ophrys, pp. 231-242. https://hal.science/hal-00599016v1

Stan Development Team. *RStan: The R Interface to Stan*. 2024, https://mc-stan.org/.

Tesnière, Lucien. *Éléments de syntaxe structurale*. Klincksieck, 1959.

Tomaszewicz, Barbara. "Against Spurious Coordination in Multiple Wh-Questions." *Proceedings of the WCCFL 28*, edited by Mary Byram Washburn, Katherine McKinney-Bock, Erika Varis, Ann Sawyer, and Barbara Tomaszewicz, Cascadilla Proceedings Project, 2011, pp. 186-195. https://www.lingref.com/cpp/wccfl/28/paper2451.pdf

Tomaszewicz, Barbara. "Wh & Wh: Syntactic and Semantic Arguments for Clausal Coordination." *Formal Approaches to Slavic Linguistics 19*, edited by John Frederick Bailyn, Ewan Dunbar, Yakov Konrod, and Chris LaTerza, Michigan Slavic Publications, 2012, Ann Arbor.

Wakefield, Jon. *Bayesian and Frequentist Regression Methods*. Springer, 2013. https://doi.org/10.1007/978-1-4419-0925-1

Whitman, Neil. *Category Neutrality: A Type-Logical Investigation*. The Ohio State University, 2002, PhD thesis, *OhioLINK,* https://doi.org/10.4324/9780203495384

Whitman, Neil. "What and How We Can Learn from Mixed-Wh Interrogatives." *Manuscript*, The Ohio State University, 2014.

Windeatt, Barry A., editor. *Chaucer's Dream Poetry: Sources and Analogues*. Boydell & Brewer Ltd, 1982.

Winter, Bodo. *Statistics for Linguists: An Introduction Using R*. 1st ed., Routledge, 2019. https://doi.org/10.4324/9781315165547

Zhang, Nina N. "The Syntactic Derivations of Two Paired Dependency Constructions." *Lingua*, vol. 117, 2007, pp. 2134-2158. *Lingua*, https://doi.org/10.1016/j.lingua.2007.02.001

APPENDIX

Whole dataset model coefficients

Family: categorical

Links: mumixed = logit

Formula: pattern ~ 1

Data: cwh (Number of observations: 786)

Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1; total post-warmup draws = 4000

Regression Coefficients:

Term	Estimates	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
mumixed_Intercept	-2.52	0.13	-2.78	-2.27	1.00	1560	2085

Draws were sampled using sample(hmc). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

Adjunct model coefficients

Family: categorical

Links: muhowwhere = logit; muhowwhy = logit; muwhenhow = logit; muwhenwhere = logit; muwhenwhy = logit; muwherehow = logit; muwherewhen = logit; muwherewhy = logit; muwhyhow = logit; muwhywhen = logit

Formula: combo ~ 1

Data: adjuncts (Number of observations: 727)

Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1; total post-warmup draws = 4000

Regression Coefficients

Term	Estimates	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
muhowwhere_Intercept	-0.90	0.19	-1.27	-0.53	1.00	3219	3204
muhowwhy_Intercept	0.57	0.12	0.33	0.82	1.00	1946	2509
muwhenhow_Intercept	-0.05	0.15	-0.34	0.24	1.00	2112	2631
muwhenwhere_Intercept	0.10	0.14	-0.17	0.37	1.00	2247	2844
muwhenwhy_Intercept	-1.41	0.23	-1.85	-0.96	1.00	3654	2924
muwherehow_Intercept	-0.60	0.17	-0.93	-0.26	1.00	2620	2460
muwherewhen_Intercept	-0.44	0.16	-0.77	-0.13	1.00	2810	2792
muwherewhy_Intercept	-2.99	0.46	-4.00	-2.18	1.00	4667	2684
muwhyhow_Intercept	-0.47	0.16	-0.79	-0.16	1.00	2746	2590
muwhywhen_Intercept	-2.65	0.41	-3.50	-1.93	1.00	5247	2786

Draws were sampled using sample(hmc). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

Adjunct & argument model coefficients

Family: categorical

Links: muobject = logit; musubject = logit

Formula: gramm_function ~ 1 + argument_word + (1 | wh_item)

Data: mixed (Number of observations: 55)

Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1; total post-warmup draws = 4000

Multilevel Hyperparameters

~wh_item (Number of levels: 5)

Term	Estimates	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(muobject_Intercept)	2.62	1.31	0.84	5.86	1.00	1394	1640
sd(musubject_Intercept)	3.79	1.92	1.16	8.46	1.00	1502	1580

Regression Coefficients

Term	Estimates	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
muobject_Intercept	2.54	1.35	-0.00	5.51	1.00	1478	1412
musubject_Intercept	0.27	1.72	-3.35	3.61	1.00	1819	1561
muobject_argument_wordwho	-4.49	1.93	-8.61	-1.27	1.00	1483	780
musubject_argument_wordwho	-1.25	2.06	-5.73	2.25	1.00	1403	803

Draws were sampled using sample(hmc). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).