


**Please cite the Published Version**

Curry, Niall  (2023) Question illocutionary force indicating devices in academic writing: A corpus-pragmatic and contrastive approach to identifying and analysing direct and indirect questions in English, French, and Spanish. *International Journal of Corpus Linguistics*, 28 (1). pp. 91-119. ISSN 1384-6655

**DOI:** <https://doi.org/10.1075/ijcl.20065.cur>

**Publisher:** John Benjamins Publishing

**Version:** Accepted Version

**Downloaded from:** <https://e-space.mmu.ac.uk/632116/>

**Additional Information:** This is an Author Accepted Manuscript of an article published in *International Journal of Corpus Linguistics*, by John Benjamins Publishing. The Version of Record can be accessed on the publisher's platform: <https://www.jbe-platform.com/content/journals/10.1075/ijcl.20065.cur> and via <http://doi.org/10.1075/ijcl.20065.cur>.

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

# **Question illocutionary force indicating devices in academic writing**

A corpus-pragmatic and contrastive approach to identifying and analysing direct and indirect questions in English, French, and Spanish

Niall Curry

Coventry University

Corpus research on questions as reader engagement markers in academic writing typically focuses on direct questions. Such questions are signalled by question marks and are relatively easily searchable in a corpus. However, indirect questions can be more challenging to identify, as they can be introduced by a range of forms. Based on a contrastive analysis of a corpus of English, French, and Spanish economics research articles, this paper provides pertinent evidence on direct and indirect questions as reader engagement markers. Firstly, it shows that direct and indirect questions as reader engagement markers are a rhetorical and generic feature of academic writing in the economics research article and, secondly, it presents a comprehensive list of indirect question illocutionary force indicating devices, valuable for future studies of indirect questions. Methodologically, this paper illustrates a replicable process for functional analysis and discusses the value of theoretically merging corpus and contrastive linguistic approaches.

**Key words:** illocutionary force indicating device, function-to-form corpus analysis, questions, reader engagement, corpus-based contrastive analysis

## **1. Introduction**

Questions as markers of reader engagement in academic writing have been an area of interest to academic discourse analysts since the early 1990s (Swales & Feak, 1994; Webber, 1994). Research in this area has largely focused on analyses of the functions question perform in a text. This has included the identification of question use in academic texts to create affect (Webber, 1994); get attention; frame the discourse; organise the text; create a niche; express an attitude and counter-claim; set up claims; and ask real questions (Hyland, 2002); increase the visibility of research (Ball, 2009); and make interesting titles (Cook & Plourde, 2016), for example. These studies have typically focused on English, with a smaller collection of studies focusing on other languages, such as Spanish (Soler, 2009), or contrastive studies of English

and French (Curry & Chambers, 2017) and English and Spanish (Curry, 2021; Lafuente-Millán, 2014).

While the continued study of questions offers many avenues for research in multilingual and cross-disciplinary contexts, an important limitation of contemporary corpus linguistic research on questions has been the almost exclusive focus on direct questions, i.e. questions signalled by the use of a question mark. This has important implications for how academic languages are perceived, as linguistic descriptions of academic languages as reader- or writer-responsible (Salager-Meyer, 2011) are typically based on their use of metadiscourse markers, like questions, in academic texts. For example, Lafuente-Millán's (2014) study shows that direct questions are used infrequently in Spanish. However, the study did not consider whether the Spanish corpus contained examples of indirect questions. Therefore, given French and Spanish academic languages' established preference for indirectness and negative politeness strategies (Loffler-Laurian, 1980; Clyne, 1994; Lafuente-Millán, 2014), the lack of consideration of indirect questions may be misrepresenting the propensity of these languages to use questions to engage readers.

The almost exclusive focus on direct questions in contemporary studies of questions in academic writing is easily understood. Questions marks, which act as “illocutionary force indicating devices” (IFIDs)<sup>1</sup> (Flöck & Geluykens, 2015: 7) for direct questions, are easily corpus-searchable. Indirect questions, however, are signalled by a wider variety of IFIDs, such as *ask* in Example (1) or *examine* in Example (2), both of which appear to combine with *whether* to create indirect question IFIDs.

- (1) Finally, we briefly characterize some stylized facts regarding our estimated stocks and *ask* whether there are trends in net foreign assets and shifts in their composition over time. (engecon02)<sup>2</sup>
- (2) Building on these predictions, we *examine* whether there is a negative relationship between the strength of FPRs and labor flows from the source country to the affiliates. (engecon04)

Indirect questions are more difficult to identify in a corpus, as their IFIDs, which function to raise questions, take a range of forms. As such, very few studies of questions in academic writing have included the analysis of indirect questions. Where they have, in Blagojević and Misić-Ilić (2012: 23) for example, the focus is usually formally restricted to “subordinate nominal clauses of *wh-* or *if/whether* type, preceded by clauses containing an illocutionary force marking device, such as a verb (*ask, wonder...*) or a noun (*question, answer...*)”. Therein

lies a key challenge for function-to-form corpus analyses, whereby, owing to form acting as the entry point in corpus linguistic analyses, searching for and finding forms to determine function and form relationships is methodologically circular. That is to say, if one searches a corpus for indirect questions composed of *wh*- complements preceded by verbs or nouns, for example, the search will find all of the indirect questions that are signalled by *wh*- complements preceded by verbs or nouns in the corpus. What it will not find, however, are any other forms or structures that may act as an IFID for indirect questions. To address this issue of methodological circularity, this paper proposes an approach that is built on theories of corpus and contrastive linguistics.

This paper presents a corpus-based contrastive analysis of the use of direct and indirect questions as reader engagement in economics research articles in English, French, and Spanish. The decision to focus specifically on the discipline of economics, first, derives from previous research on the KIAP corpus (discussed in Section 3) indicating that economics demonstrates interesting examples of metatext that are not always as prevalent in other disciplines (Dahl, 2004). Second, research on economics writing is quite limited in the wider literature on academic discourse, especially so in French and Spanish. Finally, to satisfy the *tertium comparationis* principle of contrastive linguistics (discussed in Section 2.2), the delimited focus on the discipline of economics and the genre of the research article reduces variability in the corpus data, rendering the corpora more comparable.

Focusing initially on a review of research on direct and indirect questions as reader engagement, this paper outlines the current state of the art of research on direct and indirect questions. This is followed by an overview of the literature on function-to-form corpus analyses and the relevance of contrastive linguistic theory therein. The corpus data and method for extracting direct and indirect questions is then presented, followed by the results of the corpus-based contrastive analysis of direct and indirect questions as reader engagement in English, French, and Spanish. This involves a detailed presentation of the methodological processes, and the results on the presence of direct and indirect questions and the use of indirect question IFIDs in each language. Finally, the paper closes with a discussion of direct and indirect questions as reader engagement markers in the English, French, and Spanish data, indirect question IFIDs in the English, French, and Spanish data, and emerging theoretical and methodological considerations for corpus and contrastive linguistics.

## **2. Questions as reader engagement markers: finding direct and indirect questions in function-to-form analyses**

This section offers an in-depth review of the research on questions and function-to-form analyses that have informed this study. First, in Section 2.1, questions as reader engagement devices are discussed with a view to outlining the current state of the art in question research and the gaps evident in the knowledge of indirect questions in academic writing. Second, in Section 2.2, function-to-form approaches in corpus linguistics are discussed, highlighting the theoretical synergies apparent in both corpus and contrastive linguistic approaches to the study of function.

### **2.1 Questions as reader engagement in English, French, and Spanish academic writing**

As a feature of reader engagement, questions are the “strategy of dialogic involvement par excellence, inviting engagement and bringing the interlocutor into an arena where they can be led to the writer’s viewpoint” (Hyland, 2002: 185). Broadly, questions serve to position readers in texts and involve them in the social construction of knowledge that takes place in academic discourse. For academic writers, engagement markers – including questions – have become integral to their practices for anticipating reader reactions (Lafuente-Millán, 2014), addressing readers directly and positioning them discursively for persuasive ends (Tse & Hyland, 2006). To date, research on questions as reader engagement in academic writing has largely been confined to the study of English with the almost exclusive focus on direct questions. Moreover, research has focused primarily on the functions of questions in academic writing as a metadiscourse marker, with far fewer studies considering formal and textual aspects of questions. That is the aim of this paper.

Overall, direct questions have been found to perform myriad functions. As mentioned in Section 1, these functions include generating emotion, increasing visibility, and creating interesting titles. Further studies have described the use of direct questions in academic writing as means to question certainty in scientific and popular science research articles and to protect writers by expressing “the dubious nature of results” (Pic & Furmaniak, 2014: 370). In research article titles, direct questions are described as serving to hook readers (Ruegg & Sugiyama, 2013), while elsewhere, seeing direct questions as inappropriate, as Swales and Feak did (1994), direct questions in titles have been referred to as the introduction of click-bait to academic writing (Hamby, 2015) and designated as an informal feature of academic writing by Hyland and Jiang (2019).

Interestingly, research has shown that the use of direct questions in titles of publications results in a higher degree of citations for those publications, which Cook and Plourde (2016) imply can be a motivation for writers to employ the use of direct questions. For Spanish, Soler (2009, 2011) has identified direct questions as performing a useful function in re-opening discussions or emphasising a point, and in French, Curry and Chambers (2017) applied Hyland’s (2002) framework of seven question functions to the analysis of English and French linguistics research articles. The study finds that French, like English, largely uses direct questions to organise the text; set up claims; and frame the discourse. More recent work from Hyland and Jiang (2016) on direct questions as reader engagement markers finds that, while reader engagement use appears to be decreasing over time, in the field of biological sciences there has been an increase in question use to engage readers.

From a formal perspective, research on questions as reader engagement is comparatively sparse. Each of the studies cited already in this section focuses exclusively on the analysis of direct questions, where each question studied was identified by the presence of a question mark as an IFID. For indirect questions, it is necessary to look beyond academic writing research to understand how indirect questions are formally constructed.

Karttunen (1977) proposes that indirect questions are signalled with IFIDs such as *wh*-clauses, which he defines as clauses that contain question embedding verbs followed by words such as *whether*. Table 1 shows the question-embedding verbs that function as indirect question IFIDs, as identified by Karttunen (1977).

**Table 1.** Verbs used to create indirect questions in Karttunen (1977: 384–385)

<b>Verb function for generating indirect questions</b>	<b>Example verbs in English</b>
Verbs of retaining knowledge	know, be aware, recall, remember, forget
Verbs of acquiring knowledge	learn, notice, find out, discover
Verbs of communication	tell, show, indicate, inform, disclose
Decision verbs	decide, determine, specify, agree on, control
Verbs of conjecture	guess, predict, bet on, estimate
Opinion verbs	be certain about, have an idea about, be convinced about
Inquisitive verbs	ask, wonder, investigate, be interested in

Verbs of relevance	matter, be relevant, be important, care, be significant
Verbs of dependency	depend on, be related to, have an influence on, be a function of, make a difference to

---

Building on research on indirect questions, Romero (2005: 687–688) proposes “concealed questions” as a type of indirect question, where sentences that do not hold typical question or embedded question syntax correspond to embedded interrogative clauses. He argues that the existence of a question is related to the illocutionary intention to reveal information as opposed to its formal construction. Therefore, in this view, indirect questions do not need to have a *wh*-complement but are defined by their relationship to the propositional content. As such, indirect questions become more difficult to identify, as the IFIDs are not necessarily linked to a *wh*-complement or a specific verb. Example (3) shows a *wh*- complement indirect question in English and (4) shows indirect question in English without a *wh*- complement, extracted from the corpus analysed in this paper (introduced in Section 3.2).

(3) It is not *clear whether* inflation rises or falls with the amount of time left in office. (engecon27)

(4) However, as international negotiations concentrate on a country's domestic policies, one unanswered question remains the proper means of enforcement. (engecon38)  
 (“What are the proper means of enforcement?”)

Following Romero (2005), the indirect questions in Examples (3) and (4) are raised with the intention of revealing some information and are only identified as questions according to the context and the co-text. Therefore, indirect questions are not necessarily formally constrained and can occur freely, without *wh*- complements. In early studies of indirect questions, Karttunen (1977) delimited his study of indirect questions to those indicated by *wh*-complements and question verbs. More recently, Blagojević and Misić-Ilić’s (2012) study, which is noteworthy for its inclusion of indirect questions in its analysis of academic writing, delimits its view of indirect questions to those composed of *wh*- or *if/whether* clauses and IFID verbs or nouns. Such delimitations mean that these studies do not account for concealed indirect questions that do not contain any *wh*- elements. Moreover, they only consider verbs and nouns as indirect question IFIDs. This presents us with a limited view of indirect questions as reader engagement markers. That is to say, as direct and indirect questions as academic metadiscourse are defined by their illocutionary force within the text, their wholesale analysis should not be

limited to specific form-based searches. Such a limitation is a case of methodological circularity and may result in the presentation of incomplete claims on the formal nature of indirect questions in academic writing. In a move to address this issue of methodological circularity, the following section briefly reviews research in the domain of function-to-form corpus analysis, indicating the relevance of contrastive linguistic theory therein for conducting rigorous corpus-based contrastive analyses.

## 2.2 Function-to-form corpus-based contrastive analysis

Research in corpus linguistics has traditionally taken a “vertical-reading methodology” (Aijmer & Rühlemann, 2014: 5) where, through corpus analysis software, node words are studied in vertical lists. This approach has been integral to the study of form and syntax and the determination of formally and functionally constrained language, albeit from a form-to-function perspective. In function-to-form studies, such vertical reading is inadequate, as “it is often necessary to look at context to understand the functions, when taking a function-to-form approach to corpus pragmatics” (Curry & Chambers, 2017: 10).

Built upon both corpus linguistic and pragmatic approaches to language study, corpus pragmatics is not concerned solely with the vertical study of language in abstraction, but a combination of corpus vertical-reading and pragmatic horizontal-reading. Therefore, corpus pragmatics allows for a focus on discourse, co-text, and context. However, owing to limitations inherent in the nature of corpus linguistic techniques, function cannot be easily retrieved from a corpus. Rather, as Aijmer and Rühlemann (2014: 9) note, “only surface forms ‘orbiting’ [functions]” can be retrieved. Such surface forms can be based on studies of pragmatically annotated corpora, sampling and sifting of corpora, and bottom-up analyses of wordlists and key word searches (O’Keeffe, 2018: 598), for example. O’Keeffe (2018: 601) notes the advantages of working on small, specialised corpora for such functional analyses, reflecting long-standing attestations from Aston (2001: 30) that specialised corpora can boast advantages and be heavily patterned, allowing for the extraction of valuable insights into language use in the respective areas the corpora represent. Curry and Chambers’ (2017) work on questions in academic writing demonstrates a function-to-form corpus pragmatic approach to studying language, where the focus is on the identification of language that performs a question-asking function. In this case, the focus on direct questions involves searching for question marks as IFIDs. The question marks act as an accessible “starting point for electronic searches in



language corpora” (Flöck & Geluykens, 2015: 7). However, for indirect questions a more complex approach is needed to avoid the methodological circularity discussed in Section 2.1.

The issue of methodological circularity in function-to-form studies of language is not unique to corpus linguistics. In fact, where contrastive linguistics has followed a pattern of “success-decline-success” (Granger, 2003: 13), much of its decline was owing to perceived methodological weaknesses in earlier contrastive approaches. On this issue, Chesterman (1998), in his seminal work, *Contrastive Functional Analysis*, warns the contrastivist of putting the cart before the horse, as he finds that earlier contrastive approaches often suffered from methodological circularity (Chesterman, 1998: 53). Such methodological pitfalls arose where studies that perceived two things to be similar could be designed to prove a perceived similarity. Improved theorisation in the field moved to address these issues, with a specific focus on the tertium comparationis and equivalence.

The tertium comparationis is the comparable common ground (Connor & Moreno, 2005) and platform of assumed similarity from which a contrastivist contrasts two or more languages. In the case of this paper, the tertium comparationis is the presence of questions used as reader engagement markers. In contrastive linguistics, there can be only one tertium comparationis per analysis. Therefore, in this study, form cannot be assumed and searched in the corpus, as it is a finding. More than form, the tertium comparationis must be considered at all strata of data compared (Granger, 2010). Therefore, it is important to avoid having too many variables for comparison as effective contrastive analyses involve comparing like with like. As such, a delimited focus on a small genre- and discipline-specific multilingual corpus, with samples of each language collected using the same sampling frame and holding disciplinary genres that correspond across languages, allows for a more effective contrastive analysis. In essence this satisfies O’Keeffe’s (2018) emphasis on the value of small specialised corpora to the study of function as well as contrastive linguistic approaches to ensuring that the tertium comparationis cannot itself be a variable. With the presence of questions as a tertium comparationis, further challenges arise for identifying indirect questions IFIDs, as one cannot choose a range of forms to test for function. Instead, one should use the function of question raising to determine the forms used to raise questions.

Complementing the tertium comparationis, equivalence is used to test the tertium comparationis and to measure a degree of similarity or ‘sameness’ in form or function across languages. In the case of this paper, the equivalence used to test the tertium comparationis is the formal structure of questions as direct or indirect questions. Therefore, in order to ensure that this study avoids the methodological pitfalls of both function-to-form and contrastive

analyses, the following section presents a novel and in-depth approach to identifying indirect questions in academic writing.

Building on the review of the literature presented throughout Sections 2.1 and 2.2, this paper is guided by the following four research questions:

- RQ1. To what extent does the presence of questions as reader engagement correspond in English, French, and Spanish economics research articles?
- RQ2. How are indirect questions indicated in English, French, and Spanish?
- RQ3. How can indirect question IFIDs be identified in a corpus, following a function-to-form approach?
- RQ4. How can corpus linguistic approaches to function-to-form analyses benefit from contrastive linguistic approaches?

### **3. Data and methodology**

This section briefly presents the data analysed in this study and the method applied to extract indirect questions. Two thirds of the data in this research are taken from KIAP – Kulturell identitet i akademisk prosa/Cultural identity in academic prose (Fløttum et al., 2006). The KIAP corpus was developed to study cultural identity in academic writing. It is a multilingual, comparable corpus, and is composed of 450 research articles: 150 research articles in English, French, and Norwegian, respectively. These research articles are sub-categorised according to discipline, with 50 each in linguistics, economics, and medicine, for each language. This study used a subcorpus of research articles from economics in English (henceforth *engecon*) and French (henceforth *frecon*). For Spanish, a comparable subcorpus of Spanish economics research articles (henceforth *specon*) was created to conduct a trilingual study. A detailed description of the compilation process for KIAP is presented in Fløttum et al. (2006: 6–17). This corpus construction process was followed for the creation of the Spanish subcorpus, rendering it comparable to KIAP.

#### **3.1 Data: KIAP-EEFS**

After compiling a corpus of Spanish economics articles using the same sampling frame as the KIAP English and French economics subcorpora, the resulting dataset used in this study (henceforth KIAP-EEFS, KIAP, Economics, English, French and Spanish) comprises 1.2

million tokens. KIAP-EEFS is a small, multilingual, specialised corpus. Table 2 presents the size of each subcorpus when analysed using *AntConc* (Anthony, 2019).

**Table 2.** Tokens in KIAP-EEFS

<b>Subcorpus</b>	<b>Tokens in KIAP-EEFS</b>	<b>No. of articles in the (sub)corpus</b>
engecon	397,896	50
frecon	399,272	50
specon	412,631	50
Total	1,209,799	150

KIAP-EEFS is a comparable corpus that contains texts “collected using the same sampling frame” (McEnery & Xiao, 2008: 20). As such, it was important that the texts were equal in proportion according to language and genre and were collected within the same sampling period. The smaller size of the corpus is especially advantageous for conducting in-depth qualitative analyses (O’Keeffe, 2018: 601). Moreover, the delimitation of the corpus to a focus on economics research articles was chosen as it is lesser studied, exhibits rich metatext, and, importantly, reflects one specific academic discourse community, following Swales (1990, 1998). This community was defined by its interest in the dissemination of research on economics through disciplinary genres, following Bhatia (2004), that behave similarly across languages. This focus allowed for the maintenance of a *tertium comparationis* at each stratum of the corpus. While, of course, this limits the findings from speaking broadly about academic language across discipline and genre, it allows for a robust function-to-form analysis which can be replicated or used to inform future analyses, elsewhere. The following section presents the method for identifying and analysing questions in KIAP-EEFS.

### 3.2 Finding questions in KIAP-EEFS

This research has focused only on direct and indirect questions that form part of the authorial text and that serve to render the research article as interpersonal, i.e. questions that allow the author to communicate and interact with the reader. Only those items that occur between the opening tag <title> and the closing tag </title> and the opening tag <intro> and the closing tag </concl> were included.<sup>3</sup> This study excludes questions within examples, tables, and

quotations, as these were not considered interactive in terms of writer and reader interaction; that is to say, only those questions included by the author to pose a question in order to interact with the reader in some way were included.

Direct questions are easily searchable and can be identified by searching for the question mark character. However, as discussed in Section 2.1, identifying indirect questions is much more challenging. What follows is a documentation of the identification process for indirect questions.

- (i) In each language, all instances of *wh*- complements, relative pronouns, if conjunctions, and question words were searched e.g. *what*, *qu'est-ce* and *qué*.
- (ii) Given the quantity of items, randomly sampled concordance lines of each KWIC were captured. To generate appropriate samples, each sample was extracted at 95% confidence +/- 5%. This means that there is 95% confidence that the sample is balanced.
- (iii) Each sample was analysed to identify indirect questions.
- (iv) As indirect questions can occur without *wh*- complements, the search could not be constrained to this form. Therefore, the words used in conjunction with these question words to raise indirect questions – identified in step (iii) – were noted e.g. *ask*, *se demander*, *cuestionar*.
- (v) To identify indirect questions and not constrain by form:
  - a. All verb, adverb, adjective, and noun forms and inflections related to the identified words were searched and captured e.g. question as a noun (*question*), verb (*question*), adjective (*questioned*, *questionable*), and adverb (*questionably*) in each language.
  - b. Other words known to be used to raise indirect questions, drawn from previous research, were searched and captured e.g. the question words in Table 1. Furthermore, through a close read of the frequency lists for each subcorpus, words perceived to have the potential to indicate a question were identified.
  - c. Based on all of the possible IFIDs identified, further confidence samples were extracted from all KWIC searches at 95% confidence +/- 5%.
- (vi) These samples were analysed and all indirect questions were extracted.

This process resulted in the identification of the following number of questions in KIAP-EEFS as presented in Table 3, where WPM denotes words per million. Words per million was deemed the appropriate relative measure for cross-corpus analysis as it was necessary to identify the number of questions used in each database, the size of which are determined in words. Therefore, in every million words of language represented by the economics research articles in the English subcorpus of KIAP-EEFS, there are 505 questions. Note that at this point, no distinction is made between direct and indirect questions, as question type is a result of the analysis and is addressed in Section 4.

**Table 3.** Questions in KIAP-EEFS

<b>Language</b>	<b>English</b>	<b>French</b>	<b>Spanish</b>
Total	201	129	98
WPM	505	323	237

Once identified, the questions were analysed in terms of their frequency and dispersion across the subcorpora, in order to determine whether questions are a feature of the academic writing represented in each subcorpus in KIAP-EEFS, which establishes them as a tertium comparationis. Following that, the use of direct and indirect questions in KIAP-EEFS was identified, with a focus on the IFIDs used to generate indirect questions in each language.

## **4. Findings**

This section presents the results of the corpus-based contrastive analysis of questions in KIAP-EEFS. In Section 4.1, the analysis of the presence of questions as reader engagement in each language is presented. This is followed by the presentation of the analysis of indirect question IFIDs in English, French, and Spanish, in Section 4.2. It is important to note the findings presented herein surrounding indirect question IFIDs are constrained by the focus on economics research articles only.

### **4.1 Questions in KIAP-EEFS**

As seen in Table 3, each subcorpus was found to contain a differing number of questions and, upon further investigation, the questions within each subcorpus appear to have a relatively even

dispersion. Figure 1 shows that the mean (red bar) and median (black bar) are relatively comparable and not overly skewed by outliers, which are the blue dots above the box plots. This finding was further supported when a Juillard's D deviation test was applied to the data; the English, French, and Spanish subcorpora scored relatively highly, at 0.83, 0.84, and 0.81, respectively.

**INSERT FIGURE 1 HERE**

**Figure 1.** Dispersion of questions in KIAP-EEFS in WPM

Overall, questions appear to be a relatively evenly dispersed feature across all three subcorpora. However, questions are used to varying degrees in each subcorpus, with engecon using them more than twice as often as specon, and frecon situated in between. In fact, while there is no significant difference in question frequency between engecon and frecon, and frecon and specon, there is a significant difference between engecon and specon, as reported in Figure 2.

**INSERT FIGURE 2 HERE**

**Figure 2.** 95% confidence intervals for question use across languages in WPM

Overlapping error bars, as can be seen with engecon and frecon as well as frecon and specon, indicate that there is no significant difference in the use of questions across the subcorpora (Brezina, 2018: 31–32). However, non-overlapping error bars, as can be seen with engecon and specon, indicate a significant difference in the presence of questions within the subcorpora (Brezina, 2018: 31–32). Given that both the dispersion test in Figure 1 and the Juillard's D deviation scores indicate a relatively even and comparable dispersion across the subcorpora, the significant difference is not likely a result of outliers. Therefore, the evidence suggests significant differences between the English and Spanish data. However, this should not be taken to mean that questions are not a feature of Spanish academic writing in economics research articles. As Figure 1 shows, questions are a relatively evenly dispersed rhetorical feature of economics academic writing in each language.

Overall, there are enough examples and a sufficiently even distribution to render questions an interesting feature of academic discourse, as represented within KIAP-EEFS. That being said, when compared to the Spanish data, questions in engecon are significantly more frequent. Given that questions are a rhetorical feature of academic writing in each language in KIAP-EEFS, they are established as a tertium comparationis. To further identify similarities and differences in question use, the following section tests this established tertium comparationis in terms of question type as direct or indirect questions.

#### 4.2 Direct and indirect questions in KIAP-EEFS

Across the three subcorpora, questions are categorised according to two types: direct and indirect. Each subcorpus uses mainly direct questions, and the English data use more direct and indirect questions than the French data, while the French data use more direct and indirect questions than the Spanish data, as presented in Figure 3.

**INSERT FIGURE 3 HERE**

**Figure 3.** Direct and indirect questions in KIAP-EEFS in WPM

In terms of distribution within each subcorpus, 37.8% of English, 43.2% of French, and 33.7% of Spanish questions are indirect, as indicated in Table 4. Overall, direct questions are more common in each subcorpus and more questions in the French data are indirect when compared to English and Spanish.

**Table 4.** Percentage direct and indirect questions in KIAP-EEFS

Question Type	English	French	Spanish
Direct questions	62.2	56.8	66.3
Indirect questions	37.8	43.2	33.7

While direct questions are signalled by a question mark, indirect questions, as detailed in Section 3.2, are identified by a range of IFIDs. The word classes identified that act as indirect question IFIDs are nouns, adjectives, and verbs. In the case of each subcorpus, verbs are most

frequent, followed by nouns, and adjectives, as presented in Table 5. It should be noted that, owing to the in-depth qualitative analysis of indirect questions in KIAP-EEFS, it was possible to include concealed questions in the identification of indirect questions; however, concealed questions occur rather infrequently throughout the subcorpora. The following sections present the findings on nouns, adjectives and verbs in Sections 4.2.1, 4.2.2, and 4.3.3, respectively.

**Table 5.** IFIDs part of speech that generate indirect questions in KIAP-EEFS in WPM

IFID form	English	French	Spanish
Noun	40	15	22
Adjective	15	0	5
Verb	136	117	53

#### 4.2.1 Nouns as indirect question IFIDs in KIAP-EEFS

The English data use twice as many nouns to ask indirect questions when compared to the French and Spanish data. However, indirect questions raised with nouns account for 8% of all English questions, 4.7% of all French questions, and 10.2% of all Spanish questions in KIAP-EEFS. Therefore, nouns emerge as somewhat important for raising indirect questions in both English and Spanish economics academic writing. Table 6 shows the most common noun forms used to create indirect questions in each subcorpus.

**Table 6.** Nouns used to create indirect questions in KIAP-EEFS in WPM

Rank	English	French	Spanish
1	question	35 question	15 cuestión 5
2	investigation	3	pregunta 5
3	test	3	proporción 2
4			cuestiones 2

The noun *question* is the only finding of note when comparing the subcorpora. Examples (5), (6) and (7) show how *question* and its cognates are used in English, French, and Spanish economics academic writing, respectively.



(5) The first extension concerns the *question* of whether the proposition in Theorem 1 can be reversed. (engecon07)

(6) La *question* qui se pose alors est de savoir quel sera le "prix" de cette action évalué par les différents individus (frecon08)

("The question that then arises is to determine identify how the "price" of this action will be evaluated by the different individuals")

(7) Ahora bien, la *cuestión* es cómo conseguir estos objetivos (specon10)

("Now, the question is how to achieve these objectives")

From these examples in the English, French, and Spanish data, the words *question*, *question*, *cuestión* are being used to set up indirect questions. These indirect questions serve rhetorical purposes in each language to engage the reader, with Example (5), for example, understood as "can the proposition in Theorem 1 be reversed?". This creates a question that the author answers later in the text. A similar question strategy emerges in Examples (6) and (7), which ask, "how will the price of this action be evaluated?" and "how can we achieve these objectives?", respectively. These question functions are determined based on the evidence within the wider research article that answers are proposed, to some degree. The relationships between these indirect questions and the wider research articles in which they occur exemplify the importance of context and co-text in analysing function. In each case, it was necessary to look beyond the concordance line to determine whether or not uses of *question* behave as indirect question IFIDs. An interesting finding surrounding these cognates arises, as neither the English nor French data use *question* as a verb, while the Spanish data use both the noun and verb parts of speech of *cuestión*.

#### 4.2.2 Adjectives as indirect question IFIDs in KIAP-EEFS

Adjectives are rarely used in KIAP-EEFS as indirect question IFIDs. There are no examples in the French data and few examples in the English and Spanish data, which account for only 3% and 2% of all questions found in each subcorpus, respectively. The only examples of adjective IFIDs in engecon is the use of *interested* and *clear*, as Example (8) illustrates with *clear*, and the only example in specon is the use of *questionable* to create an indirect question, as presented in Example (9).

(8) It is not *clear* whether inflation rises or falls with the amount of time left in office. (engecon27)

(9) Resulta *cuestionable* la medida en que esta recomendación mejore efectivamente el aprovechamiento de los estudiantes (specon44)

(“The extent to which this recommendation effectively improves student achievement is questionable”)

In these examples, the negation of *clear* combined with *whether* questions the clarity and seeks clarification. The question itself is determined owing the move to clarify whether “inflation rises or falls”. Similarly, for *cuestionable*, the evaluation of the preceding noun phrase as *questionable* calls the noun phrase into question.

It is important to note that due to the minimal use of adjectives to create questions in the corpus, there is little of substance to be drawn from those that do occur. That being said, adjectives are typically excluded from studies of indirect questions and, while they are used rarely in KIAP-EEFS, they do occur. It would be interesting to explore adjectives as question IFIDs in a larger data set spanning more genres and disciplines.

#### 4.2.3 Verbs as indirect question IFIDs in KIAP-EEFS

The ten most frequent verbs used in KIAP-EEFS to create indirect questions can be seen in Table 7.

**Table 7.** Verbs used to generate indirect questions in KIAP-EEFS in WPM

Rank	English	French	Spanish
1	Ask	25 interroger	28 preguntar 12
2	Examine	20 déterminer	15 investigar 10
3	Explore	17 savoir	15 precisar 7
4	See	10 montrer	15 evaluar 5
5	Tell	8 demander	15 conocer 5
6	Check	5 tester	8 saber 5
7	Gauge	5 poser	5 examinar 2
8	Test	5 étudier	3 cuestionar 2
9	understand	5 examiner	3 ahorrar 2
10	Determine	5 analyser	3 determinar 2

Overall, there is higher reuse of verbs and examples of cognate verbs used to raise indirect questions in each language. For example, the verbs *ask*, *demander*, *interroger*, *preguntar*, and

*questionar*, which share very similar meanings to *ask*, *question*, or *wonder*, all rank quite highly among verbs used to create indirect questions. Moreover, they work in very similar ways as can be seen Examples (10), (11), (12), and (13).

(10) Finally, we briefly characterize some stylized facts regarding our estimated stocks and *ask* whether there are trends in net foreign assets and shifts in their composition over time. (engecon02)

(11) Si ces résultats devaient toutefois être confirmés, il serait légitime de se *demander* s'ils reflètent la spécificité du dispositif français de formation continue ou si leur portée est plus générale. (frecon22)

(“If these results were to be confirmed, however, it would be legitimate to *wonder* whether they reflect the specificity of the French continuing training system or whether their scope is more general”)

(12) Il nous semble donc intéressant de revenir aux fondements de la microéconomie du consommateur et *d'interroger* cette dernière à la lumière de la soutenabilité (frecon02)  
(“It therefore seems interesting to return to the foundations of the consumer microeconomy and to *question* it in the light of sustainability”)

(13) De estas las principales son de Estados Unidos con lo cual hay que *preguntarse* seriamente si el proceso de globalización no debiese llevar más bien la etiqueta “Made in USA” y ser considerado simplemente como un nuevo producto que venden las transnacionales de dicho país (specon31)

(“Of these, the main ones are from the United States, so we must seriously *ask* ourselves if the globalization process should not rather carry the label "Made in USA" and be considered simply as a new product sold by the transnational companies of that country.”)

These indirect questions were found to engage readers by outlining the research questions of the paper, as in Example (10), by indicating the content of the subsequent section, as in Examples (11) and (13), and by expressing critique in Example (12). The Spanish data also make use of the verb *investigar* with no examples of this verb cognate being used in the English data. In Example (14), *investigar* is used to form an indirect research question, which expresses the need to discover something in the research article.

(14) Primero, intentamos poner a prueba simultáneamente los pronósticos principales de dos grupos de modelos de las crisis (los llamados modelos de primera y segunda

generaciones de crisis), a fin de *investigar* si estos modelos teóricos nos han ayudado a entender por qué ocurren las crisis y elucidar la cuestión de si estos modelos son complementos o sustitutos en la explicación de la presencia de crisis. (specon37)

(“First, we try to simultaneously test the main forecasts of two groups of crisis models (so-called first and second generation crisis models), in order to *investigate* whether these theoretical models have helped us understand why crises occur and elucidate the question of whether these models are complements or substitutes in the explanation of the presence of crises.”)

The French data show the use of *analyser* to raise questions that indicate the content of the subsequent section, as illustrated in (15). This use of a verb and a question word (*analyser comment*) to reveal information is a relatively common means for posing and making use of an indirect question to indicate what will be discussed next in the text.

(15) Nous allons analyser comment les solutions en [...] de ce système d'équations varient en fonction de N. (frecon43)

(“We are going to analyse how the solutions in [...] of this system of equations vary according to N”)

The English data make more frequent use of *examine*, with *frecon* and *specon* using *examiner* and *examinar* much less frequently. However, all three create a question by indicating what the writers want to know through the proposed examination. Examples (16), (17), and (18) exemplify this.

(16) Building on these predictions, we *examine* whether there is a negative relationship between the strength of FPRs and labor flows from the source country to the affiliates. (engecon04)

(17) Au préalable, nous *examinons* quelles sont les questions spécifiques posées par le système de retraite en matière de redistribution (frecon15)

(“Beforehand, we *examine* what are the specific questions raised by the pension system in terms of redistribution”)

(18) *Examinare* ahora como es afectada la tasa de interés a largo plazo por las políticas fiscales de manera congruente con el crecimiento de estado estacionario. (specon49)

(“I will now *examine* how the long-term interest rate is affected by fiscal policies in a manner consistent with steady state growth.”)

This use of *examine* and its French and Spanish cognates signal an indirect research question that frames the paper, in (17), and a question that helps to guide the reader through the text, by raising questions to be dealt with in subsequent sections, in (18) and (19). This can be seen in Example (18), with the use of *au préalable* as an endophoric reference marker and in (19), with the use of *ahora* as a means to use an indirect question to create a sequence and offer guidance to the reader.

The English and French data also bear some similarities in the use of *test* and *tester* which does not feature in the Spanish data, as it is not a cognate form. What may be considered its close equivalent, *comprobar*, does not feature either, indicating that *test* is not a verb used to raise indirect questions in Spanish economics academic writing in economics. In Examples (19) and (20), *test* and *tester* are used to reveal information, like many of the verbs already discussed here.

(19) The final alternative specification *tests* whether Fed interventions that are coordinated with one or both of the other G-3 central banks are more influential than unilateral interventions. (engecon46)

(20) Si l'hypothèse de marchés complets est rejetée, il est donc intéressant de *tester* si le métayage favorise le partage des risques ou non. (frecon29)

(“If the hypothesis of complete markets is rejected, it is therefore interesting to *test* whether sharecropping promotes risk sharing or not.”)

The French data make greater use of verb *déterminer*, whose cognate also occurs in the list of verbs in engecon and specon, but less frequently. However, in all three, the verb seeks to prove or show something, again related to the idea of the discovery or revelation of information.

(20) Notre objectif est maintenant de *déterminer* les prix d'équilibre (frecon43)

(“Our objective now is to *determine* the equilibrium prices”)

The French data also use high-ranking verbs, such as *savoir*, to create indirect questions, as Examples (21) demonstrates. Its equivalent verbs, meaning *know*, do not feature as indirect question IFIDs in the English or Spanish subcorpora. In (21), the authors are studying the way in which decisions are checked. They want to know if banks pay too much.

(21) La question qui importe, et qui est celle étudiée dans l'article, est donc celle de savoir si une banque paie plus cher un besoin accru de financement externe (frecon48)

(“The important question, which is the one studied in the article, is therefore whether a bank pays more for an increased need for external financing”)

The English data feature verbs not evident in the French and Spanish data, such as *explore*, *see*, and *tell*, to create indirect questions, as illustrated in (22), (23), and (24) exemplify. In these examples, each verb is used to express an intention to discover the answer to something and deliver new information.

(22) We *explore* whether FPRs play a more important role in protecting knowledge that is transferred outside the source country and firm relative to knowledge inside the source country and firm. (engecon04)

(23) The aim of the present paper is to *see* whether the change of currency regime slowed the growth of the volume of Anglo–Irish trade, *ceteris paribus* (engecon28)

(24) The data will then *tell* us whether, within the theoretical framework developed, these proxies for lobbying costs influence protection. (engecon40)

Finally, *specon* makes use of the verb *precisar* to create questions. In Example (25), the authors want to “specify whether tourist activity is really profitable”. That is to say, they want to answer the question, “is tourist activity really profitable?”

(25) Es indudable que las dificultades estadísticas que habrá que superar, a fin de *precisar* si la actividad turística es realmente rentable y si ésta es preferible a las distintas alternativas existentes, serán, frecuentemente, considerable (specon11)

(“Undoubtedly, the statistical difficulties that will have to be overcome, in order to *determine* if the tourist activity is really profitable and if it is preferable to the different existing alternatives, will frequently be considerable”)

In summary, IFIDs of indirect questions in each subcorpus are performed mostly through verbs, followed by nouns. Adjectival IFIDs are negligible in general. The main verbs used to create indirect questions are *ask*, *interroger*, *preguntar*, *examine*, *déterminer*, *investigar*, and *explore*, *savoir* and, *precisar*. While there is some degree of uniformity across languages, there is also evidence of the use of certain verbs in each language that do not emerge in the other two, or if so, less frequently. For nouns, *question* and its French and Spanish cognates are noteworthy

IFIDs for raising indirect questions. The Spanish data alone exhibit the use *cuestión* as a verb *cuestionar*.

## **5. Questions as reader engagement makers: discussion of findings**

The findings of this research are discussed in this section. First, the results of the analysis of the presence of questions in KIAP-EEFS are presented. Second, the findings on IFIDs are discussed in relation to the wider literature. Finally, reflections on the methodological approach to function-to-form corpus-based contrastive analysis applied herein are presented and contextualised within wider corpus and contrastive linguistic approaches.

The finding that the English data use many questions as engagement markers is not surprising, given that researchers, such as Lafuente-Millán (2014), recognise that English makes use of many metadiscourse markers in academic writing. Hyland (2005b) has described English academic writing as explicit in nature, which would explain the quantity of questions present. However, another possible explanation derives from the propensity for English academic writing to demonstrate less certainty traditionally, due to writers anticipating rejection on the part of their readers (Hyland, 2005a). In this view, questions may reflect Pic and Furmaniak's (2014) identification of questions as being used to signal a lack of certainty.

Questions are used frequently in both the English and French data, but more so in the English data. This disparity is explicable. For example, the findings of this study correspond to Dahl (2004), who finds that metatext occurs less frequently in French economics writing when compared to English. However, it is important to recognise that such a relative perspective does not indicate that French does not make use of metatext at all. In fact, Fløttum et al. (2006) find that French economics writing, in and of itself, contains rich metatext. This is likely why, although less apparent than in *engecon*, *frecon* also contains many examples of questions used as reader engagement markers.

The relatively lower frequency of question use in *specon* is also somewhat unsurprising. For example, in her study of self-mentions, Mur-Dueñas (2007) finds much less evidence of metadiscoursal markers in Spanish than in English. This is similar to her study of evaluative connectors (Mur-Dueñas, 2011), which identifies that Spanish exhibits few examples of evaluative markers. In the same vein, Lorés-Sanz (2011a, 2011b) finds much less use of exclusive pronouns in Spanish, Pérez-Llantada (2010) finds fewer participant-oriented metadiscourse markers in Spanish, and Breeze (2012) considers that Spanish is a less “reader-friendly” language than English, owing to a lack of reader-centric metadiscourse in the former.

In fact, it has long been argued that metatext is infrequent in Spanish academic writing in general (Valero-Garcés, 1996).

A key pattern to observe in discussions of cross-linguistic analyses is the use of comparative forms like *fewer*, *greater*, *less*, etc. These demonstrate the importance of relativity when comparing items of language, like questions, across languages and the importance of not defining a language only in terms of another. In this study, Spanish uses significantly fewer questions than English. However, they are not absent, and unlike Lafuente-Millán's (2014) study of directives and questions in business management research articles, this research finds questions in Spanish to be a valuable rhetorical feature of economics research articles, albeit a less used resource in Spanish than English. This brings the findings of this research to bear important considerations for the widely accepted designation of French and Spanish as writer-oriented languages that place responsibility on readers to navigate texts (Salager-Meyer, 2011). Based on the analysis presented here, such a designation is questionable in the context of academic writing in economics research articles. Relatively, it is reasonable to say that French and Spanish are less reader-oriented than English. However, it would be inaccurate to say that French and Spanish do not engage readers at all, as questions as reader engagement are a definitive rhetorical device in each subcorpus of KIAP-EEFS. With this in mind, future contrastive studies should endeavour not only to describe languages as they relate to one another, but also offer individual descriptions of languages, to set relative perspectives in their contexts and readdress thinking on reader- and writer-oriented languages.

Overall, it is clear that questions as reader engagement markers occur to varying degrees across the English, French, and Spanish subcorpora of KIAP-EEFS. These questions differ in frequency, with more in English than French and Spanish, and more in French than Spanish. That being said, they occur across a range of texts and their differences in frequency are understandable as they reflect contextual and cultural factors in academic writing. In each language, questions were found to reflect a range of functions identified in the literature (discussed in greater detail in Curry, 2021), including signalling research questions (Hyland, 2002), helping guide readers through texts (Tse & Hyland, 2006), critiquing literature (Pic & Furmaniak, 2014) and revealing information (Romero, 2005).

The overall findings surrounding direct and indirect questions indicate that questions occur most frequently as direct questions in each language. English uses more of each question type, followed by French and Spanish. However, proportionally, French uses a higher percentage of indirect questions when compared to other languages, followed by English, and Spanish. Reflecting on the wider literature, very few studies of questions in academic writing have made



the distinction between direct and indirect questions. While Blagojević and Misić-Ilić (2012) include a focus on indirect questions, their focus is formally constrained. Therefore, they limit question IFIDs to specific forms. This paper offers a more detailed and expansive view of indirect question IFIDs than has yet been presented in the literature on questions as reader engagement devices in English, French, and Spanish academic writing. Moreover, while some research does exist on indirect questions in English academic writing, to the author's knowledge, no such corpus-based descriptions of indirect question IFIDs in French and Spanish exist to-date. Therefore, the IFIDs presented herein hold valuable insight into questioning practices in English, French, and Spanish academic writing in economics and are a valuable resource for future studies of indirect questions in each language.

While a primary aim of this research is to offer insight into question form and specifically indirect question IFIDs in the English, French, and Spanish data, the role of such questions therein is also of interest. The greater use of direct questions in English reflects the documented explicitness of the English language (Fløttum et al., 2006). Conversely, the presence of direct questions being greater than indirect questions in French and Spanish is somewhat surprising given these languages' preference for indirectness and negative politeness strategies (Loffler-Laurian 1980; Clyne 1994; Lafuente-Millán 2014). For French, the greater proportional use of indirect questions can be linked to the French language's preference for negative politeness strategies, while for Spanish, the convergence towards English language norms (Ciapuscio & Otañi, 2002) may account for the increased use of direct question. However, owing to the lack of research on indirect questions in academic writing in English, French, and Spanish, further research investigating writers' motivations for using indirect questions is needed to better understand their role in this context.

Among the key insights on indirect question IFIDs revealed herein is the shared use of the following verb IFIDs: *ask*, *demand*, *interroger*, *preguntar*, *cuestionar*, *determine*, *déterminer*, *determinar*, *test*, and *tester*. While each subcorpus converges on the use of certain cognate or sense-sharing verbs, there was also evidence of differences in how each subcorpus evokes indirect questions. In engecon there is noteworthy use of *explore*, *see*, and *tell* as indirect question IFIDs that do not feature in frecon or specon. For frecon, the verbs *analyser* and *savoir* also demonstrate unique uses whose cognates and translations are not used in engecon and specon to raise questions. Finally, for specon, the verbs *investigar* and *precisar* are used to generate indirect questions. Similarly, the cognates and direct translations of these verbs do not occur in engecon and frecon.

Adjectives, which appear to be rarely used as IFIDs in these data, occur with words like *clear*, *interested*, and *questionable* to raise questions. While this research offers little evidence of adjectival use to raise questions, it would be interesting to determine the use of adjectives as indirect question IFIDs in a larger corpus of academic writing. To the author's knowledge, no such work on adjectives as indirect question IFIDs in English, French, and Spanish academic writing exists to-date. For nouns, the noun *question* and its cognates are frequently used to raise indirect questions in each subcorpus. This demonstrates a clear example of shared practices in indirect question creation across English, French, and Spanish, in KIAP-EEFS.

It must be noted that the findings presented herein surrounding indirect question IFIDs contribute only to knowledge on economics research articles. However, the discipline- and genre-specific nature of this study of a small specialised corpus has produced a search inventory of possible indirect question IFIDs that can now be used to search larger corpora in each language.

In reflecting on the methodological process applied herein, this paper presents a useful roadmap for effective function-to-form corpus-based contrastive analyses. In moving to identify the linguistic items orbiting indirect question IFIDs, the analysis presented herein acknowledged the limitations of function-to-form studies which typically accept the irretrievability of function in corpus analyses, and instead opt for the use of formal findings from previous pragmatic analyses (Aijmer & Rühlemann, 2014; O'Keeffe, 2018) or corpus sampling, sifting, and bottom-up processing (O'Keeffe, 2018) to identify function. The guidance from studies in corpus pragmatics were integral to the development of the method, presented in Section 3.2, and the analysis of questions, which required both vertical and horizontal reading to determine whether the forms identified behaved as indirect question IFIDs. This is because it was often necessary to look within the research articles to determine the existence of concealed indirect questions.

In combination with such guidance from the corpus linguistic literature, this study sought support in contrastive linguistic theories, which allowed for the development of a corpus-based contrastive analytical method that reduces methodological circularity when searching for forms in a function-to-form study. More specifically, contrastive linguistic theory informed the corpus approach herein through the identification of a *tertium comparationis* and equivalence, which prohibited the study from constraining its search to specific forms. As a result, the multi-staged method of question extraction from KIAP-EEFS, presented in Section 3.2, allowed for the identification of a range of noun, adjective, and verb indirect question

IFIDs that existing approaches in corpus pragmatic and contrastive linguistic analyses of IFIDs would not have identified.

## **6. Conclusion**

This paper has identified that questions are a rhetorical feature of economics research articles in English, French, and Spanish, based on KIAP-EEFS. Moreover, from a contrastive perspective, questions occur in descending frequency in English, French, and Spanish and are significantly less frequent in Spanish when compared to English in KIAP-EEFS. This study has determined that indirect questions are a generic feature of each subcorpus, accounting for more than 30% of questions identified in each language. In identifying indirect question IFIDs, this paper has presented a comprehensive list of nouns and verbs used as indirect question IFIDs. This reflects a valuable addition to the field where, to-date, no such corpus-based compilation of indirect question IFIDs in English, French, or Spanish exists.

Methodologically, this paper has presented a novel and detailed approach to function-to-form corpus analysis. By reducing the need for constraints to specific formal searches and by adopting the contrastive linguistic theories of *tertium comparationis* and equivalence, this study has reduced methodological circularity apparent in function-to-form corpus studies. By establishing a *tertium comparationis* and using it to determine the data and analytical process employed herein, the study was able to test the *tertium comparationis* of the presence of questions as reader engagement markers with the equivalence of indirect question IFIDs. As a result of this process, this paper has produced form-based findings that correspond to a cross-linguistic functional analysis.

Overall, there remain a number of areas of research on questions as reader engagement markers that warrant further investigation. For example, while functional studies of questions reflect typical approaches to research on questions as metadiscourse, formal analyses remain largely under-developed. Further research could consider formal elements of questions beyond direct and indirect binaries, such as content questions or polar questions, for example. Moreover, in the context of writing cultures, this research has shown that each language, to varying degrees, appears to engage readers. Therefore, research revisiting designations of academic writing in different languages as writer- or reader-responsible would offer a welcome development to the field and could serve to revolutionise existing thinking in this area. Finally, further studies of languages other than English and contrastive analyses based on function are

needed to help advance corpus-pragmatic and contrastive linguistic approaches and to better understand the purpose of indirect questions in academic writing. This will allow for further methodological development within corpus-based contrastive analyses in particular.

## Notes

1. IFID are “any element of natural language which can be literally used to indicate that an utterance of a sentence containing that element has a certain illocutionary force or range of illocutionary forces” (Searle et al., 1985: 2).
2. Each example extracted from the corpus presented herein is identified according to the text in the corpus from which it was taken. For the second research article in the English, French, or Spanish economics subcorpora, this will be signalled by engecon02, frecon02, and specon02, respectively. As there are 50 articles in each subcorpus, the number at the end of the identifier can range from 01–50.
3. Where there is no explicit conclusion section, the search was confined to the closing tag </mid>

## References

- Anthony, L. (2019). *AntConc* (Version 3.5.8) [Computer Software]. Waseda University. Available from <https://www.laurenceanthony.net/software>
- Aijmer, K., & Rühlemann, C. (2014). Introduction. Corpus pragmatics: Laying the foundations. In K. Aijmer & C. Rühlemann (Eds.), *Corpus Pragmatics* (pp. 1–28). Cambridge University Press.
- Aston, G. (2001). Learning with corpora: An overview. In G. Aston (Ed.), *Learning with Corpora* (pp. 7–45). CLUEB.
- Ball, R. (2009). Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966–2005. *Scientometrics*, 79(3), 667–679. doi: <https://doi.org/10.1007/s11192-007-1984-5>
- Bhatia, V.K. (2004). *Worlds of Written Discourse: A Genre-based View*. Continuum.
- Blagojević, S., & Misic-Ilic, B. (2012). Interrogatives in English and Serbian academic discourse – A contrastive pragmatic approach. *Brno Studies in English*, 38, 17–35. doi: <https://doi.org/10.5817/BSE2012-2-2>

- Breeze, R. (2012). *Rethinking Academic Writing Pedagogy for the European University*. Rodopi.
- Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.
- Chesterman, A. (1998). *Contrastive Functional Analysis*. John Benjamins.
- Clyne, M. (1994). *Inter-cultural Communication at Work: Cultural Values in Discourse*. Cambridge University Press.
- Ciapuscio, G., & Otañi, I. (2002). Las conclusiones de los artículos de investigación desde una perspectiva contrastiva [Contrastive perspectives on research article conclusions]. *Revista del Instituto de Investigaciones Lingüísticas y Literarias RILL*, 15, 117–133.
- Connor, U., & Moreno, A.I. (2005). Tertium comparationis: A vital component in contrastive rhetoric. In P. Bruthiaux, D. Atkinson, W. Eggington, W. Grabe, & V. Ramanatan (Eds.), *Directions in Applied Linguistics: Essays in Honour of Robert B. Kaplan* (pp. 153–164). Multilingual Matters.
- Cook, J.M., & Plourde, D. (2016). Do scholars follow Betteridge's law? The use of questions in journal article titles. *Scientometrics*, 108(3), 1119–1128. doi: <https://doi.org/10.1007/s11192-016-2030-2>
- Curry, N. (2021). *Academic Writing and Reader Engagement: Contrasting Questions in English, French and Spanish Corpora*. Routledge.
- Curry, N., & Chambers, A. (2017). Questions in English and French research articles in linguistics: A corpus-based contrastive analysis. *Corpus Pragmatics*, 1(4), 327–350. <https://doi.org/10.1007/s41701-017-0012-0>
- Dahl, T. (2004). Textual metadiscourse in research articles: A marker of national culture or of academic discipline? *Journal of Pragmatics*, 36(10), 1807–1825. doi: <https://dx.doi.org/10.1016/j.pragma.2004.05.004>
- Flöck, I., & Geluykens, R.I.E. (2015). Speech acts in corpus pragmatics: A quantitative contrastive study of directives in spontaneous and elicited discourse. In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics* (pp. 7–37). Springer.
- Fløttum, K., Dahl, T., & Kinn, T. (2006). *Academic Voices across Languages and Disciplines*. John Benjamins.
- Granger, S. (2003). The corpus approach: A common way forward for contrastive linguistics and translation studies? In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* (pp. 17–31). Rodopi.
- Granger, S. (2010). Comparable and translation corpora in cross-linguistic research, design, analysis and applications. *Journal of Shanghai Jiaotong University*, 2, 14–21.
- Hamby, S. (2015). On scientific writing in the information era: Tailoring papers for Internet searching and other 21st century realities. *Psychology of Violence*, 5(2), 103–111. doi: <https://dx.doi.org/10.1037/a0039008>
- Hyland, K. (2002). What do they mean? Questions in academic writing. *Text*, 22(4), 529–557.

- Hyland, K. (2005a). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. doi: <http://dx.doi.org/10.1177/1461445605050365>.
- Hyland, K. (2005b). *Metadiscourse: Exploring Interaction in Writing*. Continuum.
- Hyland, K., & Jiang, F.K. (2016). “We must conclude that...”: A diachronic study of academic engagement. *Journal of English for Academic Purposes*, 24, 29–42. doi: <https://doi.org/10.1016/j.jeap.2016.09.003>
- Hyland, K., & Jiang, F. K. (2019). *Academic Discourse and Global Publishing: Disciplinary Persuasion in Changing Times*. Routledge.
- Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25(2), 156–177. doi: <http://dx.doi.org/10.1093/applin/25.2.156>.
- Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, 1(1), 3–44.
- Khedri, M., Chan, S.H., & Helen, T. (2015). Interpersonal-driven features in research article abstracts: Cross-disciplinary metadiscoursal perspective. *Pertanika Journal of Social Sciences & Humanities*, 23(2), 303–314.
- Lafuente-Millán, E. (2014). Reader engagement across cultures, languages and contexts of publication in business research articles. *International Journal of Applied Linguistics*, 24(2), 201–223. doi: <http://dx.doi.org/10.1111/ijal.12019>.
- Loffler-Laurian, A.M. (1980). L’Expression du locuteur dans les discours scientifiques [Authorial stance in scientific discourse]. *Revue de Linguistique Romane*, 44, 135–157.
- Lorés-Sanz, R., (2011a). The construction of the author’s voice in academic writing: The interplay of cultural and disciplinary factors. *Text & Talk*, 31(2), 173–193. doi: <https://doi.org/10.1515/text.2011.008>
- Lorés-Sanz, R., (2011b). The study of authorial voice: Using a Spanish-English corpus to explore linguistic transference, *Corpora*, 6(1), 1–24. doi: <https://doi.org/10.3366/cor.2011.0002>
- McEnery, T., & Xiao, R. (2008). Parallel and comparable corpora: What is happening? In G. A. Anderman & M. Rogers, (Eds.), *Incorporating Corpora: The Linguist and the Translator* (pp. 18–31). Multilingual Matters.
- Mur-Dueñas, P.M. (2007). Same genre, same discipline; however, there are differences: A cross-cultural analysis of logical markers in academic writing. *ESP across Cultures*, 4, 37–53.
- Mur-Dueñas, P. (2011). An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics*, 43(12), 3068–3079. doi: <http://dx.doi.org/10.1016/j.pragma.2011.05.002>
- O’Keeffe, A. (2018). Corpus-based function-to-form approaches. In A. H. Jucker, K. P. Schneider, & W. Bublitz (Eds.), *Methods in Pragmatics* (pp. 587–618). Mouton de Gruyter.
- Pérez-Llantada, C. (2010). The discourse functions of metadiscourse in published academic writing: Issues of culture and language. *Nordic Journal of English Studies*, 9(2), 41–68.

- Pic, E., & Furmaniak, G. (2014). Questioning certainty in research articles and popular science articles: A case-study of modalized wh-interrogatives. In A. Zuczkowski & R. Bongelli (Eds.), *Communicating Certainty and Uncertainty in Medical, Supportive and Scientific Contexts* (pp. 371–389). John Benjamins.
- Romero, M. (2005). Concealed questions and specificational subjects. *Linguistics and Philosophy*, 28(6), 687–737. doi: <https://doi.org/10.1007/s10988-005-2654-9>
- Ruegg, R., & Sugiyama, Y. (2013). Organization of ideas in writing: What are raters sensitive to? *Language Testing in Asia*, 3(1), 1–13. doi: <https://doi.org/10.1186/2229-0443-3-8>
- Salager-Meyer, F. (2011). Scientific discourse and contrastive linguistics: explicitness and the concept of reader/writer responsible languages. *European Science Editing*, 37(3), 71–72.
- Searle, J.R., Willis, S., & Vanderveken, D. (1985). *Foundations of Illocutionary Logic*. Cambridge University Press.
- Soler, V. (2009). Títulos científicos en lengua española: estudio exploratorio [Scientific titles in Spanish: An exploratory study]. *Lebende Sprachen*, 54(2), 50–58.
- Soler, V. (2011). Comparative and contrastive observations on scientific titles written in English and Spanish. *English for Specific Purposes*, 30(2), 124–137. doi: <https://doi.org/10.1016/j.esp.2010.09.002>
- Swales, J.M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Swales, J.M. (1998). *Other Floors, Other Voices: A Textography of a Small University Building*. Erlbaum.
- Swales, J.M., & Feak, C.B. (1994). *Academic Writing for Graduate Students: Essential Tasks and Skills*. University of Michigan Press.
- Valero-Garcés, C. (1996). Contrastive ESP rhetoric: Metatext in Spanish-English economics texts. *English for Specific Purposes*, 15(4), 279–294. doi: [http://dx.doi.org/10.1016/S0889-4906\(96\)00013-0](http://dx.doi.org/10.1016/S0889-4906(96)00013-0).
- Webber, P. (1994). The function of questions in different medical English genres. *English for Specific Purpose*, 13, 257–268.

### **Address for correspondence**

Niall Curry  
 Centre for Arts, Memory and Communities  
 Coventry University  
 Priory St, Coventry CV1 5FB UK

niall.curry@coventry.ac.uk