John Benjamins Publishing Company

A crosslinguistic study of some extended uses of *what*-based interrogative expressions in Chinese, English, and Korean

Heeju Lee[1], Danjie Su[2] and Hongyin Tao[1]
[1]University of California, Los Angeles / [2]University of Arkansas

Interrogative pronouns such as *what* in English, *shenme* in Mandarin Chinese, and *mwe/mwusun* in Korean all have developed extended uses beyond interrogation. Such uses may include filling a gap in conversation, softening a speaker's epistemic stance, and indicating strong emotions such as surprises or incredulity. Yet there is little research dealing with crosslinguistic patterns with large corpora of interactive discourse data. In this paper, we investigate the extended uses based on corpora of multiple telephone calls from the three languages. We show that eight categories of extended use can be identified in the corpora and that most of the extended uses tend to fall in the negative territory. We provide a pragmatic interactive account for this phenomenon and hope that the taxonomy and coding scheme developed here can serve as a starting point for future crosslinguistic and corpus-based comparative studies of *what*-like tokens as well as of the discourse pragmatic uses of other interrogative forms.

1. Introduction

In this paper, we investigate a widespread yet only sporadically studied linguistic phenomenon whereby the interrogative 'what' and its equivalents in Chinese and Korean extends its use to perform a variety of semantic and pragmatic functions. By extended use, we mean that the interrogative form is not used by the speaker to ask a question and to solicit an answer from the addressee but rather is used for some other functions such as marking hesitation, uncertainty, disbelief, disapproval, and so forth. This can be illustrated by the following English examples of

what. In (1) *what* is used in a typical interrogative sense, whereas in (2) and (3) it is used to indicate uncertainty and incredulity, respectively:

(1)   **What** is the final score in volleyball called?   (Internet)

(2)   I mean *he scored* **what**, 129 this year? Something like that?   (Internet)

(3)   He scored **what**? That many points?!   (Internet)

Such extended uses appear to be widespread in other languages as well. For example, in Mandarin Chinese, one can say:

(4)
　　我　今天　　要　去　了解　　一下,
　　wo jintian   yao qu liaojie   yixia
　　1SG today   will go check.out   a.bit
　　'Today I'm going to check it out,'
　　拿 点　资料　什么　　的.
　　na dian ziliao   **shenme** de
　　get a few material what   PRT
　　'and get some materials and stuff like that.'

where the generalized meaning of 'stuff like that' in the second line is indicated with *shenme*, which is similar to *what not*. Similarly, in spoken Korean, one can say:

(5)   A: 전공   영어   상식   뭐   이렇게 봤–었–거든.
　　　　cenkong yenge   sangsik   **mwe** ilehkey pwa-ss-ess-ketun
　　　　major   English common.sense what like.this take-PST-PST-CORREL
　　　　'(We took tests of) such major subjects as, English, general studies, and what
　　　　 have you.'
　　　where *mwe* 'what' is also used to denote a vague reference 'what have you.'

Given the observed uses in multiple languages, a number of interesting questions can be asked: What are the range and types of derived uses that can be observed of typologically different languages? Are there common patterns of extension over time? Answers to these questions can contribute to our understanding of a common class of lexical items in the world's languages and help reveal patterns in grammaticization and linguistic universals.

   So far there is little research available to adequately answer the aforementioned questions. Two main reasons may be associated with this. First, there is no systematic taxonomy of the categories of extended use that can be relatively uniformly applied to different languages and/or for corpus-based studies. Since language can differ in many ways and extended uses can go in different directions

(e.g. vague reference vs. disapproval), there is an urgent need to define extended uses in more precise ways so that they can be applied to different languages for meaningful comparison and generalization. Second, most previous studies, to the extent relevant, are based on isolated sentences rather than corpora of natural discourse. The problem with this is that, as will be shown later, many of the extended uses are context-bound and interaction-oriented; unless we examine their surrounding interactive discourse, their functions may not be adequately revealed.

To address the shortfalls of the existent research, we use in this study corpora of telephone conversations from Mandarin Chinese, English, and Korean to develop a comprehensive coding system for the extended uses for crosslinguistic analysis. Our hope is that the categories and the coding system based on the three languages will serve as a starting point for future crosslinguistic studies of *what*-based expressions. It is also hoped that our findings can be extended to other interrogative forms (*where, who*, and *how*, and so forth) and their extended uses.

In this study, we take a corpus-driven approach (Tognini-Bonelli 2001), in the sense that our categories are strictly based on what we actually find in the corpora of the three languages. Although a corpus-driven approach can miss some types of use that are not present in the data, the distinctive advantages of this approach are that the taxonomy and coding system will be realistic and that the frequency data will show natural tendencies that may not be readily available otherwise.

2.   Review of relevant research

2.1 General and English studies

A survey of the existing literature indicates that very little research has been done in the extended uses of *what*-based interrogative expressions in actual discourse. Brinton (1996, 2008), one of the few that are concerned with the evolution of *what* and other pragmatic markers (e.g., *why, like, now, only*), touches upon the historical pragmatic development of *what*. According to Brinton, *what* develops from an interrogative pronoun/adverb/adjective indicating direct questions to a complementizer of indirect questions, and to a pragmatic marker. That is, the simple interrogative sense of *what* evolves into textual and interpersonal meanings. When *what* functions as a pragmatic marker, according to Brinton, it tends to be placed at the initial position of statements while it is freed from its syntactic position as the interrogative pronoun/adverb/adjective.

Blake (1992) identifies three interpersonal functions of *what* for Old and Middle English: (1) to cue turn-taking, (2) to elicit attention from the hearer (as in *know what*), and (3) to show the speaker's attitude such as "surprise, astonishment,

impatience, exultation or encouragement." The "surprise" sense of *what* is discussed for present-day English in Quirk et al. (1985: Section 11–15, p. 819).

Some of the literature on grammaticalization deals with interrogatives in general and investigates their extended grammatical functions. For example, in hypothesizing that grammaticalization is more likely to evolve from "less personal to more personal" than the reverse, Traugott (1982) exemplifies this with interrogatives (*where, why*, etc.) by proposing the following stages of development: interrogative (propositional) > complementizer (textual) > pragmatic marker (expressive). In the same vein, Heine and Kuteva's (2006) survey of interrogatives in European languages finds that interrogative forms evolve to become complements, adverbials, and relative clause markers. The four stages they suggest are: (1) marking word questions, (2) introducing indefinite complement or adverbial clauses, (3) introducing definite complement or adverbial clauses, and (4) introducing headed relative clauses. They contend that the development indicates the extension of question functions to other functions (i.e., syntactic relationships).

Since most of these proposals are based on historical data, it is necessary now to examine contemporary naturalistic data to substantiate or complement them, as we do for this study.

A few recent typological studies have begun to explore *what*-related expressions crosslinguistically. For example, Enfield (2003) discusses the use of *what-d'you-call-it* as oriented to the interactional nature of the speech context. However, since Enfield's study is not grounded on naturally-occurring conversation, it remains to be seen how everyday social interaction shapes the use of *what*-based expressions in conversation, as we do here.

Finally, in the field of Conversation Analysis (CA), there has been substantial research on the non-interrogative nature of question forms, including *what*-based expressions and on repair sequences associated with these expressions. For instance, Schegloff (1997) notices the use of *what*-based expressions in initiating a repair (*what when*), in prompting the advancement of a telling (*where, when*), and as a pre-telling (*Guess what/who (X), Y' know what (X)*) for the speaker to project "the possibility of telling some news or story, contingent on the recipient's response to that prospect" (p. 516). Hayashi et al. (2013) find that speakers use *what do you mean?* to "target a prior turn as a trouble source and, in the process, foreshadows a possible disagreement or dispute with it" (p. 25). Similarly, based on a survey in twenty-one languages (including English, Mandarin, Spanish, French, Dutch, and others), Enfield et al. (2013) find that the use of a question word (usually *what*) is a common strategy for open-class other initiation of repair in everyday conversation. However, most of the *what*-based expressions analyzed here, as well as in studies such as Dingemanse et al. (2014), Dingemanse and Enfield (2015), and K-H Kim (1993, 1999), are used as genuine questions (but for a discussion of the

relevance of Drew (1997), see Section 6.3), and most of these studies do not go further to discuss the extended use of *what*-based expressions in a cross-linguistic context, which is the focus of the current study.

2.2 Mandarin-based studies

Turning now to the literature on Chinese, many grammarians have noted the differences between true interrogative uses and non-interrogative uses. For example, Ding et al. (1961), a well-known early reference grammar, makes a reference to generic and non-referential uses of Chinese interrogatives. Chao (1968: 651–657) states that "all interrogative pronouns, like other interrogative forms, can also be used in the indefinite sense." Chao also points out two other non-interrogative uses of *shenme* 'what.' One is what he calls listing, indicating a sense of 'such things as, things like, and so forth.' The other usage is disapproval, as in 什么睡觉了?! 晚饭还没吃呐 *Shenme shuijiao le?! Wanfan hai mei chi na?!* 'Let alone (=what) sleeping, (we) haven't had dinner yet.'

Other Chinese grammarians have provided detailed descriptions of indefinite uses of interrogatives (Li & Thompson 1981; Lü 1982, 1985; Li 1992), and of their being used to indicate surprise (Zhou 2007), grumbling (Wang & Wang 2003), and inclusivity (Zhu 1982; Lu 1986), where a complete set of reference is denoted as having no exceptions, especially when the interrogative is combined with other adverbs such as *ye* 'also, additionally' and *dou* 'all, completely.' Finally, Tang (1981) is notable in that, in an attempt to capture the range of non-interrogative uses in Chinese, a set of six subcategories is proposed: generic, non-referential, anaphoric, rhetoric, exclamation, and other.

While the Chinese literature surveyed so far is based mostly on isolated and/or made-up sentences, some of the previous studies have begun to explore the value of discourse data. For example, Biq (1990) examines several ways in which *shenme* can be used as a hedge in Chinese conversation, including what she calls interactional hedges (fillers), referential hedges (disclaimers), and expressive hedges (mitigators in negation). A follow-up study can be found in Hsieh (2005), where three types of extended uses of *shenme* in conversation are identified: listing, scaffolding/floor holding, and hedging. Shao and Zhao (1989), on the other hand, find in a corpus eight categories of non-interrogative uses. In a more comprehensive study, Wang and Wang (2003) reveal that, in naturally occurring conversations, *shenme* is rarely used to ask questions; it is more often used to express a sense of negation, doubt, and uncertainty, as well as for such discourse functions as substitution, topic management, and turn taking. While these findings are extremely useful, the size of the corpora and the number of tokens found in these studies are typically small. For example, Wang and Wang's (2003) study is based on a

collection of only 137 cases of *shenme*, so their frequency-related claims need further verification.

A related issue in Chinese studies, as with crosslinguistic studies, is that different terminologies are employed by different researchers, making it difficult to compare. For instance, Hsieh (2005) on one hand cites Biq's (1990) term "hedge" and uses it to categorize those cases in which *shenme* has been grammaticalized as an epistemic marker, while on the other hand labels a sub-category of "hedge" in Biq (1990) – the referential hedge – with another category "listing." In addition, the use of "listing" in Hsieh (2005) is sometimes referred to as "filler." Such non-conforming uses of terms, as mentioned at the beginning of this paper, make it difficult to conduct meaningful comparison both within and across languages.

2.3   Korean studies

The Korean literature is very similar to the Chinese and English literature in that most of the early works are based on isolated sentences while some recent works have a tendency to focus on the process of grammaticalization, using data from diverse spoken and written modes. C-H Kim (2000) and M-H Kim (2005), based on literary works and spoken corpus data (i.e., semi-spontaneous data from television dramas or news and spontaneous data from monologues or natural conversations) respectively, point out that, since the 18th century, Korean interrogatives have evolved into markers of indirect questions, indefiniteness, and, in some cases, discourse markers. In her corpus-based data, M-H Kim finds that 93% of *mwe* 'what' tokens are used as a discourse particle, displaying the speaker's lack of commitment to the proposition of the utterance, mitigating the speaker's assertion, or filling gaps between utterances.

In another study of hers on *mwusun* 'what,' M-H Kim (2006) shows that 26.2% of *mwusun* is used as a discourse marker, either indicating the speaker's uncertainty about referents or to avoid face-threatening actions by softening the speaker's negative tone in the utterance.[1] Using data from interviews and monologues, J-A Lee (2002) similarly contends that *mwe* has evolved from an information-seeking form to a discourse marker with four functions: connecting prior and subsequent utterances, foregrounding information, listing items that the speaker is vaguely thinking of, and filling gaps as a filler word.

There are also studies focusing on the function of interrogative forms in Korean based on the CA approach. K-H Kim (1993, 1999) shows that *mwe* can be

---

1.   In this study, two *what*-like tokens, *mwe* and *mwusun,* in Korean are examined, both of which mean 'what' in English. *Mwe* is an interrogative pronoun, and *mwusun* is an interrogative determiner. Detailed difference of the two tokens is described in 3.2.

used to target specific trouble source in the previous turn, for example, when the speaker is not certain about what the interlocutor has referred to. He also points out that as a strategy for other-initiated repair *mwe* further extends its use to project the speaker's negative actions such as disagreements whereby the speaker not only points to the trouble source of the previous turn but also displays his/her disagreement.

Recently, Koo and Rhee (2013) discuss what they call an emerging paradigm of sentence-final particles of discontent (SFPDs), noting that although the newly emerging sentence-final particles, *-tam, -lam, -kam*, and *-nam*, have undertaken different pathways of grammaticalization, they can all be seen as having originated from interrogative words.

Particularly relevant studies on interrogative forms in Korean discourse include Nam and Cha (2010), and Cha (2010). Nam and Cha (2010), calling *mwe* a discourse marker, examine the non-interrogative uses of *mwe* in the Sejong Spoken Corpus. *Mwe* is identified as serving as a stance-taking marker and displays the speaker's negative stance when it occurs along particular sentence enders at the clause boundary position. On the other hand, when *mwe* occurs freely within one clause, it is seen as being used to list items, give examples, and ask rhetorical questions. While the patterns they reported are worth noticing, their study nonetheless leaves room for further exploration. For example, the authors discarded 40% of the tokens found in their database due to difficulties in categorizing them. By contrast, the current study is corpus-driven and will deal with all instances of the interrogative forms found in our data. In another study, Cha (2010) focuses on factors, prosodic and non-prosodic, that affect the stance-taking (mostly negative) function of *mwe*.

Overall, the Korean studies have made substantive inroads into this area of inquiry, especially concerning how some of the pertinent tokens evolve over time and are distributed across spoken genres. However, the Korean literature, as with Chinese and other studies, is also replete with diverse terminologies and incompatible conclusions, as well as limitations in data sampling size and coding transparency.

Thus, as our quick review shows, while existent research has produced valuable insights for our understanding of the extended uses of *what*-based expressions, a number of drawbacks can be noted. First, crosslinguistic studies, particularly those with interactive discourse data, are rare, as most of the studies deal with individual languages alone. As a result, secondly, there is little consensus as to what categories are to be used to best characterize the range and types of extended uses of *what* and other tokens of interrogative forms. Third, as mentioned earlier, few studies have used large quantities of interactive spoken discourse data, if at all. Even though some have used written corpus data, the quantity and scope

are usually limited. The present study will address these limitations by examining three languages and a large amount of actual spoken discourse samples, which we describe next.

3. Data and methodology

3.1 Data

Data for this study for all three languages come from corpora of natural conversations. For Mandarin Chinese, the CallFriend Mandarin conversational corpus was used. CallFriend consists of 100 unscripted telephone conversations (60 official transcripts plus additional 40 locally transcribed calls), or over 233,000 words, lasting between 5 to 30 minutes each. In each conversation, all parties are native speakers of Mandarin Chinese from Mainland China (Canavan & Zipperlen 1996a). A total of 1,654 occurrences of 什么 *shenme* 'what' were found in the corpus. 34 uncertain hearings, as indicated on transcription, were eliminated. The remaining 1,620 cases were selected for coding.

For Korean, the CallFriend Korean conversation corpus (Canavan & Zipperlen 1996b) was used. Similar to the Mandarin telephone conversation corpus, the Korean CallFriend corpus consists of one hundred telephone conversations between people in close relationships such as family members or close friends. Each call lasts up to thirty minutes. Participants are native speakers of Korean who grew up in diverse regions in Korea. However, they were living in the United States or Canada at the time of the recording. There are approximately 290,000 words in the corpus, and 1,776 instances of *what* tokens (including both *mwusun* and *mwe*) were coded for analysis.

For English, the Switchboard Corpus (Godfrey et al. 1992) was used. This is a collection of telephone conversations among strangers from all regions of the United States. For this study, a total of 2,099 tokens of *what* (1,945) and *whatever* (154) were found in the corpus and coded for analysis.

3.2 Identifying non-interrogative uses and related issues

The focus of the current study is on interrogative forms of *what* expressions that are not meant for the speaker to ask a question and solicit an answer from the addressee. This means that any instances of non-question forms will be included in this study. The basic forms of the *what* expressions are *shenme* in Mandarin Chinese, *what* and *whatever* in English, and *mwe* and *mwusun* in Korean.

In the case of Korean, while both *mwe* and *mwusun* resemble *what* in English, there are morpho-syntactic differences between the two tokens. As an interrogative pronoun, *mwe* can be used as a stand-alone word, but more frequently, it is combined with particles or suffixes and appears in various morphologically combined forms. In this paper, we examine all the morphological variants of *mwe* and their contracted forms (e.g., *mwe-lul* → *mwe-l*, or *mwe-nun* → *mwe-n*). *Mwusun,* on the other hand, is an interrogative determiner and takes a noun at a pre-noun position as in *mwusun umsik cohahay-yo?* what food like-POL 'What food do you like?.'

A clarification about our use of the terms "*what*-based expression" and "interrogative" is in order. While it is most practical to take the word to be the basic unit in identifying the interrogative form in question (*shenme, what/whatever*, and *mwe/mwusun*), we often find that in actual discourse lexical expressions do not occur alone but instead combine with other lexical (e.g. *shenme de* 'what + particle,' in Mandarin Chinese, or *or what* in English) and/or morphological forms (e.g. *mwe-ci* 'what + sentence ender' in Korean). Thus, even though we base our coding decisions essentially on the lexical form of *what* in each of the three languages, we align with many other researchers in taking the theoretical position that the word may not always be taken as the best unit of meaning in language use (see e.g. Sinclair 1991, 1996). Also, even though we may use the term "interrogative" freely, as other authors may have used the term "interrogative pronoun" and such, the target form of this study is actually the non-interrogative uses. The term "interrogative" is used here simply for expository convenience.

A more general issue to be discussed is that, as what might be called a form-based approach, whereby interrogative pronouns such as *what* in English and the equivalents for other languages constitute the basis for our initial investigation, this methodology has both advantages and disadvantages. One of the advantages is that we can identify tokens of interest relatively quickly from corpora of different languages and then investigate their functions. A potential disadvantage is that each form in a particular language may have multiple functions that may not be identified for another language and that some of the conversation or pragmatic functions we identify as extended uses may be carried out by more than one form, rendering the comparison somewhat problematic. However, we believe that for our purpose here the form-based approach is necessary, as for any crosslinguistic comparison based on corpora, a tangible form can always help define the scope of investigation in a straightforward manner, thus serving as a useful starting point from which patterns of convergence and divergence can be further revealed. This of course does not preclude other studies to approach the issue from a different perspective, e.g. an activity- or function-centered one, whereby a wide range of forms can be investigated across multiple languages.

Finally, as mentioned earlier, our coding system is ostensibly corpus-driven, in the sense that we first identify all instances of *what* expressions that are not in the given context used to solicit an answer from the co-participant, and then try to separate these instances into distinct categories. When ambiguity arises, we try to use explicit criteria to distinguish one from another. Through this procedure, our proposed coding system eventually yields eight categories: (1) generic, (2) indefinite, (3) disapproval, (4) general extender, (5) fillers, (6) exclamation, (7) softener, and (8) avoidance. Again, these are by no means a finite set, but merely intended as a starting point for crosslinguistic and/or corpus-based studies of a very elusive discourse grammatical phenomenon. In the following section, then, we define and illustrate each of these categories with examples from the three languages.

4.   The functional coding categories

In this section, we first discuss the eight functional categories using Mandarin and Korean data for illustration. We then discuss the application of these categories to English with respect to two tokens, *what* and *whatever*, as English represents a rather different case from Chinese and Korean (to be detailed in Section 5 and 6).

4.1   Defining the categories (with Mandarin and Korean data as illustrations)

4.1.1   *Generic*
The generic category refers to cases where the *what*-based expression has essentially the same meaning as the standard use of the English pronoun *whatever*, referring to an entire set of membership. Examples (6) from Chinese and (7)–(8) from Korean are cases in point.

> (6)   B: 现在   北京   的话, 什么   都   有.
>   xianzai Beijing dehua, **shenme** dou you
>   now   Beijing TOP   what   all   exist
>   'Nowadays one can find just about anything in Beijing.'

In the following Korean example, *mwusun* indicates *whatever* with the particle *-tun* and the dependent noun *kan* after *iyu* 'reason,' meaning 'whatever reasons.'

(7) A: 그렇게 설그할   때   무슨   이유-든   간-에 잘
         kulehkey selkyoha-l ttay **mwusun** iyu-tun   kan-ey cal
         like.that sermon-RL when what   reason-either DN-LOC well
         경청-을   해-야-지.
         kyengcheng-ul   hay-ya-ci
         listen.carefully-AC do-must-COMM
         'When the priest preaches a sermon, for whatever reasons, (you) must
         listen carefully.'

This is similar to what other linguists have called the "inclusive" use (Zhu 1982;
Lu 1986).

   An example of *mwe* for the generic use is shown below:

(8) A: 뭘   읽어-도 기발-한   생각-은   떠오르-는 것-도
         mwe-l   ilke-to   kipalha-n sayngkak-un tteolu-nun kes-to
         what-AC read-even brilliant-RL idea-TP   come.up-RL thing-even
         없-고.
         eps-ko
         not.have-and
         'Whatever I read, any brilliant idea (regarding studying) never comes up
         in my mind.'

With the particle *-to* 'even' in *ilke-to*, the *mwe-l* combination indicates 'whatev-
er.' The literal translation of *mwe-l ilke-to* is 'even I read everything.' Notice also
that *mwe* here is combined with the accusative particle *-lul* in the contracted
form *mwe-l*.

4.1.2 *Indefiniteness*
In this category, the *what*-based expression essentially has the same meaning as
the indefinite pronoun *something*, referring to entities that are unspecified or un-
known. For example,

(9) B: 你 那边   呢, 随便   注册 一 个 什么   公司.
         ni nabian ne suibian   zhuce yi ge **shenme** gongsi
         2SG that.side PRT randomly register a CLF what   company
         'As for you, (you can) just register a company.'

In Chinese, there is a formulaic construction concerning this usage of *shenme*
'what.' The formula is often realized in the form of "X1 *shenme* X2 *shenme*" and
"*shenme* X1 X2 *shenme*," which basically means 'X2 whatever X1.' For example, 你
想吃什么就买什么 *ni xiang chi **shenme** jiu mai **shenme*** '(You can) buy whatever
you desire to eat' (Li & Thompson 1981: 530). The formulaic use of *shenme* in this
construction is coded as a special case of the indefinite use.

Sometimes the generic and the indefinite uses may not be easily teased apart. The general rule of thumb is that unless there is some element indicating a general scope of reference, the *what* form will be considered indefinite rather than generic. Thus, the use of *shenme* in the following example is judged to belong to the indefiniteness category rather than the generic category as the scope of the reference is limited by the use of the modifying element *zhengqian* 'money making,' which limits the scope of the reference to this particular domain rather than keeping it wide open.

(10) A: 现在　就 是 什么　挣钱　　做 什么　呀
          xianzai jiu shi **shenme** zhengqian zuo **shenme** ya
          now   EM is what   profitable do what   PRT
          'Nowadays (people) do whatever that makes (more) money.'

In Korean, *mwusun* can also be used to indicate a sense of indefiniteness, as the following example illustrates.

(11) A: 무슨　　애기-를 하-다-가　이제 케이터링하-는　애기-가
          mwusun yayki-lul ha-ta-ka   icey keyithelingha-nun yayki-ka
          what   talk-AC do-DC-and now catering-RL   talk-SBJ
          나-왔-어.
          nao-ass-e
          start-PST-INT
          '(We were) talking about something, and then (we) started to talk about the catering thing.'

The following example shows a case of indefinite use of *mwe*, indicating *something*:

(12) A: 근데　　뭐 좀　　마시-고　일하-지?
          kuntey   **mwe** com masi-ko   ilha-ci?
          by.the.way what a.little drink-and work-COMM
          'By the way, why don't we drink something and work?'

4.1.3   *Disapproval*

Moving on to the disapproval use, in this category the *what*-based form indexes the speaker's unfavorable opinion or negative stance towards the proposition expressed in the utterance. We distinguish two types of disapproval: direct disapproval vs. indirect disapproval. In direct disapproval, the interrogative token is attached to the lexical item directly indicating the speaker's negative stance. In indirect disapproval, by contrast, the *what* form is used in a doubting tone, with functions similar to a rhetorical question. Let's start with cases of direct disapproval.

In Mandarin Chinese, direct disapproval is often realized as *shenme* preceding a repeated element, as in (13):

(13) A: 两千   块 钱   对 吧?
        liangqian   kuai qian   dui ba
        two.thousand CLF   money right PRT
        'Two thousand (RMB), right?'

      B: 什么 两千,   三千   多, 现在 是.
        shenme liangqian   sanqian   duo xianzai shi
        what   two.thousand three.thousand more now   is
        'Not "two thousand" at all. (It costs) over three thousand now!'

In this example, Speaker B repeats the word *liangqian* 'two thousand,' which is originally uttered by Speaker A, and prefixes it with *shenme* to show a negative stance over the proposition expressed in the phrase *liangqian* 'two thousand.' *Shenme* is used to quote Speaker A's utterance and to pinpoint the part of the proposition that Speaker B disagrees with, i.e., not "two thousand" but "three thousand."

    Similarly, Korean *mwusun* contradicts the previous speaker's utterance with a repeated use of the problematic target element, which is to be disagreed with. In (14), previously Speaker A said that she would go back to Korea in December; and, in this excerpt, she says that she has less than three months left in America. Speaker B shows her disapproval of the way Speaker A calculates the time frame with a token of *mwusun* highlighting the wrong date ('the fifteenth') that was brought up by Speaker A.

(14) A: 삼   개월-도   안 되-지.   지금-부터-는.
        sam kaywel-to   an toy-ci   cikum-pwuthe-nun
        three month-even not become-COMM now-from-TP
        'I have less than three months from now.'

      B: 그런-가?
        kule-n-ka
        that-RL-Q
        'Is that so?'

      A: 그치. 지금 벌써 구-월   보름-이-니까.
        kuchi cikum pelcce kwu-wel   polum-i-nikka
        yes   now   already nine-month fifteenth-be-becuase
        'Yes, because it's already September the fifteenth.'

      B: 무슨   보름-이-냐.
        mwusun polum-i-nya
        what   fifteenth-be-INT
        'What "fifteenth".'
        지금 오늘-이 지금 구 월   십 일-이-구만.
        cikum onul-i   cikum kwu wel   sip il-i-kwuman
        now   today-SBJ now   September ten day-be-UNASSIM
        'Now, today is, now, it is September tenth.'

A: 응. 그렇구나.   삼   개월-도   안 되-지.
    ung kuleh-kwuna   sam kaywel-to   an toy-ci
    yes like.that-UNASSIM three month-even not become-COMM
    'Yes, I see. (I have) less than three months.'

The following example shows the disapproval use of *mwe*.

(15) B: 무섭게   생겼어.
          mwusepkey sayngky-ess-e
          scary   formed-PST-INT
          'Your face was scary-looking (in the video).'
      A: 허허허
          hehehe
          'HAHAHA'
          뭐-가   무섭게   생겼어.   원래   그렇게
          mwe-ka mwusepkey sayngky-ess-e   wenlay   kulehkey
          what-SBJ scary   formed-PST-INT originally like.that
          생겼지.
          sayngky-ess-ci
          formed-PST-COMM
          'I'm not scary-looking. I've always looked like that.'
      B: 몰라.   아니야.
          molla   ani-ya
          not.know no-INT
          'I don't know. No. (You looked different in the video.)'

      ((B specifically points out changes in A's appearance))

The speakers in this conversation are cousins who are living in different parts of
America. Prior to this excerpt, Speaker A asked whether Speaker B saw the video
of A's family that A sent to B. In the exchange the focal point, *mwusepkey sayn-
gkyesse* 'your face was scary-looking', is recycled from B's previous assessment and
repeated in A's disagreement marked by *mwe-ka*.

    If the direct type of disapproval cases can be said to be blunt, the more in-
direct type of disapproval mostly appears as some doubt being cast, much like a
rhetorical question but without necessarily reducing the negative force implied. A
Mandarin example of an indirect disapproval can be seen in (16),

(16) 1. A: 这个 吃苦   么 这个 不必   谈
              zhege chiku   me zhege bubi   tan
              this bear.hardships PRT this   no.need talk
              'There is no need to talk about bearing hardships.'

2. 这个 有 什么 好
   zhege you **shenme** hao
   this have what good
   'What good is this (i.e. talking about suffering)?'
3. B: 那 我 这个 我 不 把 他 摆
   na wo zhege wo bu ba ta bai
   then 1SG this 1SG NEG PRT 3SG place
4. 给 你们, 我 给 哪 个 呐?
   gei nimen wo gei na ge na
   to 2PL 1SG to which CLF PRT
   'If I had not given him to you (i.e., leaving my son with you and asking you to raise him), whom else could I have given him to?"

In this example, Speaker B, the son of Speaker A, is working in North America and has left his own son with his parents (Speaker A) in China. Speaker A is very unhappy and is complaining about the tremendous effort it takes to take care of the grandson. By saying 'what good is this (i.e. talking about suffering)?' (line 2), the grandfather implies that there is no need to talk more about "suffering" and that the most important thing is to stop the "suffering" itself, to which the son (Speaker B) responds by saying that he has no choice but leaving the child with him.

The Korean counterparts are illustrated below.

(17) B: 무슨 소리 하-는-거-야= 네-가 잘 몰라서 하-는-
   mwusun soli ha-nun-ke-ya ney-ka cal molla-se ha-nun
   what sound do-RL-thing-INT you-SBJ well not.know-so do-RL
   'What are you talking about? You're saying like that because you don't know really know (about Korea).'
   한국-이 훨씬 더 문란하다니까 지금,
   hankwuk-i hwelssin te mwunlanha-ta-nikka cikum
   Korea-SBJ much more disorderly-DC-you.know now
   'What I'm saying is Korea is too much these days.'

Previously, Speaker A, a graduate student in America, said that she wanted to go back to Korea after all because she thought American people were "too much" for her. In response, in the excerpt, Speaker B problematizes A's entire talk by asking the rhetorical question 'What are you talking about?', casting his doubt that Korea is any better than America. Note that this kind of disapproval in the form of a rhetorical question often appears in Korean as a formulaic chunk. Other formulaic expressions include *mwusun mal-ha-nya?* what talk-do-INT, *mwusn soli-ya?* what sound-INT, *mwusun yayki-ha-nun ke-ya?* what story-do-RL thing-INT and other equivalent expressions, meaning 'What are you talking about?'.

Clearly, the difference is a matter of degree, so in our coding, both types are lumped under the same category.

4.1.4   *General extender*

The term general extender comes from Overstreet and Yule (1997), which refers to expressions such as 'and so on,' 'etcetera,' 'or something' in English (see also Overstreet 1999, 2005). These expressions, typically used to exemplify a set of like-items, are also variably called "summarizing phrases" (Crystal & Davy 1975: 113), "generalized list completers" (Lerner 1994), and "vague category identifiers" (Channell 1994).

What expressions can be used in such a fashion to extend the scope of a list of items (Jefferson 1990). For this study, we define a list as having minimally two items expressed or, where there are fewer than two items expressed, there must be some special formula indicating that the items are of a listing nature. In this context, the *what*-based expression may either precede (for initiation) or follow (for extending) the list.[2] In the Mandarin example below, *shenme* precedes the list:

(18) B: 营养   好 一点 的, 什么   鱼 呀, 鸡   呀.
         yingyang hao yidian de **shenme** yu ya ji   ya
         nutrition good a.little NOM what   fish PRT chicken PRT
         '(Eat some food that is) nutritious, such as fish, chicken, and so forth.'

Similar cases can be found in Korean. Below are instances of *mwusun* in medial-list position, and *mwe* in pre-list positions respectively:

(19) A: 영어,   스페인어, 독일어, 일본어, 프랑스어, 이집트어, 한국어,
         yenge supheyine tokile ilpone phulangsue iciptue   hankwuke
         English Spanish   German Japanese French   Egyptian Korean
         중국어,   베트남어, 그 다음 무슨   힌두어 이런 몇
         cwungkwuke peythuname ku taum **mwusun** hintwue ilen   myech
         Chinese   Vietnamese that next what   Hindi   these several
         언어-들-에   대해서   뭐   한-대.
         ene-tul-ey   tayhayse mwe   ha-n-tay
         language-PL-at regarding something do-RL-hearsay
         '(They said that they will) do some (research on) several languages such as English, Spanish, German, Japanese, French, Egyptian, Korean, Chinese, Vietnamese, and then what Hindi.'

---

2. For our purpose here, we are not making a fine distinction between initiating a list and completing a list, though such a distinction may be interactively important, as some of the cited literature have shown.

Here, the speaker lists relevant languages pertinent to the topic of a research project, and *mwusun* occurs in the middle of the listed languages. In the example below, the speaker lists a couple of big cities to highlight the cleanliness of Austin, and *mwe* comes before the list starts.

(20) A: 대도시 뭐   엘에이, 필라델피아 같-은   데-보다   훨씬
        taytosi **mwe** eyleyi   phillateylphia kath-un tey-pota   hwelssin
        big.city what LA   Philadelphia like-RL place-than much
        깨끗하-지.
        kkaykkusha-ci
        clean-COMM
        '(Austin is) much cleaner than big cities such as what LA and
        Philadelphia.'

4.1.5   *Fillers*

In everyday conversation, speakers may for various cognitive and interactive reasons take time to gather their thoughts, search for a particular word, and dynamically formulate their expressions (see, e.g., Goodwin 1979), resulting in gaps between turns and/or utterances. In the process, various items may be used to fill up the gap. Speakers can do so with the *what* expression serving as a place holder for trouble sources, word searching, or formulation, a function we call the filler function. Note that the entity that the place holder refers to may or may not appear in the subsequent context, and the entity designated by the *what* form may be either a nominal or a clausal unit.

   In the following examples, the entities named by the *what* expression are given in a later context. The entity is a noun.

(21) B: 他 让 你 留 下   那 本
        ta   rang ni liu   xia   na ben
        3SG ask 2SG leave COMP that CLF
        'He asked you to leave that,'
        就是 那个 什么   签 签证   的   那 指南 啊 什么
        jiushi nage **shenme** qian qianzheng de   na zhinan a   shenme
        EM   that what   vi- visa   ASSO that guide PRT what
        'that what, that Visa Application Guide.'

In this case, the nominal entity is *qianzheng zhinan* 'visa application guide', which comes right after the place holder *shenme*.

   The entity can also be an action denoted by a verb phrase. For example,

(22)   A: 他们　经常　　　那个, 什么,　在 这边儿, 用　　那个, 遥控
　　　　　tamen jingchang nage **shenme** zai zhebianr yong nage yaokong
　　　　　3PL   often   that what at here   use that remote.control
　　　　　照相　　　呢
　　　　　zhaoxiang ne
　　　　　take.photo PRT
　　　　　'They often (do) that what/thing here – using the remote control
　　　　　(function) to take a photo.'

Shenme appears here after the adverb *jingchang* 'often' and before the verb phrase 'taking remotely controlled photographs.'

In the Korean example below, the speaker corrects the word 'phone-call' to 'phone' using *mwe* without any identifiable error or reason. *Mwe* fills the gap between the initial word 'phone-call' and the subsequently corrected form 'phone.'

(23)   A: 그래서 전화통화-도　　　뭐　전화-도　자주　　　하-고
　　　　　kulayse cenhwathonghwa-to **mwe** cenhwa-to cacwu   ha-ko
　　　　　so   phone.call-also   what phone-also frequently make-and
　　　　　이러-더라-고.
　　　　　ile-tela-ko
　　　　　do.this-RT-QT
　　　　　'So (they) make a phone call what frequently make a phone (call to each
　　　　　other) and the like.'

Sometimes the speaker has trouble recalling a named entity. In the following example, the speaker is having trouble recalling a personal name.

(24)   B: 叫 黄　　什么　　的　　夫妇
　　　　　jiao huang **shenme** de   fufu
　　　　　call Huang what   ASSO couple
　　　　　'The couple who is called Huang something.'

Tokens of *mwusun* and *mwe* in this category are often used in tandem with single demonstrative elements such as *i* 'this,' *ku* 'that,' *ce* 'that over there,' and related demonstrative combinations such as *i-ke(s)* 'this thing,' *ku-ke(s)/ce-ke(s)* 'that thing,' *ilen* 'this kind of,' or *kulen/celen* 'that kind of.' Other collocations include formulaic searching expressions such as *ceki mwe-ya* there what-INT 'What is that?,' *mwe-tela* what-RT 'What was that?,' *mwe-la-l-kka* what-QT-RL-INT 'What should I explain?,' and *mwe-ci* what-COMM 'What is that?.'

4.1.6   *Exclamation*
What expressions can also be used to indicate the speaker's intense emotions in disbelief, surprise, incredulity, and the like. In this function, the *what* expression is

often used in response to a statement made by the prior speaker, indicating that the current speaker finds the previous speaker's statement extremely unbelievable, incredible, or questionable. It is usually used in a freestanding format, and in noticeably different prosodic patterns such as a high volume, a high pitch, and/or a rising intonation contour (cf. Selting 1996 for the role of prosody in initiating repairs and in expressing emotional responses such as "astonishment" and "surprise").

Below is an example of a freestanding form from Mandarin:

(25)　B:　他　死　了.
　　　　　ta　si　le
　　　　　3SG die COMP
　　　　　'He died.'
　　　　A:　什么?
　　　　　shenme
　　　　　what
　　　　　'What?'

In Korean, only *mwe* but not *mwusun* can be used for this function in our data. The following is a case of *mwe*:

(26)　A:　야 근데　　너-의　리얼 레터-는　지금　누가　　읽-고
　　　　　ya kuntey　ne-ui　liel leythe-nun cikum nwuka　ilk-ko
　　　　　hey by.the.way you-POSS real letter-TP　now　somebody read-and
　　　　　있-을 거-야.
　　　　　iss-ul ke-ya
　　　　　be-RL thing-INT
　　　　　'Hey, by the way, the letter you wrote to me is being read by somebody else now.'
　　　　B:　뭐? 못　받-았-어?
　　　　　mwe mos pat-ass-e
　　　　　what not　receive-PST-INT
　　　　　'What? You haven't received (my letter)?'
　　　　A:　못　　받-았-다-는　　말-이-야.
　　　　　mos　pat-ass-ta-nun　mal-i-ya.
　　　　　cannot receive-PST-DC-RL talk-be-INT
　　　　　'I mean I haven't received (your letter).'

Here, Speaker B shows her surprise when Speaker A reveals the news of the wrong delivery of B's letter. B increases the pitch and loudness when producing *mwe*.

4.1.7   *Softener*

What expressions can also be used as a down-toner (Quirk et al. 1985) to soften a statement or to show the speaker's non-committal epistemic stance or attitude towards the statement introduced in conjunction with the *what* expression. For example,

(27)   B: 工资 少, 也 没有 什么 奖金.
       gongzi shao ye meiyou **shenme** jiangjin
       salary little also NEG what bonus
       'The salary is low, and there isn't much of any bonus either.'

In this case, the reference to bonus is prefixed with *shenme* 'what' indicating the speaker's less enthusiastic attitude toward the meager salary situation.

The Korean examples below show a similar connotation:

(28)   A: 올 때 애들-이랑 다 연락하-고 왔-어?
       o-l ttay aytul-ilang ta yenlakha-ko o-ass-e
       come-RL when friends-with all call-and come-PST-INT
       'When you left for (Canada), did you call all of your friends (to say goodbye)?'
       B: 삼-개월 어학연수 가-는데 무슨 뭐 그럴
       sam-kaywel ehak.yenswu ka-nuntey **mwusun mwe** kule-l
       three-month language.program go-but what what do.like.that-RL
       필요 없-잖아.
       philyo eps-canha
       necessity not.exist-you.know
       'Well, you know, going for a language program for only three months, what, it doesn't seem necessary (to call all people up to say goodbye).'

Speaker B shows his disaffection toward the farewell call but softens his opposition by coloring the statement with both *mwe* and *mwusun*.

In Korean, *mwe* in particular can display the speaker's non-committal attitude as well as softens a possible bragging tone, as indicated by the following example.

(29)   A: 요새 비지니스-는 좀 어떠-세-요?
       yosay picinisu-nun com ette-sey-yo
       these.days business-TP a.little how-HON-POL
       'How's your business going these days?'
       B: 뭐 여긴 뭐 이 회사-는 꾸준히 잘 되-죠.
       mwe yeki-n **mwe** i hoysa-nun kkwucwunhi cal toy-cyo
       what here-TP what this company-TP constantly well run-COMM
       'Well, so far the company has been running rather well actually.'

In some cases, the *what* expression introduces (apparently) reported materials, and the indirectness conveyed by the reported speech helps reduce the speaker's epistemic authority (Heritage 2012) and distance the speaker somewhat from the proposition expressed in the statement. This sub-category is coded as a special case of softener.

In the Korean example below, the speaker roughly reports the utterance made by the daughter, and the exact wording of "what should I do" is thus cast as being in doubt.

(30) A: 딸내미-가 돈   없-다-고   뭐   어떡하-지 그랬-더-니,
      ttalnaymi-ka ton   eps-ta-ko   **mwe** ettekha-ci kulay-ss-te-ni
      daughter-SBJ money not.have-DC-QT what how-COMM say-PST-RT-then
      'The daughter said something like "What should I do?" because she did not have money.'

The *what* expression can also function to show the speaker's uncertainty and approximate estimation, either about the point that the speaker is making or something else, e.g. estimation of numbers (ages, sizes, etc.). This is yet another sub-category of "softener." For example,

(31) A: 可   可能   一 个 星期 什么   没有   信   什么
      ke   keneng yi ge xingqi **shenme** meiyou xin   shenme
      may- maybe one CLF week   what   NEG   letter what
      'It's possible that there isn't any letter in a week or so.'

The following Korean examples show uncertainty and proximate estimation.

(32) A: 도시-가 무슨   우리나라 남대문-서부터   명동   바닥
      tosi-ka **mwusun** ulinala   namtaymwun-sepwuthe myengtong patak
      city-SBJ what   Korea   Namdaemun-from   Myengdong area
      정도   밖에   안 되-니까,
      cengto   pakkey an toy-nikka
      approximately only   not become-so
      'The (size of Montreal) city is, what, approximately only the size from Namdaemun market place to Myengdong area in Korea so.'

(33) A: 한   뭐   여섯 일곱 시간 정도   자-는데,
      han   **mwe** yeses ilkop sikan cengto   ca-nuntey
      around what six   seven hour approximately sleep-but
      '(I) sleep around, what, approximately six or seven hours (a day).'

4.1.8   *Avoidance*

The last category of extended uses in our coding system is what we call the avoidance use, where the *what* expression stands in for something that the speaker appears to avoid mentioning due to various undesirable connotations.

In the following Mandarin example, Speaker A, a former teacher in her home country, is now working as a part-time housekeeper in North America. The word being avoided here in the second line and repeated in the third (translated as 'something/someone') is "housekeeper," a job that some Chinese speakers may not consider very highly, as implied in the conversation below.

(34)   A: 也 不 象   以前   在 国内.
        ye bu xiang yiqian   zai guonei
        also NEG like   previously in within.the.country
        'Unlike in (our) home country,'
        你 要   在 国内   想   出去 当   个 什么.
        ni yao   zai guonei   xiang chuqu dang   ge **shenme**
        2SG COND in country.in want go.out become CLF what
        'if you want to be 'something/someone' (like this),'
        呃呀, 什么   当   一 个 都 觉得 挺   不好意思的.
        eya   **shenme** dang   yi ge dou juede ting   buhaoyiside
        PRT   what   become one CLF PRT feel   quite embarrassed
        'well, you would feel embarrassed.'
        在 这 好象   大家   都 习以为常   了.
        zai zhe haoxiang dajia   dou xiyiweichang   le
        at here seem   everybody PRT be.accustomed.to PFV
        '(But) here (in this country) nobody feels bad about it (lit. everybody is
        so used to it).'
        好象   谁   都 要   从 这 地方 开始 似的.
        haoxiang shui   dou yao   cong zhe difang kaishi shide
        as.if   everyone all   have.to from this place   start   as.if
        'As if (they know that) everyone starts from nothing.'

In Korean, only *mwe* is found to be used for the avoidance use. In the following example, Speaker A blames the hearer for not taking her phone calls by emphasizing the number of calls that she has made. However, A avoids explicitly complaining about the hearer and, instead, implies her negative stance with a use of *mwe* in reference to her calls:

(35) A: 몇   번-이나   전화했-는데   완전   애브리타임
          myech   pen-ina   cenhwahay-ss-nuntey wancen   aypulithaim
          how.many time-as.many call-PST-but   completely every.time
          뭐.
          mwe
          what
          '(I called you) many times but (you) every time completely what
          (ignored my call).'

We have now described the eight general categories of extended use, with some having subcategories. A special note about repeated tokens: in handling repeated tokens, each of the repeated tokens is counted as an independent instance. A repetition is defined as a *what* expression uttered at least twice without any other intervening materials between the two, or appearing in disjointed units that are in close proximity, typically in the same turn, as exemplified by the Mandarin example just discussed in (34).

4.2 Applying the categories to the English data

English differs from both Chinese and Korean in that it has two related tokens: *what* and *whatever*, with the latter being a compound form normally serving a generic and exclusive function that is carried out in Chinese and Korean by single *what* tokens. For this reason, we conducted coding separately for these two tokens.

In addition, both *what* and *whatever* in English can serve as a relativizer in a relative clause construction. Just as with true interrogative uses, such relativizing uses are excluded from the extended use category, although, as the literature has indicated, such a grammaticalized use is usually deemed an extension of the interrogative use (Heine & Kuteva 2006).

For English data coding, we have used the same categories discussed in 4.1.1 through 4.1.8. Some quick examples are given below without further discussion.

(1) Generic: referring to the entire set of membership.

(36) but I think it is. I think it'd show up no matter **what**. If you doing it, then it's going to show up.

(37) and if he has to learn just by kinesthetic, we're supposed to teach **whatever** way that that child has to have it, learns the best way.

(2) Indefinite: referring to entities that are unspecified or unknown.

(38) And I don't know if that's just a pure volumes number or, or **what**, but, uh. Sometimes I think the jury is ignorant in the facts of law.

(39)   figured on a **whatever** basis how much it costs to actually support them for a year.

(3)   Disapproval: expressing the speaker's unfavorable opinion. Most of the examples found in the data are of the indirect, rhetorical type.

(40)   it's like, you know, God, **what** are they doing.

(41)   and I said, **what** is with this mower, I can't even push it around the yard.

(4)   General Extender: reference extended to unspoken items.

(42)   So our stuffing, was, um, lot of times was, uh, like sausage or **what** not.

(43)   there were always E P A people and **what** not were always telling us that, uh, farm chemicals and **what** not were destroying our water system and all that.

(5)   Fillers: filling in the gap in speech.

(44)   Well, I tell you **what**, I'm a kind of, uh, a history nut. I'm trying to think back now,

(45)   He was, he was a, a, **what** do you call it, abuser.

(6)   Exclamation: reaction to something unbelievable or incredible.

(46)   A: What would you have these people do if they were brought in?
        B: **What**, into the service?
        A: Well, yeah.

(47)   B: so he, uh, he had to stay home.
        A: Oh, **what** a shame.
        B: Yeah, yeah, it really was.

(7)   Softener: to downplay a statement or reduce authority.

(48)   A: and many have not been tested yet.
        B: Hm, well, it's only been **what** a year? Two years?
        A: Uh, a couple of years, yeah.
        B: Huh-uh

(49)   A: so that was part of, I guess, his character building or **whatever**, but,
        B: I think it adds a, probably adds a little bit of depth.

(8)   Avoidance: substituting something sensitive or negative. No examples are found.

4.3 Summary

In this section we discussed the eight functional categories of the extended uses of *what* forms. Our taxonomy is derived from what were found in the corpora and hence is corpus-driven. The next section will describe some general tendencies observed in the crosslinguistic corpus data.

5. Findings

Our report of the findings will start with several general distribution patterns, comparing typical interrogative uses with their extended uses.

5.1 Overall findings

Table 1. Interrogative vs. extended uses in Chinese and Korean

	Interrogative	Extended uses	Total
Mandarin Chinese	344 (21%)	1276 (79%)	1620 (100%)
Korean (combined)	241 (14%)	1535 (86%)	1776 (100%)
mwusun	118	268	386
mwe	123	1267	1390

In both the Mandarin and Korean cases, the non-interrogative uses constitute the vast majority. That is, in Mandarin, the true interrogative function of *shenme* accounts for only about 1/5 (21%) of its use, while the non-interrogatives take up 4/5 (79%). In Korean, this tendency is even stronger: 86% of the cases are non-interrogative. This confirms Wang and Wang's (2003) assertion about the abundance of extended uses of the interrogative forms for Chinese.

In contrast with those in Mandarin and Korean, *what* forms in English have much less extended use (13%). As Table 2 shows, in English, the relativizer function (relative clause: RC) makes up the majority of *what* and *whatever*.

Table 2. Interrogative vs. extended uses in English

	What (N)	*Whatever* (N)	Total (N)	%
Interrogative	573	0	573	27.3%
RC	1129	56	1185	56.5%
Extended	183	97	280	13.3%
Other	60	1	61	2.9%
Total	**1945**	**154**	**2099**	**100%**

5.2 Subcategories of extended uses

Turning now to the specific categories of the extended uses, Tables 3–4 show the subcategories of extended uses in the three languages.

Table 3. Distribution of subcategories of extended uses in Mandarin and Korean

	Mandarin (N)	Mandarin (%)	Korean (N)	Korean (%)
(1) Generic	89	7.0%	2	0.1%
(2) Indefinite	269	21.1%	184	12.0%
(3) Disapproval	103	8.1%	137	8.9%
(4) General Extender	281	22.0%	78	5.1%
(5) Fillers	332	26.0%	339	22.1%
(6) Exclamation	4	0.3%	2	0.1%
(7) Softener	188	14.7%	784	51.1%
(8) Avoidance	10	0.8%	9	0.6%
Total	**1276**	**100%**	**1535**	**100.0%**

In Table 3, four categories in Mandarin stand out. They are (2) Indefinite, (4) General Extender, (5) Fillers, and (7) Softener. These categories, each constituting over 10% of the instances found in the corpus data, are quite close to each other in frequency. Nevertheless, the hierarchical order in Mandarin appears to be the following:

> Fillers (26%) > General Extender (22%) > Indefinite (21.1%) > Softener (14.7%)

In Korean, three categories constituting over 10% of the examples stand out and their order seems to be:

> Softener (51.1%) > Fillers (22.1%) > Indefinite (12.0%)

Unlike in Chinese, softener uses in Korean make up a large chunk of the data.

For English, we need to separate the two tokens and count them independently first. The results show, perhaps not surprisingly, that *what* and *whatever* exhibit different tendencies.

Table 4. Subcategories of the extended uses in *what* and *whatever* in English

Category	WHAT (*N*)	WHAT (%)	WHATEVER (*N*)	WHATEVER (%)	Total (*N*)	Total (%)
(1) Generic	12	6.6%	59	60.8%	71	25.4%
(2) Indefinite	3	1.6%	1	1%	4	1.4%
(3) Disapproval	31	16.9%	0	0%	31	11.1%
(4) General Extender	15	8.2%	14	14.4%	29	10.4%
(5) Fillers	96	52.5%	0	0%	96	34.3%
(6) Exclamation	20	10.9%	0	0%	20	7.1%
(7) Softener	6	3.3%	23	23.7%	29	10.3%
(8) Avoidance	0	0%	0	0%	0	0%
Total	183	100%	97	100%	280	100%

As Table 4 shows, in the case of *what*, the three prominent categories and their hierarchy are:

> Fillers (52.5%) > Disapproval (16.9%) > Exclamation (10.9%)

In the case of *whatever*, the three prominent categories and their hierarchy are:

> Generic (60.8%) > Softener (23.7%) > General extender (14.4%)

If we lump these two tokens together, the resulting hierarchy is:

> Fillers (34.3%) > Generic (25.4%) > Disapproval (11.1%) > General Extender (10.4%) > Softener (10.3%)

The following figure summarizes the distributional patterns across the three languages based on the consolidated data just presented.

Chinese: Fillers (26%) > G. Extender (22%) > Indefinite (21.1%) > Softener (14.7%)
English: Fillers (34.3%) > Generic (25.4%) > Disapproval (11.1%) > G. Extender (10.4%) > Softener (10.3%)
Korean: Softener (51.1%) > Fillers (22.1%) > Indefinite (12.0%)

Figure 1. Overall tendencies in Chinese, English, and Korean

This shows that Mandarin Chinese and English have remarkable similarities in that three out of the top four categories are identical, and they also have relatively similar hierarchical orders. On the other hand, we have also seen that Mandarin Chinese and Korean are similar in that the extended uses predominate over the interrogative uses. We will try to explain this and other patterns in the next section.

6.   Discussion

Given the patterns described so far, the following general issues need to be addressed: similarities and differences among the languages and general development tendencies.

6.1 Chinese/Korean vs. English

As we have seen in the previous section, a large number of Mandarin Chinese and Korean *what* expressions (between 79–86%) have non-interrogative (and non-grammatical) uses, while in English only a mere 13.3% do. We believe that this vast discrepancy can be attributed to the fact that in English *what* and *whatever* are grammaticized for a special grammatical function: relativizing. This function has been treated as the last phase of grammaticalization in what Heine and Kuteva (2006) describe as a four-step development:

marking word questions > introducing indefinite complement or adverbial clauses > introducing definite complement or adverbial clauses > introducing headed relative clauses

Given this pervasive grammatical function, even if the tokens keep evolving into a pragmatic marker, as both Brinton (1996) and Traugott (1982) as well as Quirk et al. (1985) have described, their pragmatic use is still a minor one in terms of frequency of occurrence.

Mandarin Chinese and Korean, on the other hand, do not use *what* expressions to serve as the head (or marker) of a relative clause and, as a result, *what* expressions are free to develop into the pragmatic domain with much higher frequency than in English. The fact that English has a compound form in *whatever* also shows that English focuses much more on the inclusivity distinction as well as the grammatical functions than the other two languages.

6.2 Chinese/English vs. Korean

Earlier, in Section 5.2, we showed that Chinese and English are similar in that they share three of the top four extended subcategories, and they further share a relatively similar hieratical order: filler on top, generic and general extender in the middle, and softener rounding up the top forms. Korean, on the other hand, has an unusually high concentration of softeners, which make up over 50% of all the extended uses. Our speculation is that this is due to the availability of other devices in Chinese and English while Korean relies heavily on the interrogative form in serving the top ranked extended functions shown in Figure 1.

When we examine the Korean softener uses of the *what* forms, we can see that they serve a wide range of functions in addition to the softening function, which we may call "softening-plus" functions. There are discourse uses of *mwe* used to show discourse meanings (e.g., something similar to *well* or *by the way* in conversational English) in the course of conversations. *Mwe* may or may not co-occur with related discourse particles or conjunction words, mostly at the turn-initial or secondary positions. When used with these particles or conjunctions, *mwe* conveys similar discourse meanings to the co-occurring discourse particles or conjunction words. As reviewed earlier, Nam and Cha (2010) find similar patterns involving *mwe* working together with other discourse particles and conjunction words. The following are some of the examples from our corpus:

a.   *Mwe* with collocations:

(50)   *Mwe* functioning as 'well' in conjunction with the discourse particle *kulccey* 'well':

A: 왜 이렇게 다운   됐-냐?
    way ilehkey tawun   tway-ss-nya
    why like.this depressed become-PST-INT
    'Why are you so depressed?'
    몸   때문에   그렇-겠-지.
    mom   ttaymwuney kule-keyss-ci
    body.condition due.to   such-DCT.RE-COMM
    'I guess it is due to your body condition.'
B: 뭐 글쎄 꼭   몸   때문에   그러-나.
    mwe kulccey kkok   mom   ttaymwuney kule-na
    what well   exactly body.condition due.to   such-Q
    'Well, it is not only because of my body condition (but also because of my graduate school life).'

(51)   *Mwe* functioning as 'by the way' in conjunction with the linking word *kuntey*:

A: 근데   뭐 혜경이   유치원   보낸-다-며?
    kuntey   **mwe** hyeykyengi yuchiwen   ponay-n-ta-mye
    by.the.way what NAME   kindergarten send-RL-DC-hearsay
    'By the way, I heard that you are going to send Hyekyung to a kindergarten, aren't you?'

In (51), the conjunction word *kuntey* 'but' has a discourse meaning of 'by the way,' and *mwe* also helps to convey the same discourse meaning.

(52)   *Mwe* functioning as 'by the way' together with the interrogative form *ettekey* 'by the way':

A: 어떻게 뭐 밥-은   항상   한식   해 먹-냐?
     ettehkey **mwe** pap-un   hangsang hansik   hay mek-nya
     how   what meal-TP always   Korean.style do eat-INT
     'By the way, do you always eat Korean style meals?'

b.   *Mwe* without collocations:

(53)   *Mwe* functioning as a discourse marker 'well' (Schiffrin 1986):

A: 그러면   너 어떡할 거-야?
     kulemyen ne   ettekha-l ke-ya
     then   you how-RL thing-INT
     'Then, what are you going to do (after you graduate)?'
B: 뭐 어플라이-를 해 볼   거-야.
     mwe ephullai-lul   hay po-l   ke-ya
     what apply-AC   do try-RL thing-INT
     'Well, I will try to apply (for graduate schools).'

(54)   *Mwe* indicating 'by the way' or signaling a topic shift:

A: 아휴   아…..
     ahyu   ahhhh
     'Sigh. Ahhhhh.'
B: 그렇게   답답해?
     kulehkey   taptap-hay
     that.much stuffy-INT
     'Are you that much worried about (your school life in general)?'
     (4.3 seconds pause))
B: 뭐   너네 이제 미드텀   다 끝났-지?
     mwe neney icey   mituthem ta   kkuthna-ss-ci
     what your now   midterm all over-PST-COMM
     'Now your midterm is all over?'
A: 미드텀   없었어.
     mituthem eps-ess-e
     midterm   not.have-PST-INT
     'I did not have a midterm.'

In (54), Speaker B uses *mwe* at the initial position of the utterance to gain the recipient's attention before starting a new topic in conversation.

The sample functions just illustrated in Korean are often carried out in Chinese and English with other forms. For example, in English, discourse markers such as *well*, conjunctions such as *so, then, and*, and even *because* ('cause) can be used to perform some of the functions that the Korean *mwe* or *mwusun* would.

In Mandarin Chinese, similar tokens such as the demonstrative/conjunction *na* 'that, then,' turn initiators such as *en, o/ou* (Tao 1996), other conjunctions such as *ranrou* 'then,' *jieguo* 'then' (Tao 2011), and *yinwei* 'because' (Song & Tao 2009) can do similar work.

In other words, while Chinese and English have more options to share with the *what* form in performing the softener function, Korean uses the *what* form as the go-to device for a wide range of functions alongside the softener function. Although Korean does not lack corresponding conjunction words or discourse particles denoting various discourse meanings, the *what* form frequently replaces conjunction words or discourse particles in those contexts. Thus, the exact discourse meaning of the *what* form is discernible only in specific conversational contexts.

6.3 Shared properties among all three languages

One common property among the derived uses in all three languages is that most of the non-interrogative uses fall in the negative territory, encoding such marked stances as uncertainty, lack of commitment, disapproval, and incredulity. This negative tendency is also mentioned in some of the earlier studies such as K-H Kim (1993, 1999), Nam and Cha (2010), and Cha (2010). It is most clearly demonstrated by the following types of extended uses found in our corpora: disapproval, avoidance, softener, and, to a lesser degree, filler, general extender, indefinite, exclamation, etc. All of them show in one way or another and in different degrees a mitigated nature of the reference or events being described. This tendency, we contend, can be characterized as an increase in the degree of subjectification (Traugott 1989, 2010).

The process of increased subjectification can be understood in this way. Interrogation by itself can be either objective (in the case of requesting information) or subjective (in the case of casting one's doubt). However, language use over time induces a skewed pattern: the categories of subjective use we observe in the cross-linguistic data overwhelmingly favor the latter, with a large portion of the speaker's personal judgment and involvement being encoded in the expressions as interrogative forms are repeatedly used in everyday discourse.

Although we are in no position here to conduct a comprehensive study on the mechanism of change in increased subjectification, we can nevertheless speculate that this is rooted in some of the general tendencies in the use of interrogatives in social interaction. Specifically, interrogation is generally a way of indicating lack of information and/or commitment, or the appearance of them. This negative propensity provides the basis for the utterance to be perceived negatively. Given the rather negative pragmatic association, then, it would be natural to expect that

interrogative forms such as interrogative pronouns and the larger grammatical expressions in which they appear develop in a negative, as opposed to positive, direction, hence the vague and negative qualities associated with most of the extended uses described here. This is very much in line with Haspelmath's (1997) speculation about the connection between bare interrogatives and their indefinite use. While Haspelmath acknowledges that this connection is a puzzle that defies conventional explanation, he supports the idea that interrogatives and indefinites share the feature of representing "information gaps." We want to take a step further by suggesting that there is a social interactional aspect involved as well in that interrogatives can indicate a lack of commitment or negative epistemic stance on the part of the speaker, which gives the impetus for further negative extension of the tokens and utterances involved.

This kind of analysis is further supported, at least partially, by conversation sequence-based studies such as Drew (1997). Drew (1997) shows that, *what?* in English, along with some other forms (such as *sorry*? *pardon*?) can be used as Next Turn Repair Initiators (NTRI). Significantly, what he shows is that, among other environments, NTRIs can appear in contexts where (1) the repairable turn does not appear to connect referentially with its prior turn, and (2) the repairable turn, although topically connected, is somewhat inappropriate as a response to the prior turn. Drew also notes that sometimes these forms may be used simply because the current speaker has not heard what the prior speaker said. In any case, all of these circumstances point to the negative characteristic of interrogatives use in conversational interaction, which, we are proposing, constitutes the interactive basis for the pragmatic development of the cross-linguistic properties under discussion.

To summarize, then, we have shown that the three languages we have examined exhibit parallel as well as divergent patterns of development. The diversity is mainly associated with typological features such as whether or not the interrogative forms have been assigned a prominent grammatical function or the extent to which interrogative and other lexical forms share the burden of serving related discourse functions. The convergence is most preponderant in the notion of negative pragmatic connotations associated with nearly all of the derived uses of interrogative forms. Such connotations typically show properties of distancing, uncertainty, disapproval, or disbelief.

7.   Conclusions

In this study, we used natural conversation data from three languages to investigate the extended uses of the interrogative forms *what* and related forms. Based on corpus data, we developed a tentative taxonomy and coding system for the analysis of

non-interrogative *what* expressions cross-linguistically. Our investigation shows interesting parallel phenomena as well as variations across the languages.

By way of conclusion, we would like to briefly discuss some remaining issues. The first is the implication of this study for language typology. We believe that, while standard approaches based on isolated sentence and lexical-grammatical forms have yielded valuable insights, it is critical to employ corpus data comprising natural spoken discourse to supplement standard typological studies. We have hopefully shown that natural discourse-based typological studies can be a productive endeavor. This is demonstrated in the development of taxonomies and coding systems, which have been sorely lacking in existing research. Without comparable large corpora, it would be difficult to develop a realistic coding system and make it available for crosslinguistic applications. However, we readily acknowledge the limitations of this approach, including (1) the reliance on forms as a starting point, (2) the lack of CA-style scrutiny of each of the tokens in context, and (3) the extremely small number of languages we have analyzed. There are precisely the areas where future research projects are called for.

Finally, we have looked at one interrogative form, namely the "what" type of expressions. There is evidence that many other interrogative forms (*where, who, how*, etc.) have developed extended uses along similar lines. The tendencies and explanations we have proposed here again need further validation as far as these other forms are concerned.

Acknowledgements

References

Biq, Yung-O. 1990. Question words as hedges in conversational Chinese: A Q and R exercise. In Lawrence B. Bouton and Yamuna Kachru eds., *Pragmatics and Language Learning, Monograph Series* vol. 1, 141–157. Urbana-Champaign: University of Illinois.

Blake, Norman F. 1992. *Why* and *what* in Shakespeare. In Toshiyuki Takamiya and Richard Beadle, eds., *Chaucer to Shakespeare: Essays in Honour of Shinsuke Ando*, 179–93. Cambridge: D. S. Brewer.

Brinton, Laurel. J. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin and New York: Mouton de Gruyter.   doi: 10.1515/9783110907582

Brinton, Laurel. J. 2008. *The Comment Clause in English: Syntactic Origins and Pragmatic Development*. Cambridge: Cambridge University Press.   doi: 10.1017/CBO9780511551789

Canavan, Alexandra and George Zipperlen. 1996a. *CALLFRIEND Mandarin Chinese-Mainland Dialect*. Linguistic Data Consortium, Philadelphia.

Canavan, Alexandra and George Zipperlen. 1996b. *CALLFRIEND Korean*. Linguistic Data Consortium, Philadelphia.

Cha, Jihyeon. 2010. Interaction of prosody and discourse function: Based on the Korean discourse marker *mwe* in the sentence-final position. *Language Facts and Perspectives* 25, 227–256.

Channell, Joanna. 1994. *Vague Language*. Oxford: Oxford University Press.

Chao, Yuan Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Crystal, D. and D. Davy. 1975. *Advanced Conversational English*. London: Longman.

Ding, Shengshu, Shuxiang Lyu, Rong Li, Dexuan Sun, Xiechu Guan, Qian Fu, Shengzhang Huang, and Zhiwen Chen. 1961. *Xiandai Hanyu Yufa Jianghua [A referential grammar of modern Chinese]*, Beijing: The Commercial Press.

Dingemanse, M., Blythe, J., and Dirksmeyer, T. 2014. Formats for other-initiation of repair across languages: An exercise in pragmatic typology. *Studies in Language* 38(1), 5–43. doi: 10.1075/sl.38.1.01din

Dingemanse, M. and N. J. Enfield. 2015. Other-initiated repair across languages: towards a typology of conversational structures. *Open Linguistics* 1, 96–118. doi: 10.2478/opli-2014-0007

Drew, Paul. 1997. 'Open' class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28(1), 69–101.   doi: 10.1016/S0378-2166(97)89759-7

Enfield, N. J. 2003. The definition of what-d'you-call-it: semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1), 101–117.   doi: 10.1016/S0378-2166(02)00066-8

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., Drew, P., Floyd, S., Gipper, S., Gisladottir, R. S., Hoymann, G., Kendrick, K. H., Levinson, S. C., Magyari, L., Manrique, E., Rossi, G., San Roque, L., & Torreira, F. 2013. Huh? What? – A first survey in 21 languages. In M. Hayashi, G. Raymond, & J. Sidnell, eds., *Conversational Repair and Human Understanding*, 343–380. New York: Cambridge University Press.

Godfrey, J., E. Holliman, and J. McDaniel. 1992. SWITCHBOARD: A telephone speech corpus for research and development. *Proc. ICASSP*, 517–520.

Goodwin, Charles. 1979. The interactive construction of a sentence in natural conversation. In G. Psathas, ed., *Everyday language: Studies in ethnomethodology*, 97–121. New York, Irvington Publishers.

Haspelmath, Martin. 1997. *Indefinite Pronouns*. Oxford: Clarendon Press.

Hayashi, Makoto, Geoffrey Raymond, and Jack Sidnell. 2013. Conversational repair and human understanding: An introduction. In Hayashi, Makoto, Geoffrey Raymond, and Jack Sidnell, eds., *Conversational Repair and Human Understanding*, 1–40. Cambridge: Cambridge University Press.

Heine, Bernd and Kuteva, Tania. 2006. *The Changing Languages of Europe*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199297337.001.0001

Heritage, John. 2012. Epistemics in action: action formation and territories of knowledge. *Research on language and social interaction* 45(1), 1–29. doi: 10.1080/08351813.2012.646684

Hsieh, Fuhui. 2005. The use of *shenme* in interaction, *Tatung Journal of General Education*, 1, 207–222.

Jefferson, Gail. 1990. List-construction as a task and a resource. In George Psathas, ed. *Interaction competence*, 63–92. Washington, D.C.: University Press of America.

Kim, Chung-Hyo. 2000. *Kwuke-ui uimwunsa-wa pwucengsa yenkwu* [A study on interrogative and indefinite words in Korean]. Seoul, Korea: Pakiceng.

Kim, Kyu-hyun. 1993. Other-initiated repair sequences in Korean conversation as interactional resources. *Japanese/Korean Linguistics*, 3, 3–18. Stanford: CSLI Publications.

Kim, Kyu-hyun. 1999. Other-initiated repair sequences in Korean conversation: Types and functions. *Discourse and Cognition*, 6, 141–168.

Kim, Myung-Hee. 2005. Kwuke uimwunsa-ui tamhwa pyoci-hwa [Pragmaticalization of interrogative words in Korean]. *Tamhwa-wa Inci* 12(2), 41–62.

Kim, Myung-Hee. 2006. Kwuke uymwunsa 'mwusun'uy tamhwa pyoci kinung [Discourse functions of *wh*-word *mwusun*]. *Tamhwa-wa Inci* 13(2), 21–42.

Koo, Hyun Jung and Rhee, Seongha. 2013. On an emerging paradigm of sentence-final particles of discontent: A grammaticalization perspective. *Language Sciences* 37, 70–89. doi: 10.1016/j.langsci.2012.07.002

Lee, Jeong-Ae. 2002. *Kwuke hwayongpyoci-ui yenkwu* [Research on pragmatic marker in Korean]. Seoul, Korea: Welin.

Lerner, G. H. 1994. Responsive list construction: A conversational resource for accomplishing multifaceted social action. *Journal of Language and Social Psychology* 13(1), 20–33. doi: 10.1177/0261927X94131002

Li, Audrey Y-H. 1992. Indefinite *what* in Mandarin Chinese, *Journal of East Asian Linguistics* 1(2), 125–55. doi: 10.1007/BF00130234

Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Lu, Jianming. 1986. Zhoubianxing zhuyuju ji qita [Sentences with inclusive subjects], *Zhongguo Yuwen* [Studies of The Chinese Language] 192(3): 161–167.

Lü, Shuxiang. 1982. *Zhongguo Wenfa Yaolue* [Essentials of Chinese grammar], Beijing: The Commercial Press.

Lü, Shuxiang. 1985. *Jindai Hanyu Zhidaici* [Modern Chinese pronouns], Shanghai, Xuelin Chubanshe [Academia Press].

Nam, Kilim and Cha, Jihyeon. 2010. Usage patterns and functions of discourse marker *mwo*. *Hangul* 288, 91–119.

Overstreet, M. 1999. *Whales, Candlelight and Stuff like that: General Extenders in English Discourse*. Oxford: Oxford University Press.

Overstreet, M. 2005. and stuff und so: Investigating pragmatic expressions in English and German. *Journal of Pragmatics* 37, 1845–64. doi: 10.1016/j.pragma.2005.02.015

Overstreet, M. and G. Yule. 1997. On being inexplicit and stuff in contemporary American English. *Journal of English Linguistics* 25, 250–258. doi: 10.1177/007542429702500307

Quirk, Randolf, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Schegloff, Emanuel A. 1997. Practices and actions: Boundary cases of other-initiated repair. *Discourse Processes* 23.3:499–545. doi: 10.1080/01638539709545001

Schiffrin, Deborah. 1986. *Discourse markers* [Studies in interactional sociolinguistics, 5]. Cambridge: Cambridge University Press.

Selting, Margret. 1996. Prosody as an activity-type distinctive cue in conversation: the case of so-called "astonished" questions in repair initiation. In Elizabeth Couper-Kuhlen and Margret Selting, eds., *Prosody in conversation : interactional studies*, 231–270. New York: Cambridge University Press. doi: 10.1017/CBO9780511597862.008

Shao, Jingmin and Xiufeng Zhao. 1989. Shenme fei yiwen yongfa yanjiu [A student on the non-interrogative usage of shenme], *Yuyan Jiaoxue yu Yanjiu* [Language Teaching and Linguistic Studies] 1: 26–40.

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, John. 1996. The search for units of meaning. *Textus: English Studies in Italy* 9, 75–106.

Song, Zuoyan and Hongyin Tao. 2009. A unified account of causal clause sequences in Mandarin Chinese and its implications. *Studies in Language* 33(1), 69–102. doi: 10.1075/sl.33.1.04son

Tang, Tingchi. 1981. Guoyu yiwenju de yanjiu [A study on Mandarin Chinese interrogative sentences], *Shida Xuebao* [The Journal of National Taiwan Normal University] 26:219–277.

Tao, Hongyin. 1996. *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*. Amsterdam: John Benjamins. doi: 10.1075/sidag.5

Tao, Hongyin. 2011. *Working with Spoken Chinese*. Center for Advanced Language Proficiency Education and Research (CALPER) Publications, Pennsylvania State University. State College, PA.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins. doi: 10.1075/scl.6

Traugott, Elizabeth Closs. 1982. From propositional to textual and expressive meanings: some semantic-pragmatic aspects of grammaticalization. In Winfred Lehmann and Yakov Malkiel, eds., *Perspectives on historical linguistics*, 245–71. Amsterdam: John Benjamins. doi: 10.1075/cilt.24.09clo

Traugott, Elizabeth Closs. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 57, 33–65.

Traugott, Elizabeth Closs. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. In Kristin Davidse, Lieven Vandelanotte, and Hubert Cuyckens, eds., *Subjectification, inter-subjectification and grammaticalization*, 29–71. Berlin and New York: De Gruyter Mouton. doi: 10.1515/9783110226102.1.29

Wang, Haifeng and Tieli Wang. 2003. Ziran kouyu zhong shenme de huayu fenxi [Discourse analysis to *shenme* in spoken conversation], *Hanyu Xuexi* [Chinese Language Learning] 2: 21–29.

Zhou, Rong. 2007. Yiwen daici shenme ge yongfa zhijian de guanxi xintan [Relations among different uses of interrogative pronoun *shenme*], *Yuwen Xuekan* [Journal of Language and Literature Studies] 14, 120–123.

Zhu, Dexi. 1982. *Yufa Jiangyi* [Lectures on grammar], Beijing: The Commercial Press.

Appendix A. Transcription conventions

Transcription representations generally follow the original transcription in the LDC corpora. Transcription is fairly broad, with each line representing a loosely defined prosodic unit. For more information about transcription and data sources, the reader is referred to Canavan and Zipperlen (1996 a&b) for the Chinese and Korean data and Godfrey et al. (1992) for the English data.

Appendix B. List of gloss abbreviations used in the Chinese examples

1SG	1st person singular	EM	emphasis marker
2SG	2nd person singular	NEG	negator
2PL	2nd person plural	NOM	nominalizer/nominalization
3SG	3rd person singular	PRT	particle
3PL	3rd person plural	PASS	passive
ASSO	associative marker	PFV	perfective
CLF	classifier	PST	past tense
COMP	complementizer		

Appendix C. List of gloss abbreviations used in the Korean examples

AC	accusative particle	SBJ	subject particle
COMM	committal	POL	polite speech level
CORREL	*-ketun* correlative	POSS	possessive particle
DCT.RE	deductive reasoning	PST	past/perfect aspect suffix
DC	declarative suffix	Q	interrogative
DN	defective noun	QT	quotative particle
HON	honorific suffix	RL	relativizer
INF	infinitive suffix	RT	retrospective mood suffix
INT	intimate speech level	TP	topic particle
LOC	locative	UNASSIM	unassimilated

Corresponding author's addres

Heeju Lee
University of California, Los Angeles
Department of Asian Languages and Cultures
290 Royce Hall
Los Angeles, CA 90095-1540
USA

heeju0829@gmail.com