

## The apprehensive: Fear as endophoric evidence and its pragmatics in English, Mandarin, and Russian

### ABSTRACT

*This study attempts to explicate the cross-linguistically widespread uses of fear expressions that can be understood as serving pragmatic purposes, here dubbed the “apprehensive”, as in **I’m afraid I must ask you to leave now**. We argue that such uses instantiate “endophoric evidential” (Plungian, 2001) in the sense that the emotion of fear is recruited as direct personal source of information in the construction of epistemological stance. Quantitative collocation data from English, Mandarin, and Russian suggest that the apprehensive strongly attracts a proposition of high-certainty, such as denied ability, asserted predictability, and recognized obligation or necessity. The results call for a pragmatic analysis of the apprehensive as a discourse device of expressing epistemological stance. It prepares the listener of the undesirable news for which the speaker has reliable evidence. The psychology of fear as a basic emotion is discussed as the experiential basis of the pragmatics of the apprehensive. The study suggests that pragmatics is shaped by human cognitive and affective processes and is therefore embodied.*

*Keywords: pragmatics, embodiment, emotion, evidentiality, epistemological stance*

### 1. Why are they afraid?

The study presented here began eight years ago in the middle of a goodnight story. In the picture book by Ruth Krause, titled *The Carrot Seed*<sup>1</sup>, which the first author was reading to her two-year old daughter, a little boy planted a carrot seed. He diligently attended to it every day, eagerly anticipating the appearance of the carrot plant. But his family was anxious that the seed would not grow, which gave rise to their rather discouraging comments:

*His mother said: “I’m afraid it won’t come up.”*

*His father said: “I’m afraid it won’t come up.”*

Just as the first author finished the mother’s comment, a look of alarm appeared in her daughter’s face. By the time she spoke the father’s utterance, her daughter asked a question: “Why are they afraid?” All of a sudden, the first author was taken aback, as much by the simplicity of the question as by its complexity, both at the level of language use and at the level of human psychology. An attempt has been made ever since to find an answer to the child’s question, and this study offers one from a perspective that takes into account the semantics, pragmatics, and the psychology of fear.

---

<sup>1</sup> *The Carrot Seed*. Story by Ruth Krause, Pictures by Crockett Johnson. Harper Festival, 1993.

## 2. The apprehensive

Despite controversies in psychology as to which emotions are basic, the recognition of fear as a basic emotion is unequivocal. Part of being alive is being afraid. Naturally, being an emotional universal, fear is a semantic universal and is encoded in all languages (Wierzbicka, 1999), although speakers of different languages may experience fear with differing intensity (Lewandowska-Tomaszczyk, 2011). Interestingly, in many languages of the world an expression of fear (hence EOF) may have a non-literal interpretation that is continuous with its literal sense, as illustrated by the following examples from Finnish, German, Mandarin, Japanese, Spanish, English, and Russian, respectively (with EOF underlined):<sup>2</sup>

- (1) *Pelkään etten tiedä mitä rakkaus on.*  
I'm afraid I do not know what love is.
- (2) *Ich fürchte, ich muss Sie enttäuschen.* [deWAC, 12059804]  
I'm afraid I have to disappoint you.
- (3) *Zhe kongpa shi shijie shang zui chou de mao le.* [club.astro.sina.com.cn]  
I'm afraid this is the ugliest cat in the whole world.
- (4) *Yabun osoku osore-irimasu.*  
I'm sorry (afraid) to be calling so late.
- (5) *Me temo que no estoy de acuerdo contigo.*  
I'm afraid I don't agree with you.
- (6) *I'm afraid if you look at a thing long enough, it loses all of its meaning.* [Andy Warhol]
- (7) *but I fear it will all be in vain.* [BNC, 39192]
- (8) *Боюсь, у вас будут крупные неприятности.* [RWC, 139188342]  
I'm afraid you will have major problems.
- (9) *Боюсь, я не могу согласиться.* [RWC, 9790356]  
I'm afraid I can't agree.
- (10) *Боюсь, вам не понравится то, что я скажу.* [RWC, 134262733]  
I'm afraid you won't like what I will say.

The widespread non-literal usage of the EOF across languages requires an explanation. Specifically, we have to answer the following questions: (1) how to characterize the non-literal EOF in terms of its place in the universal semantic space, (2) how to characterize its pragmatic function, and (3) what are the psychological mechanisms behind its pragmatic function? These questions trigger a series of more fundamental questions relevant to our experience of emotion and our experience with language use in interaction and, more importantly, how the two experiences relate to each other.

Several previous studies touch upon the usage of the EOF that may or may not be literal. Palmer (2001:133-134) suggests the term “timitive” for fear expressions and argues that they express “little more than an unwelcome possibility” or epistemic uncertainty. Givón (2001:52-53) argues that verbs of fear, as with verbs of wish and hope, are used as “preference verbs” expressing (dis)preference for/against possible future events and that such a use is deontic. When used with a past event, Givón says, they express “epistemic anxiety” about the unresolved

<sup>2</sup> We owe the Finnish example to Thomas E. Payne, and the Spanish to Roger Gilabert Guerrero.

certainty of that event. Contrary to Palmer and Givón, who see uncertainty behind the use of the EOF, Leech (1983) says that by using *I'm afraid* the speaker warns the addressee of the upcoming inevitable bad news. Similarly, popular writer George Mikes (Mikes, 1984) sees the expression as a hedge to soften the impact of the unwelcome truth to be told. If the intuition of Leech and Mikes were right, one would expect certainty as the epistemic status of the event introduced by an EOF rather than the lack of certainty. The reason is obvious: it is unlikely that one is uncertain about something that one knows will inevitably happen and for which it is necessary to soften the blow or prepare one's addressee. Owen (1983) considers *I'm afraid* an apologetic expression as part of a remedial exchange. Although Owen does not explicitly argue for the speaker's certainty about his/her proposition, it is reasonable to infer that one will only utter a proposition that is likely to require an apology if one is certain about the proposition. To put it simply: why risk offending your listener if you are not even certain about what you say? The terse literature on the topic suggests two views on EOF uses. On the one hand, Palmer and Givón are primarily concerned with the level of knowledge representation in terms of the speaker's state of knowledge and source of information when s/he uses an EOF. Leech, Mikes, and Owen, on the other hand, focus on the interpersonal function of the EOF. These two levels are distinct, but interrelated.

In this paper, we separate the level of knowledge from that of social interaction. While the former pertains to principles of epistemology, the latter pertains to principles of social pragmatics and social psychology. We will attempt a comprehensive account by considering the interaction between the two levels in terms of the construal of "epistemological stance", to use a term proposed by Mushin (2001). We will show on account of corpus data that at the level of knowledge representation the EOF is associated with the reliability of information and the speaker's certainty and confidence about the information to be presented. However, the function of the EOF operates at the interpersonal level, where high confidence and certainty on the speaker's part can be socially threatening to the listener, which calls for the construal of a relativizing and distancing epistemological stance. We will argue that the EOF serves exactly the pragmatic function of construing such a stance and thereby prepares the listener for the forthcoming proposition of certainty, thus minimizing potential social risks. In this sense, it is *social anxiety* rather than "epistemic anxiety" that motivates the usage of the EOF. The goal of the EOF, in our view, is to induce the listener to take the message of the embedded clause seriously while not taking offence at the messenger. The way the EOF accomplishes this goal is by combining a high-certainty embedded clause with an expression of commiseration. The EOF does not "soften the impact of the awful truth" by pretending to be uncertain of its truthfulness but rather indicates that the speaker knows that it is awful and understands the listener's feelings. This is why we hypothesize that the embedded clause is more likely to contain expressions of certainty than those of uncertainty, which would put the reliability of the message in doubt.

The conventionalization of the EOF uses is especially clear through the observation that in most of the above examples, the EOF does not necessarily convey the real emotion of fear with all its psychological, physiological and behavioral characteristics and consequences. In other words, the fear being thus expressed is not necessarily referential and may not be taken at face value. We shall call this usage the "apprehensive" and define it in terms of EOF produced by the speaker in anticipation of their own undesirable assertion, which may be interpreted prototypically as a socio-pragmatically driven expression of non-referential fear.

It is the goal of this paper to explicate the semantics, pragmatics and experiential basis of the apprehensive in three languages: *I'm afraid*, *I am afraid*, and *I fear* in English, *kongpa* in

Mandarin Chinese, and *bojus'* in Russian.<sup>3</sup> Specifically, we consider the apprehensive as representing what Plungian (2001) calls “endophoric” evidential, that kind of evidential in which a direct personal inner state constitutes the source of information. We argue that endophoric evidentiality is a linguistic device of constructing epistemological stance. Mushin’s (2001:52-58) notion of “epistemological stance” captures the idea that evidentials are employed by speakers to construe a pragmatically meaningful stance in a given discourse context. She contends that epistemological stance “mediates between speakers’ actual acquisition of information and how they talk about it.” Indeed, the literature on evidentiality provides ample examples of evidential markers being used for pragmatic and interpersonal purposes (e.g. Babel, 2009; Biber and Finegan, 1989; Clift, 2006; Curnow, 2003; Du Bois, 2007; Hill and Irvine, 1993, to name just a few). Cross-linguistically attested uses of the apprehensive can be explained in terms of the pragmatic exploitation of endophoric evidential for the construal of epistemological stance. Given that epistemological stance is pragmatic in nature, its linguistic representation may or may not reflect the objective source of information. As Mushin points out, when there is mismatch between the two, the construed evidentiality serves as a “rhetoric tool”. Thus, in our case, the observation that the apprehensive does not necessarily express fear in the literal sense attests to the subjective nature of epistemological construal as dictated by the pressure of social interaction. But how is it that construed fear can do what it does in communication – from warning to a kind of pre-emptive self-defense – and can become conventionalized in that function? The preparatory and protective function of the apprehensive cannot be properly understood without reference to the nature and features of fear as a basic emotion, which is considered in this study to be the experiential basis of the pragmatization of EOF. A larger ramification of this study is that pragmatics is not separated from the rest of human experience, but experientially motivated and therefore embodied.

In addition to providing an explanation of the apprehensive usage by means of case studies in three languages, this study will contribute to research on evidentiality at two levels. First, it will fill a gap in the study of endophoric evidentials, a cross-linguistically salient but understudied category. Available studies on English evidentiality address evidential markers in conversation and academic discourse (Chafe, 1986), the emergence of evidential meaning from modal verbs (Traugott, 1989), speech act verbs such as *promise* and *threaten* (Traugott, 1997), cognitive verbs (Brinton, 1996), and auditory perception verbs (Whitt, 2009).<sup>4</sup> Although fear expressions in the first person are included in Biber and Finegan (1989) as affect markers for the conveyance of stance in English, there is no detailed examination of this category in terms of the construal of information source and the speaker’s intent behind it. Previous studies on Chinese (Chappell, 2001; Hsieh, 2008; Li and Liu, 2003; Zhang and Yu, 2003; Zhu, 2006) and Russian (Grigorenko, 2009; Kaksin, 2010; Kozinceva, 1994; Sergeeva, 2010) focus on sensory, inferential, and reported evidential, but affect has not been recognized as a source of information. The notion of the apprehensive as endophoric evidential provides a diagnostic and descriptive model for a widespread phenomenon across languages as well as an explanatory paradigm for understanding the phenomenon at the semantic, pragmatic, and experiential level.

---

<sup>3</sup> Given our uncertainty about the interchangeability of *I’m afraid* and *I am afraid*, these were treated as distinct apprehensives.

<sup>4</sup> It may be argued that the verb *threaten* is an indirect endophoric evidential because it describes threat, i.e. that which causes fear, rather than fear per se, as the source of information.

Second, we separate semantics – here understood narrowly as truth-conditional meaning – and pragmatics – understood as language use in social interaction and its effects on others – as two distinct yet closely related levels of analysis<sup>5</sup>. In so doing we are able to reconcile seemingly contradictory views in the literature on the role of evidentials in modifying the sincerity condition of a proposition. On one side, it has been asserted that the speaker’s first-person evidence adds a “sincerity condition” to the assertion being made. Faller (2002), for instance, argues for this position with regard to the direct evidential *-mi* in Cuzco Quechua.<sup>6</sup> On the other side, evidentiality – regardless of directness – has been linked to the weakening of the sincerity condition of a proposition. Chung (2010), for example, associates the use of evidentials in Korean with weakened reliability and commitment, contending that speakers use an evidential, direct or indirect, if they intend to “underspecify” their belief in the truth value of the proposition (p. 947). While this may be true of some evidentials, we suggest that this is not the case for the apprehensive, which serves as a cue to an upcoming reliable assertion that the speaker feels to be something the listener does not want to hear. Thus, instead of a simple concern with the linear amplification or reduction of the sincerity condition, our analysis suggests a more complex view of the use of evidentials as a pragmatic solution to a recurrent social situation in which the interlocutors’ face, image, and relationship are at stake.

In Section 3, we explore the semantics of the apprehensive by identifying fear as endophoric evidence within the semantic field of evidentiality. Following this, in Section 4, the collocational behavior of the apprehensive is examined in terms of the degree of certainty associated with the proposition introduced by the apprehensive as an evidential. Specifically, we analyze the strengths by which different modal expressions and other lexical grammatical categories related to modality collocate with the apprehensive in British English (4.2), Mandarin Chinese (4.3) and Russian (4.4). In Section 5, after a brief discussion of the findings in the three languages, we offer an analysis of the pragmatic functions of the apprehensive, noting both crosslinguistic commonalities and culture-specific language use conventions. In Section 6, we discuss the psychological characteristics of fear as the experiential basis of the apprehensive to account for the pragmatic function of the apprehensive. Section 7 concludes the paper and discusses its implications for the typology of evidentiality and linguistic pragmatics.

### 3. Fear as direct source of information

In linguistics, definitions of evidentiality largely fall into two senses. In the narrow sense of the term, evidentiality refers to the source of information behind an assertion, that is, “how one has knowledge of what one is saying”, to use Hardman’s (1986:115) words. In its broad sense, the notion of evidentiality refers both to source of information and to the speaker’s evaluation of the reliability or likelihood of what s/he is saying, which is the domain of epistemic modality (Chafe, 1986). In this paper we will follow Plungian (2001) and adopt the narrow sense of evidentiality, treating it as a semantic category distinct from, though related to, epistemic modality. In his classification of evidential values, Plungian first makes a general distinction between direct and indirect evidence. Given the prominence of sight in human experience, direct evidence is further

---

<sup>5</sup> Opinions on how to define semantics and pragmatics are diverse and the two areas of interest are often overlapping (Crystal 1997:120). While emphasizing the different foci of semantics and pragmatics, the definitions adopted here are no doubt oversimplifications of two much richer and more complicated linguistic domains.

<sup>6</sup> Similar findings can be found in Floyd (1993) and Weber (1986).

divided into visual and non-visual evidence, the latter including sensoric and endophoric evidence. While sensoric evidence subsumes senses such as auditory and olfactory perceptions, an endophoric value refers to that which one feels as an inner state, either mental or physical. What we are concerned with in this paper is endophoric evidence. Specifically we deal with the apprehensive, which involves the emotion of fear as a direct source of information. Figure 1 is reproduced from Plungian (2001:353):

**Figure 1. The classification of evidential values**

<<INSERT FIGURE 1 HERE>>

Aikhenvald (2004) points out that epistemic extension from evidentiality is common in languages that employ lexical means of encoding source of information. There is however no clear consensus on the relationship between evidentiality and epistemic certainty (Dendale and Tasmowski, 2001; Connillie, 2009). Plungian states that the directness of evidence corresponds to the reliability of the proposition being asserted. Thus, the more direct the evidence, the more reliable is the information, and the more certainty is signaled by the speaker. However, because Plungian's categorization does not imply a gradation of directness, the reliability of information based on endophoric evidence relative to information of other direct sources remains an empirical question.

On Givón's (1982:43-4) scalar approach to evidential strength, feeling (endophoric evidence) is lower in strength than other senses along the sensory gradient. But it is high on the scales of person, directness, and proximity due to its first-hand immediacy. Furthermore, Givón sees evidentiality as inherently tied to midrange epistemic certainty because for him neither high certainty nor low certainty requires evidence. The former is beyond challenge while the latter is beneath challenge. Evidence is only required when certainty is in the middle and therefore open to challenge. It is true that assertions about absolute truths such as *The earth is round* and those of absolute nonsense such as *All birds are mammals* would require no mediation or evidence because their respective validity and falsity is beyond question. However, real-life communications are usually not about total truisms and extreme absurdities like these. The majority of meaningful human communications involve precisely fine gradations of reliability and certainty. Furthermore, as Mushin (2001) argues, what really matters in communication is not so much the actual source of information as the speaker's construal of source of information for social purposes. Thus, it will be interesting to see whether and to what extent the speaker is certain about what s/he is saying when s/he uses an apprehensive, and what the speaker really wants to do by using it. This is what we will attempt to determine with a corpus-driven quantitative method in the next section.

#### **4. The apprehensive and reliability of assertions**

In this section, we present data from British English, Mandarin Chinese, and Russian to examine the relationship of the apprehensive as an evidential to the reliability of the assertion introduced by the apprehensive. More specifically, we will look for lexical and/or grammatical tendencies in the collocation patterns of the apprehensive as cues into the degree of reliability of the information introduced by the apprehensive.

#### 4.1 Corpora and methods

British National Corpus (BNC, XML edition, 2007), Mandarin Chinese Internet-ZH Corpus (I-ZH), and the Russian Web Corpus (RWC) were adopted for the purpose of this study. All three are balanced corpora consisting of written and spoken discourses in a variety of genres. BNC is a part-of-speech tagged 96-million-word corpus consisting of written and spoken language materials representative of modern British English.<sup>7</sup> I-ZH is a tokenized and part-of-speech tagged Chinese Web corpus of 198 million words covering a wide range of text types in simplified script. RWC is a part-of-speech tagged and lemmatized Web corpus designed to provide a general language resource for modern Russian. It contains 188 million tokens and 1.5 million words. Both I-ZH and RWC are collected by Serge Sharoff as BNC-like corpora in the sense that the language materials are not genre-biased but representative of general language use. However, while spoken language materials in the BNC are transcripts of actual speech, the core of “spoken” language in the two Web corpora I-ZH and RWC are texts from online forums, blogs, and chat rooms.

Turning now to the quantitative analysis of the reliability of an assertion, it should be noted that there are multiple ways of evaluating this semantic dimension. First, it can be evaluated by the degree of subjectivity across modal categories. Generally, the more subjective a modal category, the less reliable the information it describes. Among the three modal types, dynamic modality describing ability is considered relatively less subjective than deontic modality, which in turn is relatively less subjective than epistemic modality (Traugott, 1989; Peng, 2007). However, although the gradation of subjectivity may be useful in comparing the modal types as whole categories, it does not effectively measure the subjectivity level of individual modal auxiliaries, as other factors such as polarity must also be taken into consideration. A second measure of reliability is the relative strength or degree of a scalar (e.g. deontic and epistemic) modality as described by Huddleston and Pullum (2002), which may include a dimension of polarity as described by Nuyts (2006). On such a scale, modal types on the two extremes correspond to high reliability, which decreases toward the middle of the scale. For example, within epistemic modality, absolute necessity is a stronger modality and indicates higher reliability than probability. Likewise, absolute impossibility is a stronger modality and indicates higher reliability than improbability. To put this in social context, in situations where circumspection and moderation are valued, for example, one is more likely to incur social risks when stating that something is certain or impossible than when stating that something is (im)probable. Similar scalar measurements apply to other modal types. A third measure of reliability is the reality and actuality of an event as marked by tense and aspect markers. To be specific, realis events in the present or past are more reliable than irrealis events that are talked about as possible worlds.

The Word Sketch Engine (WSE) was used in the present study as the primary tool of generating concordances and ranking collocates of the apprehensive constructions by log-likelihood. These functions were performed on the online interface <http://beta.sketchengine.co.uk>. For *I'm afraid* and *I am afraid* in English, we ran a query for the phrases *I'm afraid* and *I am afraid*, respectively. The query for the former yielded 1967 concordances, and the query for the latter yielded 303 concordances. Following this, three negative filtering operations were conducted in the +1 slot (first word to the right of the node phrase) to eliminate instances of

---

<sup>7</sup> For further details on BNC, consult the Reference Guide for the British National Corpus (Burnard, 2007).

*afraid to*, *afraid of*, and *afraid for* from the concordances of both apprehensives. As a result, a total of 1907 (17 per million) tokens of *I'm afraid* and a total of 271 (2.4 per million) tokens of *I am afraid* remained and were included in the analysis. For *I fear*, a query for the key phrase *I fear* yielded a total of 436 tokens. Extensive manual annotations were used to exclude false positives such as *I fear for X*, *I fear nothing*, and *all I fear is*. A total of 327 (2.9 per million) tokens were maintained after the filtering. We manually annotated each instance of each of the three apprehensives in a 250-token random sample to determine whether it involved real fear (using a three-value scheme of “yes”, “no”, and “maybe”). Both authors provided independent judgments, which agreed reasonably well with each other (1.2% with real fear for *I'm afraid*, 4.2% for *I am afraid*, and 41.2% for *I fear*).<sup>8</sup>

For Mandarin Chinese, we ran a keyword query for 恐怕 (*kongpa*). This apprehensive has over time undergone categorical change from a full verb to an adverb in modern Mandarin and is rarely used as a full verb, although there may be categorical ambiguity in certain cases. A total of 13317 (47.9 per million) concordances of *kongpa* was retrieved and included in the study. Manual annotations of a 250-token random sample showed only two instances involving real fear and 99.2% with no real fear.

For Russian, we searched for the apprehensive form (боюсь ‘I fear’) followed by a comma, which indicates a sentence internal clause or embedded element (interjection, discourse marker, extraposed adjective, etc.) boundary. The inclusion of the comma eliminates uses of *боюсь* with a non-clausal direct object. Results were then manually inspected to eliminate cases in which *боюсь* is used simply to mean ‘I am afraid’ with no specification of what is feared in the same sentence. A total of 3898 (20.7/million) tokens of the apprehensive were identified.

We employed collostructional analysis for each language to measure the collocation strengths of the modal collexemes of the respective apprehensive constructions. This method is also called collexeme analysis (Stefanowitsch and Gries, 2003). Set in the theoretical framework of Construction Grammar, it aims to investigate the interaction of a grammatical construction with lexical items entering a particular slot of the construction – known as the collexemes of the construction. Specifically, it explores the collocational patterns and preferences of a grammatical construction by ranking its collexemes in the order of the strength by which they are attracted to or repelled by the construction. Attraction is measured using Fisher’s exact test: collexemes attracted to a construction are significantly more frequent in the construction than elsewhere. Collexemes repelled by a construction are significantly less frequent in the construction than elsewhere. The ranked list of collexemes allows fruitful qualitative interpretation of the semantic properties of the construction. We also employed a multiple distinctive collexeme analysis on *I'm afraid* and *I am afraid* to determine their differential preferences of contracted collexemes. This method allows an investigation of the nuanced differences between synonymous constructions. We used Collo.analysis 3.2 (Gries, 2007) for the collostructional analysis and the multiple distinctive collexeme analysis.

In addition to the analysis of collexeme strengths for each language, we also conducted cross-language comparisons with regard to a number of other collocational properties, for which we used a 250-token sample for each apprehensive in each language.<sup>9</sup>

<sup>8</sup> The inter-rater agreement was 99.6%, for *I'm afraid* and *I am afraid*, and 91.2% for *I fear*.

<sup>9</sup> The first 250 manually proved tokens of a 500 token random sample automatically generated in WSE were used for *I'm afraid*; for *I am afraid* and *I fear*, the first 250 manually proved tokens of their respective filtered

## 4.2 Data on British English

In this section, we will determine the collocation strengths of the collexemes of the two British English apprehensive constructions by way of collostructional analysis. The patterns of attraction of the collexemes will help determine the extent to which the apprehensive constructions are associated with (expressed) information reliability and speaker certainty. In addition, similarities and differences between the two constructions can be identified on account of their collocation behaviors.

### 4.2.1 *I'm afraid* and *I am afraid*

Two steps were taken in identifying the modal collocates of *I'm afraid*. First, we used the “collocations” function in WSE to generate a list of collocates ranked by log-likelihood. The target of the search was word forms rather than lemmas to ensure that relevant semantic distinctions carried by inflections are not lost (e.g. *may* and *might* differ in epistemic strength). For *I'm afraid*, modal expressions are most likely located somewhere between the subject and the lexical verb of the embedded clause next to the node phrase. Therefore we set the search window on both sides of the node, between words -5 and 5. Then we selected from the resulting list of collocates all the modal auxiliaries with a minimum frequency of 3 as the basis of further refined search.<sup>10</sup> Based on this list, we conducted a positive filtering of the 1907 concordances of *I'm afraid* between the search window -5 and 5 for each modal on the list. Subsequently, we manually annotated the results of the positive filtering for each item. However, this procedure is still incomplete because it cannot retrieve items that are phrasal in form but are functionally modal-like such as *have to* and *be going to (gonna)*, which do not appear in the word-based list of collocates.

To solve the problem of incomplete retrieval, we ran additional positive filtering of the 1907 concordances of *I'm afraid* for instances of *have to*, *has to*, *going to*, and *gonna*, respectively, within the -5 to 5 window. The results of each filtering were manually examined for false positives. For *have to*, only finite forms were included in the data, e.g. instances of *will/would have to* were eliminated. Then the results on *have to* and *has to* were combined as instances of one modal. Similarly, the results on *going to* and *gonna* were combined as instances of one modal. The same procedure of collocate retrieval was applied to *I am afraid*.

Based on the refined results of retrieval we performed collostructional analyses with regard to the modal auxiliary (Mod) slot of the complex construction [*I'm afraid (that)* Sub Mod VP] and [*I am afraid (that)* Sub Mod VP], respectively. The respective results are shown in

---

concordances (of 271 and 327 tokens, respectively) were adopted. An automatically generated 250-token random sample was used for Mandarin *kongpa*. Because of the particular nature of the Russian apprehensive, two separate 250-token samples out of all tokens of the apprehensive (N=3898) were used: (1) the first 250 tokens that came from English speakers and (2) the first 250 that came from Russian speakers and were not literal expressions of fear.

<sup>10</sup> The automatic retrieval yielded *ca* and *wo* as part of *can't* and *won't*, respectively, as the contracted negative forms are treated as two separate words by the system. Because negated modals differ from their unnegated counterparts in terms of modal strength and speaker stance, we treated them as separate items in the subsequent collostructional analysis. It should be noted that this treatment applied to all modals under investigation in English and Chinese. For Russian, the negator is often not adjacent to the modal and cannot be argued to form a single lexeme with it. Thus instead of treating negated and non-negated modals as different lexemes, we examine the proportion of times sentences containing a given modal are negated.

Table 1 and Table 2.<sup>11</sup> In the column of collexemes, capitalized items represent both contracted and non-contracted forms of the same morpheme, e.g. NOT for *n't* and *not*, and *WILL* for both *'ll* and *will*.

**Table 1. Modal auxiliaries as collexemes of *I'm afraid***

<<INSERT TABLE 1 HERE>>

**Table 2. Modal auxiliaries as collexemes of *I am afraid***

<<INSERT TABLE 2 HERE>>

In order to interpret more precisely the communicative functions of the modal auxiliaries in the apprehensive, we also coded the dominant modality type in which the modal auxiliaries are used, and within the dominant modality type the dominant grammatical person of the referent about which the judgment is made.<sup>12</sup> The rightmost column in Table 1 provides each modal auxiliary's dominant (> 50%) modality type and dominant (>50%) grammatical person in the

---

<sup>11</sup> Collocation strengths above 2.64 are significant at the .05 level with Bonferroni correction.

<sup>12</sup> This method is preferable in dealing with English and Mandarin Chinese due to their morphological impoverishment, and was applied to both languages in this study. The reason for the preference lies in the technical requirements of the collostructional analysis. Specifically, the collostructional analysis requires the frequency of a collexeme in the entire corpus in addition to its frequency in the construction under investigation. This method avoids complications of manually coding modality type and grammatical person for each collexeme in the entire corpus. For Russian, modality type and grammatical person were coded in separate lemmas for the collostructional analysis. Proportion of modality types was calculated based on a 100-token random sample of each ambiguous modal and then multiplied by the number of tokens of the modal in the corpus to estimate how often, e.g., an epistemic *may* occurs in the corpus.

dominant modality type by percentage wherever applicable. In coding modal type, we followed Palmer's (1986) tripartite classification into dynamic (ability and volition), deontic (necessity, obligation and permission), and epistemic (probability and possibility) modals. These roughly correspond to the participant-internal, participant-external, and epistemic modals in terms of Van der Auwera and Plungian (1998). Following Huddleston and Pullum (2002:190) in viewing the connection between futurity and modality as intrinsic, we treat futurity markers *will/shall* as indicating confident prediction (relative to other modals).

When we look at the most strongly attracted modal auxiliaries in Table 1, we see clear lexical and grammatical tendencies. Specifically, the apprehensive prefers the dynamic modal in its negative form (*canNOT*) indicating absolute inability. The negative dynamic modal is more often than not used to refer to a first person, typically the speaker. By contrast, its positive counterpart *can* is repelled by the apprehensive. This contrast suggests that it is the denial, but not the assertion, of ability and possibility that is at the center of the discourse function of the apprehensive. This denial serves the purpose of evading expectation or responsibility. Similarly, the collocation strength of *canNOT* is significantly higher than that of its past counterpart *couldNOT* denoting attenuated inability or impossibility ( $\chi^2(1)=462.957$ ,  $p<.000001$ ).

On the other hand, the future markers associated with confident prediction are strongly preferred by the apprehensive. The high rankings of future markers attest to the high certainty of the assertions introduced by the apprehensive.<sup>13</sup> This is because futurity pertains to the predictability of an event whose factuality can be confirmed in the future. Huddleston and Pullum (2002: 189) maintain that future *will* has “the same semantic strength” as epistemic *must* which conveys high certainty of the speaker's belief or inference. Comrie (1985:44) pushes the point even further by arguing that futurity is “distinct from modal constructions that make reference to alternative worlds”. For him futurity even exceeds the strongest of epistemic modals in certainty: while an assertion in the future tense is “a very definite statement about a state of affairs to hold at a certain time subsequent to the present moment”, epistemic modality even in its strongest term still pertains to alternative worlds. Biber and Finegan (1989: 98) call both *will* and *shall* “predictive modals” and consider them markers of certainty. In comparison, the medium modality markers such as *may/might/can/could* rank lower in collocation strength (attraction rates even for the closest modals from the two groups, *might* and *shall*, are highly significantly different  $\chi^2(1)=7.6$ ,  $p=.006$ ). The deontic modals expressing strong obligation *must* and *have to* are also preferred, and tend to be used in the first person indicating that the speaker is internally or externally compelled to do something. As markers of strong modality, *must* and *have to* rank much higher than *should*, which represents medium modality. According to Biber and Finegan (1989:98), *should* indicates doubt. Thus the fact that *should* is repelled by the apprehensive construction can be attributed to the incompatibility between the apprehensive and the reduced certainty of *should* relative to both deontic and epistemic necessity. It is also interesting to note that the desire verbs *want* and *need*, both of which have a sense of dynamic modality, are repelled. We may account for this by suggesting that desirable things or events are semantically incompatible with the use of the apprehensive which anticipates the undesirability of the upcoming proposition. This incompatibility may also account for the repulsion of deontic *should*.

---

<sup>13</sup> This is not simply because we usually fear what is yet to happen (i.e., future events), and other modals are not always used with future events: *I'm/am afraid* are mostly used non-literally and are in fact used with irrealis events most of the time (70%).

The data in Table 2 show that *I am afraid* has a collocation pattern similar to its contracted counterpart in terms of the ranking of collexeme strengths. But because its overall frequency in the corpus is much lower than *I'm afraid*, it attracts the same collexemes with a lower strength. Nevertheless, the two strongest collexemes of *I am afraid* are *cannot* and *will*, expressing speaker inability and predictability of third-person events, respectively. Second, deontic *must* and *have to* are strongly attracted. Similar to *I'm afraid*, *I am afraid* repels modals of possibility *could* and *can*, as well as *should*, *need*, and *want*. What differentiates the two apprehensives seems to be stylistic. During data annotation we noticed that the contracted apprehensive *I'm afraid* goes together with contracted negations of modals (e.g. *can't* and *won't*), whereas the uncontracted *I am afraid* goes together with uncontracted negations of modals (e.g. *cannot* and *will not*). As shown in Table 3, a multiple distinctive collexeme analysis of pairs of contracted and uncontracted modals and negations of modals confirms this observation. Bolded cells show significant preferences (based on *p*-values, collocation strength > 1.3, *p*<0.05).

**Table 3. Contraction of modal collexemes and apprehensive constructions**

<<INSERT TABLE 3 HERE>>

Table 3 indicates that contracted and uncontracted modals including negated modals are differentially preferred by the two apprehensive constructions. The differential preference is significant with the two high-frequency modals *cannot/can't* and *will'll*. Specifically, *I am afraid* prefers *cannot* and *will*, but *I'm afraid* prefers *can't* and *'ll*. To what extent this contrast can be attributed to style or register difference remains to be seen and further evidence is required for a definitive conclusion. We will return to this question in a later part of the analysis (4.2.2) with contrastive data on preferences of register, grammatical person, and event type. For now, let us look at some examples of the two apprehensives with modal auxiliaries:

- (10) *I'm afraid I can't afford professional rates.* [BNC: 43148]
- (11) *I'm afraid you'll never love me.* [BNC: 33137]
- (12) *I'm afraid it will be tougher, yes.* [BNC: H9H]
- (13) *Well, I'm afraid we won't be needing you.* [BNC: 55848]
- (14) *I'm afraid he's going to be disappointed.* [BNC: 117547]
- (15) *I'm afraid I must insist on fulfilling my contract.* [BNC: JXT]
- (16) *I'm afraid we shall have to keep this room locked for a day or two.* [BNC: HWP]
- (17) *I'm afraid we have to accept the fact that Marius did die from natural causes.* [BNC: GUF]
- (18) *I'm afraid it may be no more than the zoological equivalent of chauvinistic snobbery* [BNC: H7X]
- (19) *Oh, no, thank you, but I am afraid I have to hurry home.* [BNC: AD1]
- (20) *I am afraid I cannot comment at all.* [BNC:K5M]
- (21) *I am afraid we will be slaughtered and embarrassed.* [BNC:EB3]

On the whole, based on the information in Table 1 and Table 2, and given what is known about the relative degree of subjectivity conveyed by different types of modality, the following observations can be highlighted with regard to the reliability of the assertion associated with the apprehensives *I'm/am afraid*. First, since dynamic modality is the least subjective (Hoye 1997:43) and the apprehensives are most strongly attracted to the dynamic *can't/cannot* asserting the speaker's own inability, one can derive the minimal subjectivity and high reliability of the assertions associated with the apprehensives involving *can't/cannot*. Second, the high attraction of futurity (*'ll, won't, be going to, will, and shall*) to the apprehensives (relative to epistemic modals of lower certainty) can be seen as an indicator of the inevitability of the event being predicted, thus the reliability of the assertion. The high attraction of the deontic *must* and *have to* to the apprehensives also bespeaks the certainty with which the speaker makes the assertion about a compelling pressure.

In addition to modal auxiliaries, lexical modals communicate the speaker's judgment about the event introduced by the apprehensive. Table 4 lists three items resulting from the collocation search by log likelihood within the 1-5 window to the right of the node phrase *I'm afraid*.

**Table 4. Lexical modals as collocates of *I'm afraid***

<<INSERT TABLE 4 HERE>>

From Table 4 we can see that the lexical modals are all in a negative form, expressing either impossibility or inability. For examples, consider (22)-(24):

- (22) *I'm afraid it's impossible for me to give you a definite answer.* [BNC: JXS]  
 (23) *But I'm afraid it's not possible. You cannot leave your parents just because you want to.*  
 [BNC: CH4]  
 (24) *I'm afraid I wasn't able to do a great deal more on that.* [BNC: FLS]

Note that in (22) and (23) the embedded clause is a realis statement with an empty subject. The empty subject is predicated by the copula and the adjectival modal in the present tense, followed by a first-person oblique argument and an infinitival complement stating the impossibility of an event. According to Nuyts (2001), this structure differs from its modal counterpart in the form of *I can't give you a definite answer* in that it signals intersubjectivity rather than subjectivity. Because (22) is a realis statement and demotes the first person to an oblique position, we see it as being more impersonal, and take Nuyts' notion of intersubjectivity to imply increased objectivity.

In addition to modals, the apprehensives also attract an array of negators, both absolute and implicit.<sup>14</sup> In Table 5, below, we provide the negators collocating with *I'm afraid*, ranked by the strength of their attraction to the apprehensive.<sup>15</sup>

---

<sup>14</sup> We follow Frawley (1992) in treating *rather* as an "implicit" negator.

**Table 5. Negators as collocates of *I'm afraid***

&lt;&lt;INSERT TABLE 5 HERE&gt;&gt;

Before attempting to make sense of this result, we shall address the relationship between negation and modality. Frawley (1992:384-436) defines modality as concerning “the factual status of information”, and treats negation as part of modality because it pertains to the denial, rejection or reversal of the assertability of a proposition. Given that positive polarity is the default polarity (Huddleston and Pullum, 2002), negation presupposes positive polarity. It can be assumed that to be able to deny, reject, or reverse a positive proposition, the speaker has to be certain of the invalidity, falsity or deniability of the proposition. In other words, denials, rejections, and reversals of an assertion require certainty because these negative propositions signal “total divergence of the reference and expressed world”, as Frawley puts it. Recent empirical research on negation confirms this view. Based on natural language data and experimentation, Giora (2006) argues that negation is pragmatically and functionally similar to affirmation. In a comprehension experiment using a lexical decision task, Hasson and Glucksberg (2006) found that negations are perceived as affirmations for the initial 1000ms of comprehension. These studies suggest a symmetrical view of negation and affirmation in terms of certainty. Following this view, the collocations of the apprehensive and the different forms of negation, as shown in Table 3, imply a high degree of certainty on the part of the speaker. Consider the following examples:

- (25) *I'm afraid there's no real doubt about it* [BNC: B20]  
 (26) *I'm afraid I don't speak French at all.* [BNC: HGD]  
 (27) *I'm afraid nothing of the sort is hidden from him.* [BNC: FRH]  
 (28) *I'm afraid you're going nowhere, young lady!* [BNC: H97]

The high certainty associated with negation is particularly notable in the negation of existential quantifiers in the data. In BNC all the instances of the existential quantifiers *any* and *anything* identified as collocates of the apprehensive occur with negation. According to logicians, philosophers, and semanticists (Fitch, 1979; Lyons, 1977), negated existential quantification (e.g. *not any*, *not anything*) is a semantic equivalent to universal quantification, which is a conceptual equivalent to necessity in the domain of quantity. Consider (29)-(30) as examples:

---

<sup>15</sup> We focus on *I'm afraid* because of its overall higher frequency of use than *I am afraid*. But since the contraction of a negator is stylistically relevant, we will compare *I'm afraid* and *I am afraid* in term of negator contraction preference in section 4.2.2, where we discuss register in detail.

- (29) *No, I'm afraid there isn't any cure for this disease* [BNC: HH3]  
 (30) *Really? I'm afraid I don't know anything about that* [BNC: H8Y]

Apart from certainty, the strong attraction of the apprehensive to negators reflects the negative affective valence of the proposition being introduced. That is, the proposition is about something that is contrary to the listener's expectations or undesirable to the listener, which is a defining feature of the apprehensive.

So far, the collocation data on modal auxiliaries, lexical modals, and negation point to a high degree of certainty in the speaker's proposition when s/he introduces the proposition with the apprehensives *I'm/am afraid*.

#### 4.2.2 *I fear*

In English another apprehensive denoting endophoric evidential is *I fear*. In this section we provide collocation data on *I fear*, seeking to shed light on the extent to which the uses of *I fear* and *I'm/am afraid* converge and differ. The results of the collostructional analysis of the Mod slot of [*I fear (that)* Sub Mod V] based on manual annotations are presented in Table 6.

**Table 6. Modal auxiliaries as collexemes of *I fear***

<<INSERT TABLE 6 HERE>>

The results in Table 6 show a number of dissimilarities, alongside some similarities, between *I fear* and *I'm/am afraid* in regard to their preferences of collexemes. First, *I fear* prefers the predictive *will* more strongly than *I'm/am afraid* ( $\chi^2(1)=182.3$ ,  $p<.0000001$ ). Second, *I fear* prefers the epistemic possibility modal *may* much more strongly than *I'm/am afraid* ( $\chi^2(1)=55.7$ ,  $p<.0000001$ ). This can be taken as an indication that *I fear* is less pragmatized and remains more congruent with uncertainty associated with real fear, a proposition to be confirmed by subsequent data on the differential preference of realis events by the three apprehensives. In addition, as Table 7 below shows, *I fear* is used with first person complement clause subjects significantly less often than *I'm/am afraid* ( $\chi^2(1)=31.9$ ,  $p<.00000002$ ). At the same time, *I fear* and *I'm/am afraid* disfavor or repel roughly the same items including *can*, *need*, *want*, and *should*. Some uses of *I fear* are given in (31)-(38).

- (31) *But I fear that the differences between us will grow.* [BNC: J1L]  
 (32) *We are so far behind I fear we shall never catch up.* [BNC: AJ3]  
 (33) *I fear Lumley may need replacing.* [BNC: EVG]  
 (34) *But I fear that you may not be able to do that.* [BNC: J9S]  
 (35) *It is not certain, but I fear it must be.* [BNC: H8A]  
 (36) *... so something must be done quickly, I fear, about the management* [BNC: F7V]  
 (37) *I fear we have to blame ourselves in such matters.* [BNC: HA2]

(38) *And I fear I cannot answer that question to anyone's satisfaction.* [BNC: J56]

**Table 7. Distribution of apprehensive tokens across person and number of embedded clause subjects**

<<INSERT TABLE 7 HERE>>

We can see from Table 7 that the apprehensive *I'm afraid* is used to qualify a proposition about the speaker him/herself nearly three times as frequently as *I fear* ( $\chi^2(1)=70.3$ ,  $p<.0000001$ ). *I fear* is used 1.3 times as frequently as *I'm afraid* to describe a third-person event or situation ( $\chi^2(1)=20.5$ ,  $p=.000006$ ). This finding suggests that *I fear* is seldom used to avoid responsibility for one's own future action, a common function of *I'm afraid*. The situation with *I am afraid* is more similar to *I'm afraid* than to *I fear*. The difference between *I'm afraid* and *I am afraid* in first-person singular use only approaches significance ( $\chi^2(1)=3.7$ ,  $p=.053$ ). On the other hand, *I fear* is often used to indicate real fear. Based on our annotations of the 250-token sample of each item, only 3 tokens of *I'm afraid*, and 11 tokens of *I am afraid*, unambiguously refer to real fear, compared to 103 tokens of *I fear* ( $\chi^2(1)=86$ ,  $p<.0000001$ ). *I'm afraid* and *I am afraid* appear to be significantly further along on the path from an expression of actual fear to a marker of polite refusal.

As far as polarity is concerned, *I fear* is similar to *I'm afraid* and *I am afraid* in attracting negators. However, the three differ on which negator they attract more strongly. Table 8 presents the relative frequencies (per 100 tokens) of the two main negators – *not* and *n't* – as collocates of the three apprehensives.

**Table 8. Negators as collocates of the apprehensives**

<<INSERT TABLE 8 HERE>>

As can be seen in Table 8, *I'm afraid* prefers the contracted negator over the non-contracted, and the opposite can be said of *I am afraid* and *I fear*. This suggests a possible difference between the three in formality or register, with *I am afraid* and *I fear* being more formal than *I'm afraid*. Subsequent contrastive frequency counts across registers using the text-type function in the WSE, as shown in Table 9, provide convergent evidence of this difference ( $\chi^2(1)=24.49$ ,  $p=.0000007$  for *I'm afraid* vs. *I fear*,  $\chi^2(1)=13.1$ ,  $p=.0003$  for *I am afraid* vs. *I fear*).

**Table 9. Distributions of the apprehensives across registers in BNC**

<<INSERT TABLE 9 HERE>>

Table 9 shows that *I am afraid* is most likely used in a written text, thus the most formal of all three. A further examination of the three items with regard to their syntactic freedom suggests an inverse correlation between the degree of formality and the degree of syntactic freedom. Table 10 provides the relative frequency (per 100) of each of the three apprehensives in clause-initial, clause-medial, and clause-final position, and the percentage of clause-initial occurrences of each apprehensive that takes the complementizer *that*. The more an apprehensive occurs in a slot other than the clause-initial position, the freer it is syntactically. The bottom row of Table 10 shows the proportion of clause-initial tokens of each apprehensive that takes the complementizer *that*, which is another cue of the formality of an apprehensive construction. The calculations were based on manual counting in a 250-token sample of each apprehensive.

**Table 10. Syntactic freedom of the apprehensives**

<<INSERT TABLE 10 HERE>>

Table 10 suggests that *I'm afraid* is the freest of the three apprehensives because it occurs more likely in a non-initial clausal slot than *I am afraid* and *I fear*. However, only the difference between *I'm afraid* and *I am afraid* is statistically significant ( $\chi^2(1)=9.5$ ,  $p=.002$ ); *I fear* does not significantly differ from either *I'm afraid* ( $\chi^2(1)=3.3$ ,  $p=.07$ ) or *I am afraid* ( $\chi^2(1)=1.4$ ,  $p<.13$ ). On the other hand, *I am afraid* is the least free of all three: it is most likely to occur clause-initially and to take the complementizer *that*. It is least likely to occur clause-finally, a tendency that is associated with reduced syntactic freedom in writing. The low level of syntactic freedom of *I am afraid* is consistent with its high level of formality as suggested in Table 9. These features of *I am afraid* may also be taken as indicators that this apprehensive is used more emphatically than the other two for the purpose of drawing attention to the speaker's upcoming proposition.

So far, the data on the interaction of the apprehensive and the reliability of information are primarily based on collocation information on modal categories. However, as Givón (1993:172) rightly points out, modal auxiliaries are “irrealis operators *par excellence*”. As a marked category of epistemology they tell but part of the story in the investigation of the reliability of an assertion. A comprehensive examination of the reliability of assertion must also take into consideration collocates in the unmarked category of modality – realis. Following Givón's (1993) insight into the predictable correlations between tense-aspect and epistemic modality, we consider non-modal assertions encoded in the present tense (PRS), perfect (PFV) aspect, and past tense (PST) as realis assertions whose reliability can be taken for granted. By contrast, all events qualified by modality including predictive modality (futurity) are instantiations of irrealis events. Table 11 presents the distributions of the three apprehensives across the two proposition types based on a 250-token sample of each apprehensive.

**Table 11. The apprehensives and realis vs. irrealis in BNC**

<<INSERT TABLE 11 HERE>>

The information in Table 11 shows that the majority (70%) of all the events in the sample that are introduced by *I'm afraid* or *I am afraid* are realis events or facts. We may call this a factuality tendency. What it suggests to us is the speaker's confidence in the reliability of the events on the one hand. On the other hand, the factuality bias is at odds with the anticipatory nature of fear. That is, the emotion of fear is less naturally evoked by events whose certainty is already established and known than by ones which are unconfirmed in factuality and therefore potentially frightening. When an EOF is indeed used to introduce a fact, then its role must be reanalyzed as beyond mere EOF. In other words, the factuality bias that characterizes the collocation pattern of *I'm afraid* enables the assumption that the apprehensive has undergone a semantic change from referential EOF to an endophoric evidential for the purpose of constructing epistemological stance under pragmatic pressure. The plausibility of such a scenario of semantic change is even more obvious if we consider the further observation that *I'm/am afraid* are often used to introduce what one may call a *fait accompli*, something that is known as a fact at the moment of speech. Events described as temporarily coinciding with the time of speech – the present – make up a substantial portion (69%) of the realis events, as in (39)-(41):

- (39) *I'm afraid I have some bad news for you.* [BNC: HTJ]  
 (40) *I'm afraid you misunderstand me.* [BNC: H82]  
 (41) *I'm afraid there's no real doubt about it.* [BNC: B20]

Compared to *I'm afraid* and *I am afraid*, *I fear* shows an equal divide between realis and irrealis distributions (the proportion of irrealis uses is significantly lower for *I fear* compared to *I'm/am afraid*:  $\chi^2(1)=30.4$ ,  $p<.000001$ ). The lack of a factuality tendency suggests that this apprehensive is less prototypical, that is, less pragmatized as a discourse marker. This is a finding consistent with the differential distributions of the two apprehensives across grammatical persons, the differential collocation strengths of modals expressing epistemic possibility in the two constructions, and the differences in proportion of tokens referring to actual fear. Note that the factuality tendency of *I am afraid* is closer to *I'm afraid* than to *I fear*, suggesting that it is at a level of pragmatization similar to its contracted counterpart (*I am afraid* and *I'm afraid* are not significantly different from each other in their preference for factual complements  $\chi^2(1)=.5$ ,  $p=.49$ , but both are significantly different from *I fear*). This confirms that the difference between *I am afraid* and *I'm afraid* is primarily of stylistic nature.

To summarize the collocation patterns of the three apprehensive constructions in BNC, the following generalizations suggest themselves: (1) all three apprehensive constructions strongly prefer the predictive modal *will/ll* which signals high certainty. (2) *I'm afraid* and *I am afraid* also attract *can't/cannot* in its first person dynamic use. (3) Compared with *I'm afraid* and *I am afraid*, *I fear* is much more strongly attracted to the epistemic *may*, suggesting its stronger compatibility with uncertainty and less advanced stage of pragmatization. (4) The differential distributions across grammatical persons and event types indicate that *I'm afraid* and *I am afraid* differ from *I fear* in discourse function: *I'm/am afraid* are more often used to introduce a realis (specifically, present) state of affairs than an irrealis one. By contrast, *I fear* is equally divided between realis and irrealis events. This difference, again, may be attributed to the differing

degrees of pragmatization of the apprehensives. (5) Stylistically, *I'm afraid* is less formal and more suitable for speech than *I am afraid* and *I fear*, which occur mainly in writing.

#### 4.3 Data on Mandarin Chinese

Before launching a collostructional analysis of the Mandarin apprehensive, a general observation of *kongpa* vis-à-vis the apprehensives in the other two languages is in order. Table 12 shows the relative frequencies (per million) of the apprehensives in the three languages in the three corpora I-ZH, BNC, and RWC.

**Table 12. Relative frequencies (per million) of the apprehensives**

<<INSERT TABLE 12 HERE>>

The Mandarin apprehensive *kongpa* is used more than twice as frequently as *I'm afraid* and *bojus*<sup>1</sup>, and far more frequently than *I fear* and *I am afraid*, which has the lowest frequency of all four. Given the results obtained so far, the low frequencies of *I fear* and *I am afraid* may be due to the relatively strict register-constraint they are subject to. As we will discover in a later part of this section, the Mandarin apprehensive shows a wider range of pragmatic functions including the introduction of non-negative events in addition to negative events. The extension into the appetitive and neutral domain of affective meaning may contribute to the overall relative frequency of *kongpa*.

The same data retrieval procedure as in 4.2 was followed to retrieve modal auxiliaries from I-ZH. In this process, post-query manual annotations of automatically generated collocates eliminated retrieval errors. For instance, 856 concordances of *yao* were retrieved automatically by WSE, but the actual number of *yao* as a modal turned out to be 794 after manual filtering. Also, modals that could not be generated automatically were added to the list using manual or semi-automatic annotations. For instance, the grammatical construction V-*bu*-Vres, known as the negative potential construction (NPC), expresses the speaker's negative assessment of the potential to achieve a desired state of affairs (Chao, 1968: 452; Li and Thompson, 1981:477) and thus should be included as a modal. However, it cannot be automatically retrieved because both the V and the Vres slot are lexically productive and the corpus is not designed for queries of syntactic schemas. Tedious semi-automatic annotations were therefore necessary to retrieve as many tokens of the construction as possible. Specifically, the resultative markers as part of the resultative verbs listed in Thompson (1973:377-379) as a closed lexical class were consulted because these are the candidates of the Vres slot of NPC. A combination of *bu* 'not' and each of these resultative morphemes was then entered as keyword in the item-by-item positive filtering of the automatically generated 13317 concordances of *kongpa* within the 1-5 window for instances of the NPC with different resultative types. In this process, 28 of the queried resultative types yielded a total of 342 tokens of NPC.

What makes the retrieval of modals in the Mandarin corpus most difficult, however, is the linguistic fact that the category of modal auxiliaries in Chinese is not clearly defined due to the lack of morpho-syntactic marking. As a consequence, there is no consensus on the scope, composition, and part of speech of modal auxiliaries (Song, 2004; Peng, 2007). In this study we adopted Wang's (2003:20) relatively comprehensive list of Mandarin modal auxiliaries based on collections by previous scholars, recognizing that the category is prototype-based rather than

defined in terms of necessary and sufficient features. In addition to items on this list, we included in the broad modal category *wufa* ‘cannot’, *hen-nan* ‘hardly can’, *bu-yi* ‘cannot easily’, *meifa* ‘cannot’, *bu-rongyi* ‘cannot easily’, *nanyu* ‘hardly can’, *weibi* ‘not necessarily’, and *nanmian* ‘unavoidably’ on account of their syntactic behavior and semantics. Table 13 presents the results of the colostruational analysis on the MOD slot of the construction [*kongpa* MOD VP] based on I-ZH.

**Table 13. Modal auxiliaries as collexemes of the apprehensive in Mandarin**

<<INSERT TABLE 13 HERE>>

Some general information can be gleaned from Table 13. First, modals of high certainty precede those of attenuated certainty or doubt. For instance, epistemic *yao* ‘will’, *hui* ‘will’ and *bu-hui* ‘will not’ are stronger collexemes ( $\chi^2(1)=842.7$ ,  $p<.0000001$ ) than *nenggou* ‘can’, *keyi* ‘may/can’, and *keneng* ‘can’, which are repelled by the apprehensive. Guo (2003) and Peng (2007) point out that in its epistemic sense *hui* expresses the speaker’s judgment of high probability of a future state of affairs (cf. Alleton, 1983). In fact, Lin (2005) argues that *hui* can be seen as the Mandarin counterpart of *will* as a future marker. In any case, *hui* makes a confident prediction of a state of affairs and therefore expresses high certainty. Similarly, epistemic *yao* expresses necessity in the sense of the speaker’s belief in the truth value of the occurrence of a future event (Peng, 2007). Also among the strongest collexemes are the dynamic modal V-*bu*-Vres expressing a denial of ability, and the deontic *dei* ‘have to’ expressing externally imposed necessity. Examples (42)-(45) illustrate uses of the apprehensive with these high-strength collexemes.<sup>16</sup>

- (42) 恐怕那女孩要天上的月亮，他都会想办法摘给她哩。 [I-ZH:7472]  
*kongpa na nvhai yao tian-shang de yueliang, ta dou hui xiang banfa zhai gei ta li.*  
 KONGPA that girl want sky-up ASSOC moon, he will think way pick for her MP  
 ‘I’m afraid even if that girl wanted the moon in the sky, he’d try to get it for her.’
- (43) 这次手术恐怕要花费很多钱。 [I-ZH: 21088]  
*the ci shoushu kongpa yao huafei hen duo qian*  
 this CL surgery KONGPA will cost very much money  
 ‘This surgery will cost a fortune, I’m afraid.’
- (44) 要让他们吃惊恐怕得杀人才行。 [I-ZH: 12097]

<sup>16</sup> Throughout the analysis of the Mandarin data, the following abbreviations are used: ASSOC=associative, CL=classifier, COP=copula, CRS=currently relevant state, DUR=durative, EXST=existential, LOC=locative, MP=modal particle, NEG=negator, NML=nominalizer, PN=proper name, and RES=resultative.

- yao rang tamen chijing kongpa dei sha ren cai xing*  
 want make they surprise KONGPA have:to murder person then okay  
 ‘To surprise them, I’m afraid you have to murder someone.’
- (45) 我恐怕这辈子都不会原谅他们的。[I-ZH: 15402]  
*wo kongpa zhe beizi dou bu-hui yuanliang tamen de*  
 I KONGPA this life even NEG-will forgive them MP  
 ‘I’m afraid I won’t forgive them, not in this life.’

Second, there is a third-person preference in the uses of *kongpa* with most modal expressions. Unlike the usage of *must/have to* in British English where a strong obligation is imposed on the speaker, the deontic necessity expressed by the modal *dei* is more frequently imposed on a third person or an indefinite person in Mandarin Chinese. Similarly, while in British English the speaker most frequently refers to his or her own inability when using a negative dynamic modal, e.g. *canNOT*, it is a third person that is most frequently indicated in the use of dynamic modals in Mandarin. The uses of *wu-fa* ‘cannot’, *bu-neng* ‘cannot’, and *nanyi* ‘hardly can’ more often than not refer to a situational impossibility or inability involving a third person. The skewed distribution of the apprehensive across grammatical persons becomes more evident when we compare the overall collocation patterns with regard to the grammatical person of the subject introduced by the apprehensive. Table 14, expanding on Table 7 by including frequency counts on a random sample of 250 tokens of *kongpa*, provides contrastive data on the distributions of the apprehensive structures across grammatical persons in English and Mandarin<sup>17</sup>. As can be seen in Table 14, the Mandarin apprehensive shows a strong preference for a third person subject, which is similar to *I fear* ( $\chi^2(1)=.2$ ,  $p=.64$ ), which is itself used with embedded clause subjects in the third-person much more than *I’m afraid* and *I am afraid* ( $\chi^2(1)=88.5$ ,  $p<.0000001$ ).

**Table 14. Distributions of the apprehensives across grammatical persons**

<<INSERT TABLE 14 HERE>>

Third, there is a preponderance of negation in the collexemes of the Mandarin apprehensive, which echoes what we have observed in the British English data. In particular, the Mandarin apprehensive attracts a variety of expressions of impossibility and inability. There is a whole array of negated modals in the list of attracted collexemes. Some are explicitly negative, such as

<sup>17</sup> Because zero anaphora is a common referential strategy in Mandarin (Chen 1986), the referent of a zero anaphor had to be recovered through contextual inference before the coding of its grammatical person.

*bu-hui*, *V-bu-Vres*, *bu-keneng*, *wu-fa*, *mei-fa*, *bu-rongyi*, *bu-neng*, *bu-keneng*, *bu-mian*, and *bu-yi*, where a negator is an integral part of the unit. Others are more implicitly negative, *hen-nan*, *nanmian*, and *nanyi*, for examples, communicate impossibility and inability by way of the conveyance of difficulty; *zhi-neng* and *cai-neng*, on the other hand, use a restrictive marker to limit the scope of possibility and ability. In addition to the clusters of negative modals that are strong collexemes, free negators are also preferred by the apprehensive, as shown in Table 15.

**Table 15. Negators as collocates of *kongpa* in I-ZH**

<<INSERT TABLE 15 HERE>>

The negators *bu* and *meiyou* are in complementary distributions, with the former negating unbounded situations and the latter bounded (Shi and Li, 2000). Likewise, the Classical Chinese negators *fei* and *wu* are in complementary distribution, but are largely used in idiomatic constructions or lexical chunks consistent with a literary style. In addition, *bu-shi*, the negative form of the copula *shi*, is another device of expressing denial, rejection, and reversal of an assertion. The attraction of negative modals and negators to *kongpa* bespeaks the high level of certainty at which the denial, rejection, and reversal of a proposition is made. Examples (46)-(47) illustrate the use of the apprehensive with negation.

- (46) 最忆潮州粥，这东西恐怕在北京真的吃不到了。 [ I-ZH: 64726]  
*zui yi Chaozhou zhou,zhe dongxi kongpa zai Beijing zhende chi-bu-dao le.*  
 most remember PN porridge, this thing KONGPA LOC PN really eat-NEG-RES CRS  
 ‘(I) miss Chaozhou porridge the most. I’m afraid it’s impossible to find it in Beijing.’
- (47) 但有一点是肯定的：他们做的事情恐怕不是为了国家，  
*dan you yi dian shi kending de: tamen suo zuo de shiqing kongpa bu shi weile guojia,*  
 but EXST one point be sure NML: they do NML thing KONGPA NEG COP for country  
 ‘But one thing is for sure: what they did, I’m afraid, was not for the country,  
 而是为了自己。 [I-ZH: 43157]  
*ershi weile ziji*  
 but for self  
 but for themselves.’

Lastly, we compared the Mandarin apprehensive and the three English apprehensives in regard to the affective valence of the events they introduce by calculating the percentages of positive, neutral, and negative events within a 250-token random sample of each item. Unlike the linguistic notion of polarity, affective valence is a psychological concept related to pleasure or pain in our emotional motivational system. As such its linguistic coding must be interpreted in context and with reference to world knowledge and human experience. An event is coded as positive if it is described as pleasant to a relevant participant in the discourse, as negative if it is described as unpleasant to a relevant participant in the discourse, and otherwise as neutral. We

used the context expansion function in WSE to access the larger discourse context in which the target sentence with the node word occurs. The data on the affective valence of events, as presented in Table 16, show a distinctive tolerance of non-negative events including positive events by *kongpa* ( $\chi^2(1)=275.7, p<.000001$ ).

**Table 16. Affective valence of events in the use of the apprehensive in BNC and I-ZH**

<<INSERT TABLE 16 HERE>>

As can be seen in Table 16, the three English apprehensives introduce affectively negative events or states of affairs only, while the Mandarin apprehensive is used for all three affective valences. It is worth pointing out that in the English data some instances may at first blush appear neutral, but prove negative as soon as the larger discourse context is examined. One example will suffice here as an illustration [Original conversation is presented here as a numbered sequence to ease discussion].

- (48) A1: *'I asked you a question and I'm still waiting for your answer. Do you intend to stand in my way?'*  
 B1: *(He smiled a veiled smile and sat back in his seat.) 'Yes, as a matter of fact, I do.'*  
 A2: *'You can't be serious!'*  
 B2: *'I'm afraid I'm most serious.'*  
 [BNC: JXS]

Viewed in isolation, the assertion in B2 is not negative. In the discourse context of the larger exchange between A and B, however, it constitutes a confrontational statement that is unpleasant to A.

The tolerance of non-negative events by *kongpa* suggests that this apprehensive has undergone dissociation from negativity central to the original semantics of the apprehensive associated with fear and thus a negative affective valence. Consider (49)-(51) below as examples of *kongpa* introducing positive, neutral, and negative events, respectively:

- (49) 或者什么也不聊，只是对望。  
*huozhe shenme ye bu liao, zhishi dui wang*  
 or what also NEG talk:about, only face look  
 'Or talk about nothing, but looking at each other.  
 恐怕再也没有比这更亲密温暖的事了。 [I-ZH: 22187]  
*kongpa zaiye meiyou bi zhe geng qinmi wenuan de shi le*  
 KONGPA ever NEG compare this more intimate warm ASSOC thing CRS  
 I'm afraid nothing gives you more intimacy and warmth than this.'
- (50) 让孩子健康成长，恐怕是每一位父母最大的心愿。 [I-ZH: 39301]

*rang haizi jiankang chengzhang, kongpa shi meiyi wei fumu zuida de xinyuan*  
 let child healthy grow, KONGPA COP every CL parent biggest ASSOC wish  
 ‘I’m afraid getting their child grow up healthy is the biggest wish of every parent.’

(51) 小久儿，恐怕你得学着做饭了。[I-ZH: 63380]

*xiaojiuer, kongpa ni dei xue-zhe ziji zuofan le*  
 PN, KONGPA you have:to learn-DUR self cook CRS  
 ‘Little Jiu, I’m afraid you have to learn to cook for yourself now.’

As we can see, (49) describes a pleasant state of affairs and the speaker uses a universal statement to strengthen the positive assertion about it. Similarly, in (50) which is a neutral statement, the superlative in the statement conveys high certainty and absolute knowledge. Compared with (49) and (50), (51) describes a negative situation – the speaker’s regret that she cannot cook for the listener anymore. As we will argue in section 5, when the information introduced by the apprehensive is non-negative, the primary purpose of the apprehensive is nothing but to downplay the epistemic force generated by the high certainty of the assertion.

In summary, as all the examples of *kongpa* show, the Mandarin apprehensive introduces a high-certainty claim or supposition about something. This is particularly clear in cases where the claim or supposition is prefaced with a concession of the impossible, as in (42), or a hyperbole, as in (44) and (45), in cases of openly declared certainty, as in (47), and in cases involving a superlative, as in (49) and (50). It is worth noting that unlike the apprehensives in English, which introduce undesirable events, *kongpa* can be used to introduce neutral or even positive events, although the majority of its uses are still associated with unpleasant news. This extension of the Mandarin apprehensive indicates its advanced level of pragmatization whereby the function of the word has risen above the original condition of usage and a more general discourse effect has emerged, to which we will return in section 5 in our discussion of the pragmatic functions of the apprehensive.

#### 4.4. Data on Russian

Just as with British English and Mandarin Chinese, the collocation data on modality were obtained for Russian in a two-step process. Collocations were then retrieved from RWC on the basis of log likelihood. The collocates had to occur in the corpus at least 10 times and in the filtered sample of apprehensives at least three times. All three times had to be an actual occurrence of the modal with an infinitival complement in the clause introduced by the apprehensive, which was ensured by manual inspection. Since there is a vast number of verbs and adverbs that can take an infinitival complement in Russian, we decided to include only the verbs and adverbs with the highest log likelihood capping the number such that the word with the minimum log likelihood in the table is still significantly collocated with the apprehensive at the Bonferroni-corrected .05 level (where  $p_{crit} = .05$  divided by the number of rows in table, and minimum log likelihood is  $-1.82 \cdot \ln(p_{crit}) - 1.6819$ ).

In this case, automatic analysis, at least as provided through the WSE, was clearly insufficient. First, we were unable to distinguish between the deontic sense of 'can/may' and the epistemic sense of 'can/may'. This requires manual inspection of the results. Second, the retrieval procedure yielded many false positives that are not actual examples of the apprehensive or that do not contain a modal in the right structural position in the clause that is modified by the apprehensive. Third, it was not obvious whether or not a modal is negated in any given instance. Thus we manually analyzed each of the hits retrieved using the procedure above, eliminating

tokens that did not meet the selection criteria, disambiguating epistemic vs. deontic modality and annotating each token for whether or not the modal was negated (using *ne*, *vrjad li*, *malo*, *trudno poverit' chto*, etc.). We then conducted collocation analysis on the results, which are presented in Table 17. The results reveal more similarities than differences between Russian and the other two languages. Like in English, the future *bud-* ‘will’ indicating strong predictability or epistemic certainty tends to be strongly attracted to the apprehensive. However, the collocation strength for *will* is much lower in Russian than in English with *I’m afraid*. This may be due to the fact that the Russian future marker is only used in the imperfective aspect (lexically specific verbal prefixes are used in the perfective), and future imperfectives are relatively uncommon. Furthermore, perfective/to-be-completed events might induce fear and require apology more often than imperfective events.

Dynamic modals indicating ability to do something tend to be negated and the negated forms are strongly attracted to the apprehensive, a tendency more salient with *I’m/am afraid* than with *I fear*. Likewise, deontic modals that indicate having to do something (due to some external force or the pressure of circumstances) are not negated and are significantly attracted to the apprehensive. Furthermore, just like in English, modals indicating desire to do something are strikingly absent. They are in fact repelled from the apprehensive (e.g., *xochu* ‘I want’ never occurs with *bojus'* in the sample and is repelled from the construction with a strength of 15.68). The one desire modal that is attracted to the apprehensive is *xochetsja*, which places the one who wants something into an oblique position (‘to me verbing will make itself wanted’). One does not claim responsibility for the regrettable events introduced by the apprehensive. Unlike the highly pragmaticized *I’m/am afraid* but like *I fear*, the Russian apprehensive attracts the epistemic modal ‘may’ (*mozhet*) quite strongly, suggesting a tolerance for uncertainty and a less advanced level of pragmatization.

**Table 17. Modal auxiliaries as collexemes of the apprehensive in Russian**

<<INSERT TABLE 17 HERE>>

Auxiliaries are not the only constructions that are reported to denote modality in Russian. Krause (2008) and Razlogova (2004) provide descriptions of modal adverbials in Russian that explicitly address speaker certainty associated with the use of these forms. In the case of Krause (2008), the hypotheses about differences in reliability indicated by the various adverbs are supported empirically by the results of a rating study involving minimal pairs of sentences differing only in the adverbial. Most of these, however, are repelled from the apprehensive, with the exceptions of ‘most likely’, ‘may be’, ‘possibly’, and ‘it appears’.

**Table 18. Modal adverbs as collexemes of the apprehensive in Russian**

<<INSERT TABLE 18 HERE>>

As Table 18 shows, attraction is weakly correlated with rated reliability of the information (Spearman  $\rho(16)=.43$ ,  $p=.08$ ), with adverbs rated as being less certain being more likely to be attracted to the apprehensive.<sup>18</sup> However, we believe that reliability is not the explanation for why some of these adverbs are attracted and some are repelled. The repelled adverbs fall into several categories that are, in various ways, incompatible with the meaning of the apprehensive. One category consists of the super-certain modals *bezuslovno* ‘undoubtedly’, *konečno* ‘certainly’, *razumeetsja* ‘of course’, and *ponjatno* ‘clearly’, which introduce old information. These are understandably repelled from the apprehensive because the apprehensive introduces new information. One does not warn the listener about something (one thinks) the listener already knows.

Another class of repelled adverbs is composed of former or current evidentials: *očevidno* ‘evidently’, *vidimo* ‘it appears’, *vidno* ‘it appears’, and *po-vidimomu* ‘by all appearances’. All of these contain the root for ‘to see’ (*vid-*) and literally mean that the information comes from visual evidence, although they have come to refer to inferential evidence as they pragmatized. *Kažetsja* ‘it seems’ can also be included in this group, as it literally means ‘shows itself’, as can *dumaju*, which means ‘I think’. The fact that these adverbs are repelled from the apprehensive is consistent with the apprehensive being an endophoric evidential, which is incompatible with other evidentials.

Another repelled adverbial is *po-moemu* ‘in my opinion’. It is not immediately obvious why this adverbial is repelled. However, *po-moemu* may mean not only ‘in my opinion’ but also ‘as I wish’, e.g., *Vs'jo budet po-moemu* ‘Everything will happen as I wish’. This, of course, is most inappropriate for an apprehensive: while the ambiguity is easily resolvable in most contexts, this is perhaps a case where even a temporary ambiguity is to be avoided due to the social consequences of misunderstanding.

Finally, *nu* ‘well’ is a grammaticalized placeholder used to buy time or signal hesitation. The fact that it is repelled from the apprehensive is somewhat surprising given that one does not want to seem too eager to say the clause preceded by an apprehensive. One strategy would be to appear hesitant by inserting *nu* into the embedded clause: *Bojus', chto ty, nu, ne sovsem prav.* ‘I am afraid you, well, are not quite right.’ However, this strategy does not appear to be favored by Russian speakers. One possible reason is that the speaker does not want to make it appear that they do not yet know what to say after using the apprehensive, as this undermines the evidential meaning of the apprehensive and may suggest that the apprehensive is used disingenuously.

---

<sup>18</sup> One caveat is that the differences in collocation strength among repelled collexemes may not be particularly meaningful as they are driven exclusively by differences in word frequency. Given that these modals do not appear in the construction in the sample at all, we do not know what the actual rate of co-occurrence is: regardless of word frequency the maximum-likelihood estimate of co-occurrence probability is zero until we observe at least one instance of the word in the construction. All we can estimate from the frequency of occurrence elsewhere in the corpus is the upper bound on this co-occurrence probability, which decreases with increased word frequency. However, the upper bound is not the most likely probability of co-occurrence (which is always zero) and its decrease with increasing word frequency should rather be attributed to increasing statistical power of the co-occurrence test. Thus, we are perhaps on safer ground predicting repulsion vs. attraction rather than degree of repulsion as measured by the significance value of the Fisher test.

An interesting property of the Russian corpus is that it contains a large number of either translations of English novels, or Russian novels that feature English-speaking characters. These characters speak in Russian but can be identified by their English names and the uses of typically English terms of address: 'sir', 'miss', 'mister', etc. that are not used in Russian except in depicting English speakers. It is thus possible to compare apprehensive uses by real English speakers in the BNC to uses by depicted English speakers in Russian, and to apprehensive uses by Russian speakers. This affords us the possibility of conducting a cross-linguistic comparison within a single-language corpus. Interestingly, markers used by English speakers and Russian speakers in Russian differ in their distribution across grammatical persons, tenses, animacy and the presence of real fear, suggesting differences in the degree to which the apprehensive is pragmatized across the two languages, with Russian showing a lower degree of pragmatization.

We drew a sample by taking all tokens of the apprehensive (N=3898, 20.7/million) and finding the first 250 tokens that came from English speakers and the first 250 that came from Russian speakers and were not literal expressions of fear. Literal expressions of fear that were intermixed with the obtained uses of *bojus'* as a discourse marker were also included in the sample. Fear was coded based on the second author's reading of the sentences. When there was any uncertainty regarding whether the speaker/writer means that s/he is afraid of the event expressed in the embedded clause, the fear value was coded as 'maybe'. Sentences that are marked 'yes' are thus not interpretable as non-literal uses of *bojus'* in the judgment of the second author.

As Table 19 shows, English speakers are more likely to use the apprehensive in sentences that are not easily interpreted as a literal expression of fear regarding the event described by the embedded clause ( $\chi^2(1)=19.68$ ,  $p<.00001$ ). Sentences that do not come from depictions of English speech are far more likely to retain at least a shade of fear. Many of these fear expressions involve a fear such that what is about to be said will not be understood or will be taken badly, e.g., *Bojus'*, *vam ne ponravitsja to, chto ja skazhu*. 'I fear you will not like what I am going to say'. In addition, English speakers are much more likely to use the apprehensive to introduce a clause with a first-person singular subject<sup>19</sup> ( $\chi^2(1)=9.71$ ,  $p=.002$ ) and singular animate subjects more generally<sup>20</sup>.

---

<sup>19</sup> We coded the person of the subject of the embedded clause based on case marking. The subject was usually the nominative-marked element, except for the "dative-subject" construction, e.g., *mne nravitsja* 'to me is pleasing', which is the unmarked way of expressing a liking for something in Russian. In such cases, the dative-marked participant was coded. When the subject was omitted but recoverable from agreement marking on the verb, it was coded based on the agreement marking.

<sup>20</sup> We were concerned that this may be a general property of sentences produced by (simulated) English vs. Russian speakers rather than something specific to the apprehensive. To test this hypothesis, we looked at the proportion of sentences containing 'sir', 'miss', 'missis', 'mister', or 'lady' that contained the nominative first person singular pronoun *ja* 'I' vs. the proportion of all sentences that contained *ja* 'I' in the Russian corpus. The rate of *ja* use is generally higher near the English terms of address (18.4% vs. 13.6%). The rate in the embedded clauses introduced by the apprehensive matches that in the general corpus for Russian speakers (13.6%) but is significantly higher for simulated English speakers (26.0%,  $\chi^2(1)=5.86$ ,  $p=.015$  vs. 18.4%). Thus there is some (albeit limited) data that the apprehensive for simulated English speakers is specifically associated with first person singular subjects. For Russian speakers, the apprehensive does not appear to be specifically associated with non-first-person subjects.

**Table 19. English vs. Russian speakers in the Russian corpus with respect to person, number, animacy of the embedded clause subject and the presence of real fear. Each cell contains English speakers/Russian speakers. Bolded cells show large differences between English and Russian speakers.**

<<INSERT TABLE 19 HERE>>

Table 20 incorporates distributions of the apprehensive across grammatical persons in English and Russian.<sup>21</sup> The contrastive data indicate that the uses of the Russian apprehensive that are not associated with real fear are more similar to the uses of *I'm afraid* and *I am afraid* than to *I fear*. Uses of the apprehensive in Russian that are interpretable as expressing real fear, however, are more similar to uses of *I fear* than *I'm afraid* and *I am afraid*.

**Table 20. Person and number of subjects of the embedded clauses introduced by the apprehensive**

<<INSERT TABLE 20 HERE>>

Table 21 shows another difference between English and Russian speakers in the corpus. Compared to English speakers, Russian speakers are more likely to use the apprehensive with embedded clauses in the future tense ( $\chi^2(1)=14.71$ ,  $p=.0001$ ; the proportion is much higher than for *I'm afraid*,  $\chi^2(1)=32.7$ ,  $p<.0000001$  and is similar to that for *I fear*,  $\chi^2(1)=0.13$ ,  $p=.71$ ; by contrast, simulated English speakers' use of *bojus'* is distributed across tenses much like *I am afraid*:  $\chi^2(1)=0.6$ ,  $p=.43$  and is much less commonly used with future tense than *I fear*,  $\chi^2(1)=13.5$ ,  $p=.0002$ ).<sup>22</sup> This is also consistent with the apprehensive being less pragmaticized in Russian than *I'm/am afraid* are in English and being similar to *I fear*: we usually fear what is yet to happen.

**Table 21. English vs. Russian speakers in the Russian corpus with respect to tense of the verb in the embedded clause and person of the subject. Discourse-marker uses only.**

<sup>21</sup> In Russian, some sentences have no subject, e.g., *Bojus', da*. 'I am afraid yes'. These are not included in these counts to make the proportions comparable across languages. The three Russian columns have 28, 31, and 2 such tokens respectively.

<sup>22</sup> But because *I am afraid* is stylistically marked, it is more appropriate to speak of the similarity between simulated English speakers' *bojus'* and *I'm afraid*, especially given that *I'm afraid* is the most common apprehensive in English. Thus, most instances of *bojus'* that are translations of English apprehensives are expected to be translations of *I'm afraid*.

<<INSERT TABLE 21 HERE>>

Table 22 shows that the Russian speakers' preference for using the apprehensive with future clauses holds even for instances of *bojus'* that are bleached of the original fear meaning ( $\chi^2(1)=10.64$ ,  $p=.001$ ). While bleached instances predominantly come from English speakers with present and past clauses, they mostly come from future clauses when produced by Russian speakers. Unbleached fear tokens mostly come from Russian speakers, but fears of the future are particularly unlikely to come from English speakers ( $\chi^2(1)=8.14$ ,  $p=.004$ ).

**Table 22. English vs. Russian speakers in the Russian corpus with respect to tense of the verb in the embedded clause and semantic richness. Discourse-marker uses only.**

<<INSERT TABLE 22 HERE>>

In summary, Russian speakers are associated with less bleached apprehensives, introducing a future or present clause with a third person, usually inanimate subject. A prototypical Russian apprehensive, some examples of which are shown in (52)-(56), expresses regret regarding the fact that what is expressed in the embedded clause is likely to happen. Since what is expressed in the embedded clause is often negated, it is often a regret that something desirable will not happen, contrary to expectations of the listener or the hopes of the speaker.

- (52) *Bojus', chto vozrast ne pozvolit.*  
I fear that age will not allow [that].
- (53) *Bojus', chto ocherednoe malo komu nuzhnoe meroprijatie.*  
I fear this is another useless event.
- (54) *Bojus', chto u nas v gruppe ni u kogo mozgov net.*  
I fear that no-one in our group has brains.
- (55) *Bojus', dazhe v temnote ee trudno pereputat' s toboj, ved' ona v tri raza tebja tolshche.*  
I fear it's hard to mistake her for you even in the dark for she's three times fatter than you.
- (56) *Bojus', chto vse skazannoe vyshe est' forma samozashchity.*  
I fear that everything above is a form of self-defense.

By contrast, a typical English apprehensive in the Russian corpus expresses a polite (or sarcastically hyper-polite) refusal by the speaker to do something, a promise/threat to do something unpleasant to the listener, or disagreement with what the listener has said. All of these are conveyed through an excuse of inability, a conflicting duty, or a dominant prevailing force. Some examples are shown in (57)-(61).

- (57) *Bojus', ja ne mogu soglasit'sja.*  
I'm afraid I can't agree.
- (58) *Bojus', chto poka ja ne mogu byt' dlja nego tem Garri, kakim on xochet menja videt'.*  
I'm afraid I can't yet be the Harry he wants me to be.
- (59) *Bojus', ja vynuzhden poprosit' vas prijti eshche raz, Katrina.*  
I'm afraid I am forced to ask you to come again, Katrina.
- (60) *Bojus', ja ne ulavlivaju vashu mysl'.*  
I'm afraid I don't catch your thought.
- (61) *Bojus', ja ne mogu pozvolit' miss Kerti ujtj.*  
I'm afraid I can't let miss Curty go.

## 5. Pragmatic functions of the apprehensive

The collocation data of the apprehensives in the three languages converge to the extent that the most salient collexeme is always one that expresses high certainty – be it denied ability, asserted predictability, or recognized obligation and compulsion. Thus, the collocation evidence confirms the idea of the inevitable bad news as articulated by Leech, and challenges the view of epistemic anxiety as proposed by Palmer and Givón. From a panchronic perspective, the degree of pragmatization of an apprehensive plays a role in how much uncertainty it tolerates. Thus, as the collostructional data indicate, while the Mandarin *kongpa* and the English *I'm/am afraid* introduce high-certainty assertions, the less pragmatized *I fear* and *bojus'* are compatible with lower-certainty propositions. Thus, in the larger picture of propositional certainty and reliability associated with the apprehensive as an endophoric evidential, there is a gradient of strength. There is also variability in the kind of lexico-grammatical patterns that each apprehensive exhibits. An obvious variability is observed along the dimension of grammatical person. Of all the four apprehensives, *I'm afraid* (whether used by real or depicted English speakers) is most likely to select the first person as subject of the embedded clause. This suggests that it is the most dialogical and interaction-oriented. While the differences between *I'm afraid* and *I am afraid* are of stylistic nature, the differences between *I'm afraid* and *I fear/bojus'* can be largely accounted for in terms of the degree of pragmatization as a discourse-driven diachronic process of language change from event-orientation to discourse orientation.<sup>23</sup>

As far as *kongpa* is concerned, although it is similar to *I fear* in its third-person preference, it exhibits a more advanced degree of pragmatization in that its scope of usage extends beyond affective negativity, the hallmark of the apprehensives in English and, for that matter, Russian. The use of *kongpa* to introduce non-negative information in addition to the more general use associated with affective negativity may account for the overall higher frequency we see when comparing it with English and Russian apprehensives. To account for the third-person preference and the non-preference of first-person propositions by *kongpa*, it seems, would require further investigations of factors beyond the level of pragmatization, which would be a goal for future research.

On balance, cross-language variability notwithstanding, we were able to show that, by and large, the source of information introduced by a fear-based endophoric evidential is reliable. Now the question arises: if the speaker has a reliable source of information for what s/he has to

---

<sup>23</sup> An in-depth diachronic analysis of the pragmatization of the apprehensives is beyond the scope of this paper, and will be pursued elsewhere.

say, what keeps him/her from saying it directly without the help of the apprehensive? In a similar vein, even if it is not real fear that constitutes the source of information, but the speaker construes the situation as if s/he were informed by real fear, what purpose does the construal serve in communication? By answering these questions we can provide an adequate characterization of the pragmatic functions of the apprehensive.

Previous research on pragmatics, especially politeness research, suggests that the apprehensive is an interpersonal strategic code used in conversation to smooth out the delivery of undesirable information (Leech, 1983; Mikes, 1984; Owen, 1983). Our quantitative analysis of the three languages enables us to fully spell out the functional specifics of the apprehensive. First, propositions introduced by the apprehensive as an endophoric evidential are not just generally undesirable (with the exception of some uses in Mandarin). They are highly reliable. It is the reliability of undesirable information that calls for the use of the apprehensive. Thus, the number-one function served by the apprehensive is best described as preparatory. It prepares the listener for the awful but reliable information that the speaker is about to announce in the forthcoming discourse.

Second, as is most noticeable in Mandarin, high-certainty propositions need not even be undesirable to justify the use of the apprehensive as a strategic device. Why? If the apprehensive serves to cushion the blow of bad news that the speaker has strong gut-level evidence for and is therefore highly certain about, what purpose does it serve when what is announced is not bad news? The answer is obvious: a high level of certainty on the speaker's part can be a source of concern in communication because certainty about information implies complete knowledge and authority over the reliability of information, which could undermine the speaker's image as a responsible interlocutor. Understandably, high certainty in one's assertion runs the risk of discrediting one's own performance. It may also cause one's assertion to be subject to potential doubt and challenge. The use of *kongpa* brings restraint to the speaker's level of knowledge. It counterbalances the impression of omniscience due to the high certainty and confidence of the speaker. As an evidential it also preempts the listener's potential challenge of the speaker's source of knowledge. In short, the apprehensive serves a protective function in addition to its preparatory function. It balances a strong statement by injecting a proper dose of reticence and by allowing the speaker to downplay his/her high level of certainty. In its protective function, *kongpa* can be seen as a device of self-presentation and impression management in terms of Goffman (1959).<sup>24</sup> Although the preparatory function and the protective function have differential emphases in communication, they are not entirely distinct, but jointly contribute to a smooth social exchange as the ultimate goal of communication. Even the cushioning effect of the preparatory function serves a protective purpose in making the upcoming proposition feel less blunt and abrupt, while helping the speaker to project an empathic and considerate image.

But how might an expression of fear have developed into a pragmatic marker with preparatory and protective function that provides lubricant for a successful social transaction? That is, what might the process of pragmatization look like? The answer to this question lies in the speaker-listener dynamic in socio-emotionally sensitive communication. The role of the listener and his/her inference are typically emphasized in grammaticalization theory. For instance,

---

<sup>24</sup> Although cultural values such as modesty, humility, and an aversion to explicit self-enhancement typical of a collective society (Hsu, 1981; Nisbett, 2003; Wu, 2004; Cai et al., 2011) could be invoked as an external motivation of the image protecting function of *kongpa*, an in-depth discussion of the role of culture is beyond the scope of this paper and will have to be pursued elsewhere.

the grammaticalization of “going to VERB” from literal locomotion in order to perform an action to a future marker can be accounted for by the fact that the speaker’s saying that s/he is walking somewhere in order to do A can lead the listener to infer that s/he intends to do A in the future and will likely do A in the future (Bybee et al., 1994:269). However, in the present case the source and target of pragmatization, or pragmatically oriented grammaticalization, differ in the certainty of the proposition they introduce. If the speaker says *I’m/am afraid, A* and the listener understands the statement literally, s/he would infer that the speaker is unsure whether A will happen: we are afraid of events that are not sure to happen, which is what makes “I am really scared. I’m afraid that the sun will rise in the east tomorrow” comical. However, as we have seen, pragmatized uses of the apprehensive, which include 98.8% of uses of *I’m afraid*, 95.9% of *I am afraid*, and 99.2% of *kongpa* in our 250-word samples, introduce high-certainty propositions. The listener may not infer that the speaker believes A to be certain to happen if s/he understands the speaker to mean that they are literally afraid that A will happen. Some step is missing. We suggest that the gap is filled in by the speaker, who jumpstarts the pragmatization process, illustrated in Figure 2.

**Figure 2. Decision network predicting the utterance of *I’m afraid, A* in a situation when the speaker (S) wants the listener (L) to know A, but thinks that A will displease L, and S genuinely does not want to offend L. Lines show causation (downward), the lightning bolt is inhibitory (but fails to overpower other reasons for telling L about A).**

<<INSERT FIGURE 2 HERE>>

The speaker might (unconsciously) follow the line of reasoning in Figure 2 in his/her construction of epistemological stance. The listener can then unravel the speaker’s reasoning to various degrees and react accordingly, for instance by inferring that the speaker thinks s/he does not like to hear bad news or that the speaker must have a pretty good reason to tell about A

despite thinking that the listener would not like to hear A. One possible reason is that the speaker is quite certain that A will happen, thinks that the listener does not know that A will happen, and feels that s/he should warn the listener despite the possible social danger. More generally, we concur with Traugott and Dasher (2005) in suggesting that some cases of grammaticalization, especially the pragmatization of politeness markers, might be triggered by the speaker using a construction in novel ways in order to invite the listener to make an inference desirable to the speaker. That is, pragmatization is a process of intersubjectification by which the speaker plays an active role in “rhetorical strategizing, indeed indexing and choreographing the communicative act”, to use the words of Traugott and Dasher (2005:5).

Having ventured an explanation for what triggers the development of the apprehensive as a pragmatic device for politeness, we shall be quick to add that the apprehensive in its prototypical non-literal use should not be seen as the conventionalization of some extra cognitive effort to convert the literal into the non-literal. Quite to the contrary, as Gibbs asserts, neither language nor thought is inherently literal (Gibbs, 1994). Given an experiential basis, the non-literal can be just as well motivated as the literal. This leads to our question about the psychological basis of the pragmatic development of the apprehensive. Namely, what makes fear a proper emotion to display in the situation that triggers the use of the apprehensive? In order to answer this question, let us take a close look at the emotion of fear as the experiential basis of the apprehensive.

## **6. The emotion of fear – psychological components and pragmatic consequences**

Little agreement exists about what is an emotion and definitions of emotion vary by the level at which emotion is conceptualized (Frijda, 2000). For the purpose of this study we take a functional definition of emotion and consider it “a superordinate program whose function is to direct the activities and interactions” of all human adaptive subprograms, mental and physical, in the individual’s interaction with the world (Cosmides and Tooby, 2000: 93). What distinguishes the experience of emotion from other experiences is that it is evaluative for the purpose of motivating adaptive response. That is, it “introduces value in a world of fact” by telling the good from the bad (Frijda, 2000:63), which orchestrates individual’s perception, cognition, and behavior for optimal survival outcomes. The communication of emotion, on the other hand, serves an important preparatory function with chain-effect in social interaction, as articulated by Scherer who writes (1984: 296):

“[It] not only allows other organisms to predict the most likely behavior of the emoting organism and to plan their behavior accordingly, it also provides feedback about the likely reaction of others to the intentional movement or expression, allowing in turn appropriate changes in one’s own behavioral plans.”

That is, by sharing our own experience of evaluated events, we prepare others for our own behavior and its consequences. At the same time we prepare ourselves for the likely response by others. This preparatory function has specific manifestations in the communication of fear – one of the most primal emotions – to which we now turn.

Fear is the number one threat-related emotion which motivates the fight-or-flight response for the purpose of self-protection and survival (Cannon, 1929; Arnold, 1960; Izard, 1977; Frijda, 1986; Lazarus, 1991; Panksepp, 1998). Because it calls for high action readiness, it

is of immediate survival relevance to humans and all other organisms that are programmed with a “system of defensive behaviour”, to use the words of LeDoux (2004:173-174). That the motivational potency of fear transcends species and is deeply entrenched in our biology is insightfully captured by LeDoux when he states: “when it comes to detecting and responding to danger, the brain just hasn’t changed much.” If fear mobilizes our attention and energy in order for us to act in self-protection and avoid potential threats, then it makes sense to assume that not only physical hazards evoke vigilance in us, but so do social perils. The reason is obvious: humans have evolved in and through social interaction throughout the history of the species. Individual survival depends in large part on one’s social fitness or the ability to understand what is at stake in interpersonal relationships, and the ability to perceive potential sources of social damage, and to avoid them in order to maintain a safe place in the community. This is the point we would like to make in explicating the experiential ground of the apprehensive and its pragmatics. In what follows we will discuss the components of fear and how the experience and communication of the emotion of fear map onto the pragmatic functions of the apprehensive.

At the affective level, the experience of fear triggers automatic physiological and behavioral response such that the emoting organism gets ready for self-defense. Psychologists have identified a number of dimensions along which to characterize fear. They are: (1) physiological arousal, (2) cognitive appraisal, (3) subjective reactivity, and (4) behavioral avoidance (Lang, 1970; Izard, 1973; Frijda, 1986; Lazarus, 1991). (1) Physiological arousal in fear includes bodily changes such as cold sweat, a racing heart, and increased pulse, triggered by the encounter with a fear stimulus. These bodily changes prepare the individual for fight or flight. (2) Cognitive appraisal involves the interpretation and evaluation of stimulus as either good or bad, followed by the identification of cause of the event. Recent findings in neuroscience suggest that cognitive appraisal is not a brain function independent of the brain’s bodily mechanisms. Rather it is embodied in the visceral functions of the body that “carry out the emotional and motivational control of behavior” (Tucker, 2007:14). (3) In fear, cognitive appraisal involves the instinctive judgment that the stimulus at hand is a source of danger. Subjective reactivity refers to what the individual feels about him/herself in the threatening situation. According to Izard (1973), the level of self-assurance is lower in an individual experiencing fear than in the experience of other negative emotions. The subjective reactivity caused by fear is therefore characterized by uncertainty and passivity. (4) Behavioral avoidance describes the action tendency of the individual in fear. Rather than approaching the stimulus, the individual seeks to avoid it. When fear is verbally communicated, all these components together impact how the listener processes the emotional information.

Cognitive appraisal of threat can be directly transferred from the psychological domain into the social domain of face-to-face interaction. Presumably, if the speaker has a reliable but unpleasant truth to tell the listener, s/he instinctively appraises the communication of the news as a potential threat to their relationship. By verbalizing fear the speaker communicates to the listener the negative appraisal as a warning signal. The listener is urged to predict and look for the threat to which the communication of fear can be attributed. The process of predicting and identifying threat in the assertion following the apprehensive allows the listener to infer the existence of something undesirable in that assertion. This way, the listener is prepared for the upcoming bad news.

Likewise, the behavioral tendency of avoidance inherent in fear has its correlate in the social domain. Because fear lowers self-assurance, the expression of fear is a signal that the speaker is reluctant and hesitant about what s/he has to assert, and that s/he plays a passive role

in making the assertion. The reluctance, hesitation, passivity, aversion and avoidance as behavioral manifestations of low self-assurance in fear all contribute to the effect that the speaker is stepping back from the high-certainty undesirable information s/he is presenting as an involuntary experiencer. For what s/he has to say is beyond his/her own control.

Furthermore, in reality, that which prompts the speaker to express (mock) fear is typically more threatening to the listener than to the speaker, as, for example, when the speaker rejects the listener's request. In this case the use of the apprehensive involves a shift of perspective in that the speaker is putting him/herself in the listener's shoes – I fear what you fear. That is to say, by saying that s/he is afraid, the speaker conveys the message that s/he anticipates and understands the distress on the part of the listener caused by the bad news. This conveyance of empathy is no doubt pragmatically significant and socially desirable. One doesn't throw the awful truth in someone's face. Instead, one prepares them for it and empathizes with them. This expression of empathy, or what Brown and Levinson (1978:160) call "commiserative usage", not only prepares the listener for the upcoming information, but also additionally minimizes the damaging impact of the upcoming assertion. For this pragmatic function to work, however, the expression of fear is crucial because the shift of perspective relies on emotional contagion or empathy. In the case of the protective function of the apprehensive, the same processes of mapping apply, except that the fear is not due to the negativity of the upcoming information but the very fact that the information is of high certainty.

In light of the psychological characteristics of fear, the use of the apprehensive which signals the speaker's low self-assurance, passivity, aversion and avoidance tendency despite high reliability of information is no longer paradoxical, but perfectly sensible. It is the unique psychological features of fear that enables the invited inference about the intention of the speaker beyond the mere expression of fear, which is essential to its pragmatization into a discourse marker. The seeming paradox is no doubt in line with the logic of social interaction. Goffman (1959:22-76) tells us that language use is one of the ways to construct a "social front". Our "performance" presupposes an audience whose reactions are of social importance to us. Following from this, the actuality of the fear in the apprehensive use is less important than the stance it serves to construe in the interaction situation. The non-literal usage of EOF, or mock fear, is an instantiation of Goffman's observation that people "forgo or conceal action which is inconsistent with" ideal social standards and therefore socially perilous and costly.

At this juncture, it may be instructive to go back to *The Carrot Seed* and observe the profound difference between the parents' assertion cushioned by the apprehensive and the same assertion made by the older brother, without that cushion:

*His mother said: "I'm afraid it won't come up."*

*His father said: "I'm afraid it won't come up."*

*His older brother said: "It won't come up!"*

At the truth-conditional level the parents and the older brother essentially make the same assertion about the carrot seed. What differentiate the two utterances are the pragmatic choices made and the interpersonal consequences of those choices. While the considerate parents commiserate with their youngster by using the apprehensive, the older brother flatly dumps the verdict at the little boy. The brother shows no sign of empathy and gives no warning of the unwelcome possible world while opening himself up to an evidential challenge ("How do you know it won't come up?").

In summary of this section, it is important to note that the basic components of fear in the domain of emotion have direct bearings on the pragmatics of the apprehensive in the social domain of verbal communication. All the observable pragmatic effects of the apprehensive including cushioning, being apologetic and empathetic, and its preparatory and protective function can be traced back to the basic components of fear as an emotion. Thus, it is no exaggeration to say that what we call politeness is experientially grounded in emotion both as an intra-psychoic phenomenon and as an inter-psychoic process.<sup>25</sup>

## 7. Conclusion and implications

This study identifies the semantic category of the apprehensive as a type of endophoric evidential. Drawing on corpus data from three languages, it is shown that the proposition being introduced by the apprehensive is associated with high certainty and reliability rather than uncertainty, as some assume. At the pragmatic level, the apprehensive makes the social transaction less threatening by preparing the listener for the upcoming inevitable unpleasant news and/or by protecting the speaker from sounding over-confident. That is, social anxiety rather than epistemic anxiety is the underlying force that calls for the use of the apprehensive as a hedge. Furthermore, the study demonstrates that as an endophoric evidential the apprehensive is a stance marker, one that is not merely concerned with the source of information and the strength of knowledge, but, more importantly, serves as an “index” (Clift, 2006:570; Traugott and Dasher 2005:5) of the speaker’s attitude and emotion for the purpose of regulating social interaction. That is, if we were to invoke the stance categories identified in Biber et al. (1999), the apprehensive is more a marker of “affective stance” than “epistemic stance”. On a diachronic note, our analysis converges with Babel (2009) in showing that common human psychological conditions, especially affective states, may motivate the development of interactional evidentials.

Here we should point out that within classical pragmatics, the interpersonal function of the apprehensive could be derived by way of Gricean implicatures generated from his conversational maxims (Grice, 1975). Specifically, the use of the apprehensive arguably violates the maxim of quality if literal fear cannot be realistically identified as the speaker’s true inner state at the moment of speech. It also arguably violates the maxim of manner, as the apprehensive can be taken as obscuring the real intention behind the utterance. Therefore, the listener, assuming that the speaker is a rational and cooperative interlocutor, must infer what the apprehensive is really for under the given circumstances of conversation. This way, the listener can reasonably arrive at a politeness-related interpersonal interpretation of the usage: the speaker is considering my feelings and is implicitly protecting them. However, we must recognize that the Gricean approach cannot reveal the deeper motivation of using an expression of fear as a hedge. This is because within such an approach the experiential continuity from real fear as an emotion to the apprehensive as an endophoric evidential that constructs the speaker’s epistemological stance for social purposes is neither necessary for, nor relevant to, the inference of an implicature. In other words, whether it is fear or any other emotion that is used non-literally would not matter as long as it is non-literal and triggers the search for an implicature. But,

---

<sup>25</sup> We haste to add that the mappings from the emotional onto the pragmatic domain suggest an experiential link between the two, which is to be differentiated from a direct historical link of language change. The latter cannot be definitively established without further empirical evidence from historical linguistic data. The search for such evidence will be a goal for future research.

“language has a heart”, as Ochs and Schieffelin (1989) say, precisely because much of what we say and how we say something is deeply grounded in how we feel and what attitudes we bring into the conversation. A pragmatic analysis solely based on logic is insufficient to explicate the deeper experiential motivation. It takes a query into the psychological characteristics of fear as a basic emotion to understand the experiential grounding of the verbal communication of fear in the apprehensive.

By separating the level of semantics and that of pragmatics, this analysis provides a possibility to resolve the debate about the role of direct evidentials in modifying the sincerity condition of an assertion. Viewed in isolation, the semantic certainty and reliability denoted by the apprehensive and the pragmatic weakening it brings out in interaction are paradoxical. Viewed together, the two tell a coherent story of human communication: because of the certainty and reliability of the proposition, pragmatic weakening is called for. In light of this integrative view of the two linguistic levels, Faller’s (2002) finding of a strengthened sincerity condition by *-mi* in Cuzco Quechua, and Chung’s (2010) finding of the opposite with Korean evidentials can be reconciled after all. Clearly, it is crucial for any linguistic analysis to specify the level at which the analysis is conducted. Ideally, both levels should be considered in relation to one another.

The implications of this study are beyond the semantics and pragmatics of the apprehensive as an endophoric evidential. Critically, the analysis points to the experiential continuity between physical and social threats, and the experiential continuity between psychological and linguistic expressions of emotion. As Ochs and Schieffelin (1989) eloquently argued a decade ago, the pragmatics of affect is an important and motivating factor of verbal communication as a social act. It is clear that the semantics of evidential markers is not only “fed by language-internal resources”, as Epps (2005:617) notes, but also by common human social and emotional experiences. The experiential continuity between language use and emotion suggests that linguistic pragmatics is experientially based and embodied.

We would like to conclude by emphasizing the value of quantitative collocation analysis in studying functionally similar constructions. From introspection the differences between *I’m afraid*, *I am afraid* and *I fear* would not have been discovered. Nor are the cross-linguistic similarities between *bojus<sup>j</sup>* and *I fear*, but not *I’m/am afraid*, obvious. Likewise, *kongpa* might have easily been taken to be the equivalent of *I’m/am afraid* and *bojus<sup>j</sup>*, which it is not in particular contexts. The actual usage patterns, both within language and across languages, would not have been revealed without the corpus-driven quantitative methods. Empirically buttressed, the findings not only deepen our knowledge of language use and language change as socially and psychologically driven processes, but also can be of practical use to translators in transmitting meaning across languages and cultures.

## Acknowledgments

Earlier versions of this paper were presented at the Universität zu Köln Linguistics Workshop (2007), ICLC 2011 Xi’An, and the Cognitive Linguistics Workshop (2011) at the University of Oregon. We are grateful to the audiences at these events for valuable discussions and comments. We thank Deborah Grant for her valuable contribution and editorial assistance. As ever, all remaining errors in the writing belong to the authors.

## References

- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Alleton, Vivienne. 1983. *Les auxiliaires de mode en chinois contemporain*. Paris: Maison des Sciences de l'Homme.
- Arnold, Magda B. 1960. *Emotion and Personality, vol. 1, Psychological Aspects*. New York: Columbia University Press.
- Babel, Anna M. 2009. "Dizque, evidentiality, and stance in Valley Spanish." *Language in Society* 38, 487-511.
- Biber, Douglas, Finegan, Edward. 1989. "Styles of stance in English: Lexical and grammatical marking of evidentiality and affect." *Text* 9, 93-124.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. New York: Longman.
- Brinton, Laurel J. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin/New York: Mouton de Gruyter.
- Brown, Penelope, Levinson, Stephen C. 1978. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Burnard, Lou. 2007. Reference Guide for the British National Corpus (XML Edition). <http://www.natcorp.ox.ac.uk/docs/URG/>. Published for the British National Corpus Consortium by the Research Technologies Service at Oxford University Computing Services.
- Bybee, Joan, Perkins, Revere, Pagliuca, William. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: The University of Chicago Press.
- Cai, Huajian, C. Sedikides, L. Gaertner, C. Wang, M. Carvallo, Y. Xu, E. M O'Mara, L. E. Jackson. 2011. "Tactical self-enhancement in China: is modesty at the service of self-enhancement in East-Asian culture?" *Social Psychological and Personality Science* 2(1), 59-64.
- Cannon, Walter B. 1929. *Bodily Changes in Pain, Hunger, Fear, and Rage*, 2nd Edition. New York/London: D. Appleton and Co.
- Chafe, Wallace. 1986. Evidentiality in English conversation and academic writing. In W. Chafe & J. Nichols (eds.), 261-272.
- Chafe, Wallace L. & Johanna Nichols (eds). 1986. *Evidentiality: The linguistic encoding of epistemology*. Norwood, NJ: Ablex Publishing.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley/Los Angeles: University of California Press.
- Chappell, Hilary. 2001. A Typology of evidential markers in Sinitic languages. In *Sinitic Grammar: Synchronic and Diachronic Perspectives*, H. Chappell (ed.), 56-84. London: Oxford University Press.
- Chung, Kyung-Sook. 2010. "Korean evidentials and assertion." *Lingua* 120 (4), 932-952.
- Clift, Rebecca. 2006. "Indexing stance: Reported speech as an interactional evidential." *Journal of Sociolinguistics* 10, 569-595.
- Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.
- Connillie, Bert. 2009. "Evidentiality and epistemic modality." *Functions of Language* 16(1), 44-62.

- Cosmides, Leda and Tooby, John. 2000. Evolutionary psychology and the emotions. In *Handbook of Emotions*, M. Lewis & J.M. Haviland-Jones (eds.), 2<sup>nd</sup> edition, 91-115. New York/London: The Guilford Press.
- Crystal, David. 1997. *The Cambridge encyclopedia of Language*, 2<sup>nd</sup> edition. Cambridge: Cambridge University Press.
- Curnow, Timothy J. 2003. "Nonvolitionality expressed through evidentials." *Studies in Language* 27, 39-59.
- Dendale, Patrick, and Tasmowski, Liliane. 2001. "Introduction: Evidentiality and related notions." *Journal of Pragmatics* 33, 339-348.
- Dixon, Robert M W. 1991. *A New approach to English Grammar*. New York: Oxford University Press
- Du Bois, John W. 2007. The stance triangle. In *Stance Taking in Discourse*, R. Englebretson (ed.), 139-182. Amsterdam/Philadelphia: John Benjamins.
- Epps, Patience. 2005. "Areal diffusion and the development of evidentiality: Evidence from Hup." *Studies in Language* 29, 617-650.
- Faller, Martina. 2002. "The evidential and validation licensing conditions for the Cusco Quechua enclitic –mi." *Belgian Journal of Linguistics* 16, 7-21.
- Fitch, Gregory. 1979. "Analyticity and Necessity in Leibniz." *Journal of the History of Philosophy* 17, 29-42.
- Floyd, Rick. 1993. The structure of Wanka Quechua evidential categories. Ph.D. thesis, University of California San Diego.
- Frawley, William. 1992. *Linguistic Semantics*. Hillsdale, N.J.: Lawrence Erlbaum.
- Frijda, Nico H. 1986. *The Emotions*. Cambridge: Cambridge University Press.
- Frijda, Nico H. 2000. The psychologist's point of view. In *Handbook of Emotions*, M. Lewis & J.M. Haviland-Jones (eds.), 2<sup>nd</sup> edition, 59-74. New York /London: The Guilford Press.
- Gibbs, Raymond. 1994. *The poetics of mind: figurative thought, language, and understanding*. Cambridge: Cambridge University Press.
- Giora, Rachel 2006. "Anything negatives can do affirmatives can do just as well, except for some metaphors." *Journal of Pragmatics* 38(7), 981-1014.
- Givón, Talmy. 1982. "Evidentiality and epistemic space". *Studies in Language* 6, 23-49.
- Givón, Talmy. 1993. *English Grammar: A Function-based Introduction*. Amsterdam/Philadelphia: Benjamins.
- Givón, Talmy. 2001. *Syntax: An introduction*. Vol. 2. Amsterdam/Philadelphia: Benjamins.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.
- Goffman, Erving. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Guo Zhaojun (郭昭军). 2003. "从‘会2’与‘可能’的比较看情态动词‘会2’的句法和语义." *Yufa Yanjiu he Tansuo* 12.
- Grice, Paul. 1975. Logic and conversation. In Cole, P. and Morgan, J. (eds.) *Syntax and semantics*, vol 3. New York: Academic Press.
- Gries, Stefan Th. 2007. Coll.analysis 3.2. A program for R for Windows 2.x. <http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/readme.txt>
- Grigorenko, M. Ju. 2009. Система средств выражения эвиденциальности в русском языке. *Мовознавство*, 17.
- Hardman, Martha J. 1986. Data-source marking in the Jaqi languages. In W. Chafe & J. Nichols (eds.), 113-136.
- Hasson, Uri, Glucksberg, Sam. 2006. "Does understanding negation entail affirmation? An

- examination of negated metaphors.” *Journal of Pragmatics* 38(7), 1015-1032.
- Hill, Jane H., Irvine, Judith T. 1993. *Responsibility and Evidence in Oral Discourse*. New York: Cambridge University Press.
- Hoye, Leo. 1997. *Adverbs and Modality in English*. London: Longman.
- Hsieh, Chia-Ling. 2008. “Evidentiality in Chinese newspaper reports: Subjectivity/objectivity as a factor.” *Discourse Studies* 10, 205.
- Hsu, Francis L. K. 1981. *Americans & Chinese: Passage to Differences*, 3rd ed. Honolulu: University of Hawaii Press.
- Huddleston, Rodney, Pullum, Geoffrey K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press,
- Izard, Carroll E. 1973. *Human Emotions*. New York/London: Plenum Press.
- Kaksin, A. D. 2010. Эвиденциальность как функционально-семантическая категория уровня модальности (на примере русского и хантыйского языков). *Вестник Челябинского Государственного Университета: Филология, искусствоведение* 13(43), 52-56.
- Kozinceva, N. A. 1994. Категория эвиденциальности (проблемы типологического анализа). *Вопросы Языкознания* 3, 92-104.
- Krause, Marion. 2007. *Epistemische Modalität: Zur Interaktion lexikalischer und prosodischer Marker*. Wiesbaden: Harrasowitz Verlag.
- Lang, Peter. 1970. Stimulus control, response control and desensitization of fear. In D. Levis (ed.), *Learning Approaches to Therapeutic Behavior Change*. Chicago: Aldine Press.
- Lazarus, Richard S. 1991. *Emotion and Adaptation*. Oxford: Oxford University Press.
- LeDoux, Joseph. 2004. *The Emotional Brain*. London: Orion Books.
- Leech, Geoffrey. 1983. *Principles of Pragmatics*. London: Longman.
- Lewandowska-Tomaszczyk, Barbara. 2011. Emotions: prototypes, culture and discourse meanings. Paper presented at ICLC 11, Xi’an, China, July 11-16.
- Li, Jinxia (李晋霞) and Liu, Yun (刘云). 2003. 从“如果”与“如果说”的差异看“说”的言据义(On the evidential meaning of *shuo*). *Yuyan Kexue* 3, 59-70.
- Lin, Jo-wang. 2005. “Time in a language without tense: The case of Chinese.” *Journal of Semantics* 23(1), 1-56.
- Lyons, John. 1977. *Linguistic Semantics: An introduction*. Cambridge: Cambridge University Press.
- Mikes, George. 1984. *How to be a Brit*. London: Penguin Books.
- Mortelmans Tanja, Boye Kasper, and van der Auwera Johan. 2009. Modals in the Germanic languages. In *Modals in the Languages of Europe: A Reference Work*, 11-69. Berlin: Mouton de Gruyter.
- Mushin, Ilana. 2001. *Evidentiality and Epistemological Stance: Narrative Retelling*. Amsterdam: John Benjamins.
- Nisbett, Richard E. 2003. *The Geography of Thought*. New York: Free Press.
- Nuyts, Jan. 2001a. “Subjectivity as an Evidential Dimension in Epistemic Modal Expressions.” *Journal of Pragmatics* 33, 383-400.
- Nuyts, Jan. 2001b. *Epistemic modality, language, and conceptualization: a cognitive-pragmatic perspective*. Amsterdam: John Benjamins.
- Nuyts, Jan. 2006. Modality: Overview and linguistic issues. In *The Expression of Modality*, W. Frawley (ed.), 1-26. Berlin/New York: Mouton de Gruyter.

- Ochs, Elinor and Schieffelin, Bambi. 1989. "Language has a heart." *Text* 9(1), 7-26.
- Owen, Marion. 1983. *Apologies and Remedial Interchanges*. Paris: Mouton.
- Palmer, Frank R. 2001. *Mood and Modality*, 2nd Edition. Cambridge: Cambridge University Press.
- Panksepp, Jaak. 1998. *Affective Neuroscience*. Oxford: Oxford University Press.
- Peng, Lizhen (彭利贞). 2007. *现代汉语情态研究 (A study of Modern Mandarin Modality)*. Beijing: Zhongguo Shehui Kexue Chubanshe.
- Plungian, Vladimir A. 2001. "The place of evidentiality within the universal grammatical space." *Journal of Pragmatics* 33, 349-357.
- Razlogova, Je. E. 2004. *Logiko-kognitivnye i stilisticheskie aspekty semantiki modal'nyx slov*. Moscow: Izdatel'stvo Moskovskogo Universiteta.
- Scherer, Klaus R. 1984. On the nature and function of emotion. In *Approaches to Emotion*, K.R. Scherer and P. Ekman (eds.), 293-318. Hillsdale, N.J.: Lawrence Erlbaum.
- Searle, John. 1983. *Intentionality*. Cambridge: Cambridge University Press.
- Sergeeva, E. V. 2010 *Tuna как эвиденциальный маркер в русской разговорной речи. Acta Linguistica Petropolitana* 6, 149-153.
- Shi, Yuzhi and Charles N. Li. 2000. "The Effects of the syntactic changes around the fifteenth century on the formation of the negative system of modern Chinese: The motivations for the grammaticalization of the negative MEI." *Linguistics Study* 2, 39-62.
- Song, Yong-gui. (宋永圭). 2004. 现代汉语情态动词“能”的否定研究. Ph.D. Dissertation, Fudan University.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. "Collostructions: investigating the interaction between words and constructions." *International Journal of Corpus Linguistics* 8(2). 209-243.
- Thompson, Sandra A. 1973. "Resultative verb compounds in Mandarin Chinese: A case for lexical rules." *Language* 49(2), 361-379.
- Traugott, Elisabeth C. 1989. "On the rise of epistemic meaning in English: an example of subjectification in semantic change." *Language* 65, 31-55.
- Traugott, Elisabeth C. 1997. Subjectification and the development of epistemic meaning: the case of promise and threaten. In *Modality in Germanic Languages: Historical and Comparative Perspectives*, T. Swan & O.J. Westvik (eds.), 185-210. Berlin/New York: Mouton de Gruyter.
- Traugott, Elisabeth C., Dasher, Richard B. 2005. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Trnavac, Radoslava. 2006. Aspect and subjectivity in modals constructions. Ph.D. dissertation, University of Utrecht.  
<http://www.lotpublications.nl/publish/articles/002062/bookpart.pdf>.
- Tucker, Don M. 2007. *Mind from Body: Experience from Neural Structure*. Oxford: Oxford University Press
- Wang, Xiaoling (王晓凌). 现代汉语情态动词语义研究 (A semantic study of Modern Chinese modal verbs). MA thesis, Fudan University.
- Weber, David. 1986. Information Perspective, Profile, and Patterns in Quechua. In W. Chafe and J. Nichols. (eds.), 137-155.

- Werkhofer, Konrad. 1992. Traditional and modern views: the social constitution and the power of politeness. In *Politeness in Language*, R. J. Watts, S. Ide, and K. Ehlich (eds.), 155-199. Berlin/New York: Mouton de Gruyter.
- Wierzbicka, Anna. 1995. *Emotions across Languages and Cultures*. Cambridge: Cambridge University Press.
- Whitt, Richard J. 2009. "Auditory evidentiality in English and German: the case of perception verbs." *Lingua* 119, 1083-1095.
- Wu, Yi'an. 2004. *Spatial Demonstratives in English and Chinese: Text and Cognition*. Amsterdam: Benjamins.
- Zhang, Chengfu (张成福), Yu, Guangwu (余光武). 2003. 论汉语的言据表达 – 以插入语研究为例 (The expression of evidentiality in Chinese). *Yuyan Kexue* 3, 50-58.
- Zhu, Yongsheng (朱永生). 2006. 试论现代汉语的言据性 (Evidentiality in modern Chinese). *Xiandai Waiyu* 29(4), 331-337.

Table 1. Modal auxiliaries as collexemes of *I'm afraid*

Collexemes	FRQ in con	FRQ in cor	Relation	Collocation strength	Dominant modal type Grammatical person
<i>can NOT</i>	84	53343	attraction	495.06	94% dynamic; 1st 85% (1sg 70%)
<i>WILL</i>	85	309181	attraction	218.36	97% predictive; 3rd 53%
<i>WILL NOT</i>	29	26803	attraction	144.89	100% predictive; 3rd 59%
<i>be going to/ gonna</i>	33	48650	attraction	135.58	100% predictive; 2nd/3rd 33%
<i>must</i>	32	67365	attraction	109.88	94% deontic; 1st 63%
<i>HAVE to</i>	23	53097	attraction	74.52	100% deontic; 1st 63%
<i>shall</i>	8	18606	attraction	25.51	100% predictive; 1st 100%
<i>might</i>	11	56693	attraction	19.79	100% epistemic; 1st 46%
<i>could NOT</i>	6	57031	attraction	11.30	50% dynamic; 1st 100%
<i>may not</i>	3	6719	attraction	9.74	100% epistemic; 1st 67%
<i>may</i>	12	120948	attraction	9.57	100% epistemic; 3rd 67%
<i>would NOT</i>	4	27858	attraction	5.24	100% epistemic; 3rd 50%
<i>shall not</i>	1	1208	attraction	4.41	100% predictive; 1st 100%
<i>should NOT</i>	2	10744	attraction	3.43	100% deontic; 2nd/3rd 50%
<i>might not</i>	1	2434	attraction	3.09	100% epistemic; 2nd 100%
<i>would</i>	11	251187	attraction	0.51	100% epistemic; 1st/3rd 46%
<i>could</i>	5	130938	attraction	0.03	100% epistemic; 3rd 80%
<i>must NOT</i>	0	2804	repulsion	-0.19	
<i>can</i>	7	241870	repulsion	-0.29	100% 1st dynamic/2nd deontic 43%
<i>want</i>	1	55245	repulsion	-0.56	100% dynamic; 1st 100%
<i>need</i>	1	55360	repulsion	-0.564	100% dynamic; 1st 100%
<i>should</i>	0	98421	repulsion	-6.96	

Table 2. Modal auxiliaries as collexemes of *I am afraid*

Collexemes	FRQ in con	FRQ in cor	Relation	Collocation strength	Dominant modal type Grammatical person
<i>cannot</i>	10	21750	attraction	68.84	100% dynamic; 1st sg. 100%
<i>will</i>	13	240332	attraction	36.55	100% predictive; 3rd 77%
<i>shall</i>	5	18606	attraction	28.69	100% predictive; 1st 100%
<i>must</i>	5	67365	attraction	16.36	100% deontic; 1st 80%
<i>could not</i>	3	13064	attraction	16.19	100% dynamic; 1st 100%
<i>HAVE to</i>	4	53097	attraction	13.14	100% deontic; 1st /3rd 50%
<i>be going to</i>	3	32908	attraction	10.87	100% predictive; 3rd 100%
<i>will not</i>	2	10883	attraction	9.90	100% predictive; 3rd 100%
<i>shall not</i>	1	1208	attraction	7.89	100% predictive; 1st 100%
<i>would</i>	5	251187	attraction	5.24	100% epistemic; 3rd 100%
<i>may not</i>	1	6719	attraction	4.53	100% epistemic; 3rd 100%
<i>might</i>	2	56693	attraction	3.83	100% epistemic; 3rd 100%
<i>would not</i>	1	12058	attraction	3.42	100% epistemic; 3rd 100%
<i>wouldn't</i>	1	15800	attraction	2.92	100% epistemic; 3rd 100%
<i>may</i>	2	120948	attraction	1.54	100% epistemic; 1st 100%
<i>'ll</i>	1	68849	attraction	0.605	100% epistemic; 1st 100%
<i>could</i>	1	130938	attraction	0.06	100% dynamic; 1st 100%
<i>might not</i>	0	2434	repulsion	-0.029	100% epistemic; 2nd 100%
<i>must not</i>	0	2804	repulsion	-0.034	
<i>be gonna</i>	0	12438	repulsion	-0.15	
<i>can</i>	1	241870	repulsion	-0.16	100% epistemic; 3rd 100%
<i>couldn't</i>	0	16163	repulsion	-0.19	
<i>can't</i>	0	30298	repulsion	-0.36	
<i>want</i>	0	55245	repulsion	-0.662	
<i>need</i>	0	55360	repulsion	-0.663	
<i>should</i>	0	98421	repulsion	-1.19	

**Table 3. Contraction of modal collexemes and apprehensive constructions**

Modals	FRE in CON1 <i>I am afraid</i>	FRE in CON2 <i>I'm afraid</i>	Preferred construction	Collostruction strength
<i>cannot</i>	10	6	<i>I am afraid</i>	5.08
<i>will</i>	13	31	<i>I am afraid</i>	2.29
<i>could not</i>	3	1	<i>I am afraid</i>	1.98
<i>will not</i>	2	2	<i>I am afraid</i>	0.99
<i>would not</i>	1	2	<i>I am afraid</i>	0.425
<i>wouldn't</i>	1	2	<i>I am afraid</i>	0.425
<i>can't</i>	0	78	<i>I'm afraid</i>	5.93
<i>'ll</i>	1	54	<i>I'm afraid</i>	2.96
<i>won't</i>	0	27	<i>I'm afraid</i>	1.91
<i>going to</i>	3	27	<i>I'm afraid</i>	0.47
<i>gonna</i>	0	6	<i>I'm afraid</i>	0.41
<i>couldn't</i>	0	5	<i>I'm afraid</i>	0.34
<i>should not</i>	0	1	<i>I'm afraid</i>	0.07
<i>shouldn't</i>	0	1	<i>I'm afraid</i>	0.07

**Table 4. Lexical modals as collocates of *I'm afraid***

Modal adj./adv.	Frequency	Collocation strength
<i>(not)able</i>	8	29.31
<i>impossible</i>	4	20.73
<i>(not) possible</i>	6	17.44

**Table 5. Negators as collocates of *I'm afraid***

Negation	Frequency	Collocation strength
<i>n't</i>	342	2230.92
<i>not</i>	191	895.46
<i>no</i>	62	271.58
<i>rather</i>	37	227.79
<i>nothing</i>	14	67.69
<i>never</i>	13	46.56
<i>nowhere</i>	3	20.90
<i>neither</i>	5	29.06
<i>nobody</i>	3	16.99
<i>none</i>	4	20.74

**Table 6. Modal auxiliaries as collexemes of *I fear***

Collexemes	FRQ in con	FRQ in cor	Relation	Collocation strength	Dominant modal type Grammatical person
<i>will</i>	67	309181	attraction	245.07	100% predictive, 3rd 78%
<i>may</i>	20	120948	attraction	57.39	100% epistemic, 3rd 65%
<i>may not</i>	7	6719	attraction	44.25	86% epistemic, 2nd 50%
<i>will not</i>	9	26803	attraction	37.19	100% predictive, 3rd 63%
<i>shall</i>	8	18606	attraction	36.81	100% predictive, 1st 100%
<i>must</i>	10	67365	attraction	26.25	80% deontic, 1st/2nd 75%
<i>cannot</i>	9	53343	attraction	25.66	100% dynamic, 1st 67%
<i>might</i>	7	56693	attraction	16.01	100% epistemic, 3rd 57%
<i>would</i>	12	251187	attraction	9.84	100% epistemic, 3rd 100%
<i>would not</i>	3	27858	attraction	6.11	100% epistemic, 3rd 100%
<i>could</i>	6	130938	attraction	4.50	100% epistemic, 3rd 67%
<i>might not</i>	1	2434	attraction	4.48	100%epistemic 1st 100%
<i>be going to</i>	3	48650	attraction	3.45	100% predictive, 3rd 67%
<i>have to</i>	2	53097	attraction	1.01	100% deontic, 1st 100%
<i>can</i>	6	241870	attraction	0.81	100% epistemic, 3rd 83%
<i>need</i>	0	55360	repulsion	1.87	
<i>want</i>	0	55245	repulsion	1.87	
<i>could not</i>	0	29227	repulsion	0.99	
<i>should not</i>	0	10744	repulsion	0.36	
<i>should</i>	1	98421	repulsion	0.31	100% deontic, 3rd 100%
<i>must-not</i>	0	2804	repulsion	0.09	
<i>shall-not</i>	0	1208	repulsion	0.04	

**Table 7. Distribution of apprehensive tokens across person and number of embedded clause subjects**

	<i>I'm afraid</i>		<i>I am afraid</i>		<i>I fear</i>	
1SG	88	35.2%	67	26%	30	12%
1PL	23	9.2%	20	8%	25	10%
2	29	11.6%	23	9.2%	24	9.6%
3	120	48%	140	56%	171	68.4%
Total	250		250		250	

**Table 8. Negators as collocates of the apprehensives**

	<i>I'm afraid</i>	<i>I am afraid</i>	<i>I fear</i>
<i>n't</i>	14.16	1.48	1.53
<i>not</i>	8.23	19.19	16.51

**Table 9. Distributions of the apprehensives across registers in BNC**

Register	<i>I'm afraid</i>	<i>I am afraid</i>	<i>I fear</i>
Written total	1505	266	297
	78.9%	98.2%	90.8%
Spoken total	404	5	30
	21.1%	1.8%	9.2%

**Table 10. Syntactic freedom of the apprehensives**

	<i>I'm afraid</i>	<i>I am afraid</i>	<i>I fear</i>
Initial	72.3	84.1	77.3
medial	11	12.6	15.5
final	16.7	3.3	7.2
Initial- <i>that</i>	8.91%	47.37%	40.97%

**Table 11. The apprehensives and realis vs. irrealis in BNC**

Proposition type		<i>I'm afraid</i>	<i>I am afraid</i>	<i>I fear</i>
Realis	PRS	<b>129</b>	<b>115</b>	<b>79</b>
	PFV	24	23	13
	PST	28	35	33
	Total	181	173	125
	Percentage	<b>72.4%</b>	<b>69.2%</b>	<b>50%</b>
Irrealis	Total	69	77	125
	Percentage	<b>27.6%</b>	<b>30.8%</b>	<b>50%</b>
TOTAL		250	250	250

**Table 12. Relative frequencies (per million) of the apprehensive across languages**

Apprehensive	Relative frequency
<i>kongpa</i>	47.9
<i>I'm afraid</i>	17
<i>I am afraid</i>	2.4
<i>I fear</i>	2.9
<i>bojus'</i>	20.7

**Table 13. Modal auxiliaries as collexemes of the apprehensive in Mandarin**

Modals	FRQ in cor	FRQ in con	Relation	Collocation strength	Dominant modal type Grammatical person
<i>hui</i> 'will'	792	472929	attraction	3536.55	100% epistemic, 3rd 74%
<i>yao</i> 'will'	794	668107	attraction	3024.89	73% epistemic, 3rd 71%
<i>bu-hui</i> 'will not'	443	132179	attraction	2536.79	97% epistemic, 3rd 73%
<i>hen-nan</i> 'hardly can'	242	18530	attraction	2026.68	90% epistemic, 3rd 80%
<i>V-bu-Vres</i> 'cannot'	342	106100	attraction	1923.62	100% dynamic, 3rd 53%
<i>dei</i> 'have to'	276	85646	attraction	1547.77	99% deontic, 3rd+IND 64%
<i>bu-neng</i> 'cannot'	277	157284	attraction	1228.45	43% dynamic; 1st 56%
<i>nanyi</i> 'hardly can'	152	17506	attraction	1145.63	100% epistemic, 3rd 96%
<i>zhi-neng</i> 'can only'	134	51937	attraction	688.70	95% dynamic, 3rd 53%
<i>wu-fa</i> 'cannot'	117	55349	attraction	555.65	100% dynamic, 3rd 77%
<i>weibi</i> 'not necessarily'	50	5181	attraction	386.13	100% epistemic, 3rd 98%
<i>bu-rongyi</i> 'cannot easily'	42	8879	attraction	264.85	100% epistemic, 3rd 98%
<i>xuyao</i> 'need/have to'	77	106833	attraction	209.69	100% deontic, 3rd 65%
<i>nanmian</i> 'unavoidably'	24	4207	attraction	160.12	100% epistemic, 3rd 92%
<i>bu-yi</i> 'cannot easily'	23	4058	attraction	153.15	100% epistemic, 3rd 78%
<i>bu-keneng</i> 'cannot'	35	25295	attraction	137.19	100 epistemic, 3rd 60%
<i>nanyu</i> 'hardly can'	14	970	attraction	119.26	100% epistemic, 3rd 62%
<i>N-nenggou</i> 'cannot'	17	2656	attraction	117.27	100% dynamic, IND 82%
<i>bu-gan</i> 'dare not'	36	38064	attraction	115.52	100% dynamic, 3rd 69%
<i>cai-neng</i> 'cannot unless'	30	27663	attraction	103.83	77% dynamic, 3rd 83%
<i>bujiande</i> 'probably not'	12	2023	attraction	80.95	100% epistemic, 83%
<i>mei-fa</i> 'cannot'	16	6781	attraction	79.02	100% dynamic, 1st+2nd 50%
<i>buyiding</i> 'not necessarily'	15	7945	attraction	61.62	100% epistemic, 3rd 73%
<i>geng-rongyi</i> 'more easily'	6	1957	attraction	32.68	50% epistemic, 3rd 100%
<i>wucong</i> 'cannot'	6	2207	attraction	31.27	100% dynamic, IND 50%
<i>bumian</i> 'unavoidably'	7	4419	attraction	29.19	100% epistemic, 3rd 100%
<i>yinggai</i> 'should'	30	122124	attraction	29.00	100% deontic, 3rd 60%
<i>weijiande</i> 'not necessarily'	3	126	attraction	28.57	100% epistemic, 3rd 67%
<i>budebu</i> 'have to'	10	17255	attraction	23.25	100% deontic, 3rd 50%
<i>gai</i> 'should'	14	61139	attraction	12.17	86% deontic, 1st 67%
<i>bu-nan</i> 'not difficult'	3	4194	attraction	8.09	100% epistemic, 3rd 100%
<i>you-keneng</i> 'it's possible'	3	11357	attraction	3.19	100% epistemic, 3rd 67%
<i>bu-keyi</i> 'may not'	2	6106	attraction	2.77	100% deontic, 1st/3rd 50%
<i>bu-gai</i> 'should not'	2	7531	attraction	2.15	100% deontic, 1st/2nd 50%
<i>ying</i> 'should'	7	55232	attraction	1.55	100% deontic, 3rd 57%
<i>bixu</i> 'must'	6	68331	attraction	.12	100% deontic, 2nd 67%
<i>nenggou</i> 'can'	4	58307	repulsion	-.04	100% dynamic, 1st 75%
<i>keyi</i> 'may'	19	365995	repulsion	-3.19	79% epistemic, 3rd 73%
<i>keneng</i> 'can'	4	143767	repulsion	-5.85	100% epistemic, 3rd 100%

**Table 14. Distributions of the apprehensives across grammatical persons**

	<i>I'm afraid</i>		<i>I am afraid</i>		<i>I fear</i>		<i>kongpa</i>	
1SG	88	35.2%	67	26%	30	12%	27	10.8%
1PL	23	9.2%	20	8%	25	10%	5	2%
2	29	11.6%	23	9.2%	24	9.6%	17	6.8%
3	120	48%	140	56%	171	68.4%	177	70.8%
IND	0				0		24	9.6%
Total	250		250		250		250	

**Table 15. Negators as collocates of *kongpa* in I-ZH**

<b>Negators</b>	<b>Frequency</b>	<b>Collocation strength</b>
<i>bu</i>	1062	4138.29
<i>meiyou</i>	613	2609.32
<i>bushi</i>	323	1545.29
<i>wu</i>	35	84.1
<i>fei</i>	31	110.55

**Table 16. Affective valence of events in the use of the apprehensive in BNC and I-ZH**

<b>Affective valence</b>	<b><i>I'm afraid</i></b>	<b><i>I am afraid</i></b>	<b><i>I fear</i></b>	<b><i>kongpa</i></b>
<b>positive</b>	0	0	0	10 4%
<b>neutral</b>	0	0	0	61 24.4%
<b>negative</b>	250	250	250	179 71.6%
Total	250	250	250	250

**Table 17. Modal auxiliaries as collexemes of the apprehensive in Russian**

Modals	Negated	Negated in random sample	FRQ in con	FRQ in cor <sup>1</sup>	Collostruction Strength
<i>smogu</i> 'I will be able to'	97%	51%	81	5848	196.27
<i>pridetsja</i> 'will be forced to'	2%	8%	104	19147	185.24
<i>mozhet</i> 'it/(s)he may'	4%	25%	72	58659	99.04
<i>poluchitsja</i> 'it will be possible to'	97%	34%	30	6511	58.88
<i>budet</i> 'it/(s)he will'	0%	10%	41 <sup>2</sup>	55833	47.50
<i>mogu</i> 'I can'	94%	51%	33	48803	37.51
<i>budu</i> 'I will'	33%	26%	10	21636	33.06
<i>uspeju</i> 'I will have time to'	100%	38%	12	672	31.10
<i>nevozmozhno</i> 'be impossible to'	0%	0%	22	22568	28.70
<i>stanet</i> 'it/(s)he will start/become'	10%	11%	20	16348	28.10
<i>mogut</i> 'they can'	0%	18%	22	24481	27.94
<i>pozдно</i> 'too late to'	0%	10%	13	2891	25.83
<i>udastsja</i> 'will be achievable to'	100%	41%	13	5945	21.78
<i>mozhem</i> 'we may'	0%	32%	6	16905	17.57
<i>budut</i> 'they will'	6%	4%	15	27297	16.07
<i>sumeju</i> 'I will be able to'	83%	40%	6	533	14.66
<i>smozhet</i> 'it/(s)he will be able to'	90%	46%	10	10992	13.15
<i>perestanu</i> 'I will no longer'	0%	22%	4	259	10.50
<i>smozhem</i> 'we will be able to'	100%	35%	5	3237	8.01
<i>zaxochetsja</i> 'will want'	25%	22%	4	1132	7.94
<i>nachnet</i> '(s)he will start'	0%	1%	5	3754	7.69
<i>prishlos'</i> 'was forced to'	0%	7%	8	23910	7.30
<i>nel'zja</i> 'be not allowed to'	0%	0%	10	45458	7.26
<i>vynuzhden</i> 'be forced to'	0%	0%	5	4805	7.16
<i>smozhete</i> 'you will be able to'	80%	16%	5	5186	7.00
<i>mogu</i> 'I may'	0%	51%	4	2033	6.93
<i>mozhesh'</i> 'you may'	0%	18%	6	12950	6.43
<i>sumeet</i> '(s)he will be able to'	67%	16%	3	1295	5.52
<i>stanut</i> 'they will start/become'	0%	22%	4	4747	5.47
<i>pora</i> 'be time to'	0%	0%	5	11322	5.35
<i>budem</i> 'we will'	0%	17%	6	19771	5.33
<i>nuzhno</i> 'be needed to'	0%	9%	8	80199	3.53
<i>mozhem</i> 'we can'	100%	32%	4	16905	3.35
<i>mogli</i> 'they could'	20%	26%	5	31342	3.28

<sup>1</sup> For *pora*, which is normally used as a noun, and *nevozmozhno* and *pozдно*, which are usually used as adverbs, the number of auxiliary uses in the corpus was estimated based on a 100-word sample. For *(s)mo{zhet;zhem;zhete;zhesh';gu;gut;zhno;g;gla;gli}* 'can/may', the number of epistemic uses was also estimated in the same way. For *bud-* 'will', the proportion of auxiliary uses without an accompanying modal was estimated likewise.

<sup>2</sup> These counts includes only auxiliary uses of *bud-* 'will' that are not accompanied by a modal, i.e., equivalents to 'will have to' are excluded.

**Table 18. Modal adverbs as collexemes of the apprehensive in Russian**

Modals	Uncertainty (Krause 2008: 108)	Collostruction strength <sup>3</sup>
<i>bezuslovno</i> 'undoubtedly'	1.09	-.18
<i>konečno</i> 'certainly'	1.13	-.84
NULL	1.18	
<i>razumeetsja</i> 'of course'	1.27	-.13
<i>ponjatno</i> 'clearly'	1.33	-.09
<i>nu</i> 'well'	1.67	-1.28
<i>skoree vsego</i> 'most likely'	2.59	5.56
<i>očevidno</i> 'evidently'	2.88	-.13
<i>vidimo</i> 'it appears'	3.11	-.18
<i>po-moemu</i> 'in my opinion'	3.16	-.05
<i>dumaju</i> 'I think'	3.17	-.29
<i>vidno</i> 'it appears'	3.25	-.18
<i>naverno</i> 'probably'	3.57	-.29
<i>kažetsja</i> 'it seems'	3.59	-.42
<i>po-vidimomu</i> 'by all appearances'	3.61	-.04
<i>verojatno</i> 'possibly'	3.63	1.59
<i>možet byt'</i> 'maybe'	3.86	7.18
<i>vozmožno</i> 'possibly'	4.00	1.95

**Table 19. English vs. Russian speakers in the Russian corpus with respect to person, number, animacy of the embedded clause subject and the presence of real fear. Each cell contains English speakers/Russian speakers. Bolded cells show large differences between English and Russian speakers.**

Fear	1st person		2nd person		3rd person			Total	
	Singular	Plural	Singular	Plural	Singular		Plural		
	Animate	Animate	Animate	Animate	Animate	Inanimate	Animate		Inanimate
No	<b>65/34</b>	5/7	25/19	1/1	<b>20/8</b>	<b>30/41</b>	<b>6/18</b>	5/9	<b>157/137</b>
Maybe	15/21	5/9	9/12	2/1	6/5	<b>16/28</b>	4/8	1/4	<b>58/88</b>
Yes	<b>1/13</b>	1/1	3/2	0/4	<b>12/27</b>	<b>6/15</b>	1/6	0/4	<b>24/72</b>
Total	81/68	<b>11/17</b>	37/33	3/6	38/40	<b>52/84</b>	<b>11/32</b>	<b>6/17</b>	<b>239/297</b>

**Table 20. Person and number of subjects of the embedded clauses introduced by the apprehensive**

	<i>I'm afraid</i>		<i>I am afraid</i>		<i>I fear</i>		<i>bojus'</i>		<i>bojus'-ENG</i>		<i>bojus'-FEAR</i>	
1SG	<b>88</b>	<b>35.2%</b>	67	26%	30	12%	52	23.4%	<b>80</b>	<b>36.5%</b>	13	18.1%
1PL	23	9.2%	20	8%	25	10%	16	7.2%	10	4.6%	1	1.4%
2	29	11.6%	23	9.2%	24	9.6%	33	14.8%	39	17.8%	6	8.3%
3	<b>120</b>	<b>48%</b>	<b>140</b>	<b>56%</b>	<b>171</b>	<b>68.4%</b>	<b>121</b>	<b>54.5%</b>	<b>90</b>	<b>41.1%</b>	<b>52</b>	<b>72.2%</b>
Total	250		250		250		222		219		72	

<sup>3</sup> Negated values correspond to repulsion.

**Table 21. English vs. Russian speakers in the Russian corpus with respect to tense of the verb in the embedded clause and person of the subject. Discourse-marker uses only.**

<b>Person</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>Total</b>
Future	23/31	14/23	<b>28/50</b>	<b>68/109</b>
Present	<b>45/21</b>	13/8	44/47	<b>106/84</b>
Past	<b>19/11</b>	<b>11/2</b>	18/16	<b>48/30</b>
Total	<b>87/63</b>	38/33	<b>90/113</b>	215/209

**Table 22. English vs. Russian speakers in the Russian corpus with respect to tense of the verb in the embedded clause and semantic richness (discourse-marker uses only).**

<b>Fear</b>	<b>No</b>	<b>Maybe</b>	<b>Yes</b>	<b>Total</b>
Future	<b>35/54</b>	<b>33/55</b>	<b>11/55</b>	<b>79/164</b>
Present	<b>92/60</b>	14/24	12/11	118/95
Past	<b>36/23</b>	12/7	2/5	48/35
Total	163/137	59/86	25/71	247/294

