Mediating Knowledge through Expressing Surprises: A Frame-based Analysis of Surprise Markers in Research Articles across Disciplines and Research Paradigms

Full name:	Lang Chen
Affiliation:	Nanyang Technological University, Singapore
Email address:	nie15889@e.ntu.edu.sg (institution)
Mailing address:	#11-279, Block 21B, 66 Nanyang Crescent, Singapore
Cellphone No.:	+65 82650798

Full name:	Guangwei Hu
Affiliation:	The Hong Kong Polytechnic University
Email address:	guangwei.hu@polyu.edu.hk (institution)
Cellphone No.:	+852 2766 7564

Mediating Knowledge through Expressing Surprises: A Frame-based Analysis of Surprise Markers in Research Articles across Disciplines and Research Paradigms

Lang Chen

Guangwei Hu

Abstract

Taking a cognitive approach to genre-specific language, this corpus-based study investigated the disciplinary and paradigmatic effect on the use of a specific type of attitude markers—surprise markers—with an analytical framework informed by frame semantics. A Surprise frame was generated and then used to analyze the use of surprise markers in a two-million-word corpus consisting of 320 full-length empirical research articles collected from two social sciences (applied linguistics vs. clinical psychology) cutting across two research paradigms (qualitative vs. quantitative). Results from multiple binary logistic regression analyses show that research paradigm can reliably predict the absence or presence of five categories across four frame elements of Surprise frame. This study not only extends the application of frame semantics to discourse analysis but also has the potential to create a new direction for research in English for Academic Purpose.

Keywords: attitude marker; surprise marker; frame semantics; knowledge emotion; research article

Bio-information of the authors

Lang Chen is a PhD student of applied linguistics in Nanyang Technological University, Singapore. His research interests include corpus linguistics, cognitive semantics, and academic discourse.

Guangwei Hu is Professor of Language and Literacy Education in the Department of English, The Hong Kong Polytechnic University. His research interests include academic discourse, bilingual education, and second language writing. He has published extensively on these and other areas in refereed journals and edited volume. He is currently Associate Editor of *Journal of English for Academic Purposes*.

Introduction

Research articles (henceforth, RAs) nowadays do not merely present scientific findings in a faceless manner but rely on rhetoric methods that enable writers to take authorial stance and engage readers for their persuasiveness (Hyland, 2005b, 2014). The various linguistic resources that fulfill those rhetoric functions are called metadiscourses (Hyland, 2005a). One important metadiscourse whose use in RAs has been extensively researched is attitude markers, which express the writers' affective attitude toward the propositional information (Hyland, 2005a). Previous research investigating the distribution pattern of attitude markers has been along various dimensions, including but not limited to cross-cultural comparison (e.g., Mur-Dueñas, 2010; Mu, Zhang, Ehrich, & Hong, 2015), diachronic change (e.g., Gillaerts & van de Velde, 2010; Hyland & Jiang, 2018), disciplinary difference (e.g., Abdi, 2002; Hyland, 2005b) and paradigmatic effect (e.g., Hu & Cao, 2015). Such studies have revealed to us how various social variables may affect the use of attitude markers in RAs.

However, studies like those generally suffer from two drawbacks. First, they run the risk of obscuring the distribution pattern of a specific type of attitude markers because they treated all attitude markers as a uniform category. Second, those studies overlooked the interaction between attitude markers and other metadiscoursal resources and propositional information because they tend to examine the use of attitude markers independent of other metadiscourses. As a result, even though one found that some disciplines uses more attitude markers than others (Abdi, 2002; Hu & Cao, 2015) or that RAs in some social sciences published recently employed less attitude markers than their counterparts 50 years ago (Hyland & Jiang, 2018), we do not know whether the findings are applicable to all attitudes, and neither do we know whether there is systematic difference in the tendency to hedge or booster a particular attitude between disciplines or across times. Such finer grained and more contextualized information could reveal to us how deep-seated epistemological assumptions affect the choice of rhetoric strategies in RAs because the effectiveness of rhetoric strategies adopted in academic writing hinges inevitably on the consensus of a discourse community on what constitutes knowledge and its legitimate justification. Therefore, for a deeper understanding of academic writing both as a knowledge-making practice and as a persuasive endeavor, it is necessary to focus on one specific attitude marker and examine its potential interaction with other metadiscourses and other contextual information.

The present paper reports on a corpus-based study on the use of one specific type of attitude markers, surprise markers, in two social sciences (applied linguistics vs. clinical psychology) cutting across two research paradigms (qualitative studies vs. quantitative studies). In the present study, surprise markers are defined as linguistic items that express the authors' or other people's attitude towards the unexpectedness of scientific facts. Detailed and contextualized analyses of surprise markers are merited by at least two discoveries from previous research. First, surprise is one of the typical attitudes expressed in academic contexts (Hyland & Jiang, 2017) and thus worth finer-grained investigation. Second, it has been found that surprise is typically expressed in a staged manner in RAs, forming "surprise routines" that connect surprise makers to other

contextual information (Tutin, 2015). This shows the feasibility to study the use of surprise markers by taking other textual elements into consideration. Our interest in the use of surprise markers between qualitative studies and quantitative studies comes from the potential difference between these two research paradigms "in the attitudinal parameters of expectability" (Hu & Cao, 2015, p. 18) as a result of their distinctive way of making knowledge. Two social sciences were chosen precisely because of their embrace of both qualitative and quantitative research so that enough data could be collected for both the qualitative and quantitative subcorpora. To analyze the use of surprise markers with finer granularity and in a more contextualized manner, we generated an analytical framework informed by frame semantics because this linguistic theory enables us to characterize the use of surprise markers by treating all the relevant contextual information as frame elements (FEs). After all surprise markers identified in our corpus were carefully annotated with the analytical framework, a series of statistical analyses were conducted, which yielded informative results of both methodological and theoretical value.

Related research

Narrower focus: From attitude markers to surprise markers

Just as Hyland (2005b) rightly pointed out, there is a growing recognition that academic writing is more than an objective representation of scientific facts but "a persuasive endeavour involving interaction between writers and readers" (p. 173). According to Hyland (2005a), this interaction is achieved mainly through employing two types of metadiscursive linguistic resources, namely interactive metadiscourses and interactional discourses. Interactive metadiscourses are information-oriented as it "helps readers understand a text by explaining, orienting and guiding them through the information" (Hyland 2005a, p. 75) whereas interactional metadiscourses, as the label suggest, are interaction-oriented, consisting of a range of discursive features such as hedges, boosters, self-mentions and attitude markers to show the stance of the writer and to engage with the readers (Hyland, 2005b). In the past decade, this theory has become one of the dominant analytical frameworks in academic writing research. Many studies based on this theory have indeed identified significant effect of various social variables on the distribution of those metadiscourses. For example, Abdi (2002) found that social sciences employ significantly more such interactional metadiscourses as hedges and attitude markers than natural sciences, Mur-Dueñas (2011) observed that RAs written in English contain significantly more metadiscourses than those written in Spanish, and recently Hyland and Jiang (2018) noticed a trend to use more interactive metadiscourses but less interactional metadiscourses in RAs published in a 50-year span. Findings from studies like those revealed to us how the genre of RAs is "shaped by complex social interactions in academic discourse community" (Hu & Cao, 2015, p. 12) and, therefore, enhanced our understanding of genres as textual response to social situations (Hyland, 2008; Miller, 1984).

However, research along this line generally has two downsides. First, except in few studies (e.g., Cao & Hu, 2014), each metadiscursive feature is treated as a uniformly functional category and not enough attention was paid to the semantic difference between its subcategories. For instance, when we read Khedri, Chan and Ebrahimi's (2013) finding that RA abstracts in Applied Linguistics use less transitions than those in Economics, we do not know whether this applies to all its subcategories, namely addition, comparison and consequence. Second, interaction between different types of metadiscourses tends to be overlooked. For example, when Hyland and Jiang (2016) reported a fall in the use of both hedges and boosters in applied linguistics and sociology in the past 50 years, they did not tell us whether there was a systematic pattern in the attitudes or propositions that were hedged or boosted.

When it comes to attitude markers, these two problems look more prominent. For one thing, attitude markers are intrinsically heterogeneous in meaning as they may cover a range of attitudes such as "surprise, agreement, importance, frustration, and so on" (Hyland, 2005b, p. 180). It would be simplistic to assume that a social factor affects the distribution of different attitude makers in a uniform manner. For another, attitude markers also have inherent association with many other metadiscursive features. The most obvious connection is with hedges and boosters because "a defining property of all attitudinal meanings is their gradability" (Martin & White, 2005, p. 135). Furthermore, attitude markers would in many cases be expressed with self-mentions as they indicate the writer's attitude (Hyland, 2005b). For the same reason, they are also related to the propositions that they are expressed towards. Therefore, it is safe to say that ignoring the interaction between attitude markers and other metadiscourses and propositions may result in incorrect or at best incomplete interpretation of the effect of various social factors on the former's distribution.

Though rarely has a researcher paid exclusive attention to attitude makers except for Mur-Dueñas (2010), which found no significant difference in the use of attitude markers between international American and local Spanish in business management RAs, this metadiscursive feature have been extensively researched in study that analyze metadiscourses. However, regrettably, practically no study so far has taken these two points mentioned above into consideration when analyzing the use of attitude markers in RAs. The consequence is that some more subtle patterns may be obscured by the pattern observed when attitude markers are analyzed as a homogeneous category. For example, while both Abdi (2002) and Hyland (2005b) have observed that social sciences (SS) use more attitude markers than natural sciences (NS), we do not know whether this holds true for all attitudes. It is possible that NS convey certain attitudes, for instance, "importance", more frequently than SS. Consequently, the inference that "SS writers feel free to express their own feelings, while NS writers prefer to sidestep such emotions in order to preserve the scientific nature of their information" (Abdi, 2002, p. 143) has to be received with caution. In the same vein, without knowing the change pattern of each attitude, it would be difficult, if not impossible, to correctly interpret the finding that frequency of attitude markers drops in SS but rises in NS in the past 50 years (Hyland & Jiang, 2018).

Similar thing could be said about the pattern found within discursive soft disciplines. For instance, though Hu and Cao (2015) found no significant difference in

the frequency of attitude markers among the three soft disciplines they investigated, one might still wonder whether one discipline is more likely to boost or hedge expressed attitudes than the others (e.g., *it is somehow surprising* vs. *it is surprising*) or whether there is qualitative difference between the propositions that the same attitude is expressed towards in different disciplines. By the same token, even when the discipline factor is controlled, the culture preference indicated in Abdollahzadeh's (2011) observation that Anglo-American writers tend to use more attitude markers than their Iranian counterparts in English-mediated applied linguistics RAs does not tell us the complete story.

Therefore, to avoid those afore-mentioned two problems, it is necessary to carry out detailed studies on attitude markers that fall within the same semantic field. This, in turn, means that the special semantic property of a specific attitude markers must be taken into consideration. Along this line of thinking, one would find the definition of attitude markers in Hyland (2005b) questionable because it explicitly excludes the epistemic aspect of attitude markers by stating that they "indicate the writer's affective, rather than epistemic, attitude to propositions" (p. 180). However, previous research in psychology (e.g., Foster & Keane, 2015; Teigen & Keren, 2003) and cognitive linguistics (e.g., Goddard, 2015; Kövecses, 2015) have shown that some attitudes are more epistemic than affective. One of these attitudes is surprise.

According to Silvia (2009), surprise belongs to a family of emotions called knowledge emotions because, like interest and confusion (the other two knowledge emotions), surprise is closely "associated with thinking and comprehension" (pp. 48-49). In other words, surprise is related to knowledge and knowledge-making. Considering this, linguistic expressions of surprise may be not only affectively driven but also epistemically motivated. This explains why surprise markers are not uncommon while all other basic emotions are rarely expressed in RAs (Tutin, 2015). This is also in line with Hyland and Jiang's (2017) observation that both surprise and interest are typical attitude markers employed in academic writing.

The explicit exclusion of the epistemic aspect of attitude markers may have also resulted in a lack of attention to factors that are more epistemologically related than socially relevant, for example, research paradigm. In social sciences, qualitative and quantitative research constitute two major competing and complementing paradigms due to their fundamentally different ontological and epistemological assumptions (Cohen, Manion, & Morrison, 2011; Guba, 1990). Hu and Cao (2015) is the only study available to us that examined the possible effect of research paradigms on the use of attitude markers. They found that quantitative RAs use significantly more attitude markers than qualitative RAs. The authors attributed this difference to the different epistemological assumptions between these two research paradigms and even hinted that the difference in the use of surprise markers may be a major contributor to the difference in attitude markers. However, they did not explore this further.

Partly motivated by Hu and Cao (2015), the present study aims to investigate how two epistemologically related factors, discipline and research paradigm, affect the use of surprise markers, linguistic expression of a knowledge emotion. To do this, we need an analytical framework that revolves around surprise markers and connects them to related

semantic elements. For such an analytical framework, we turn to the theory of frame semantics.

Broader context: Frame semantics and surprise markers

Frame semantics is a theory of cognitive semantics developed by Fillmore (1976, 1977, 1982). Recognizing the cognitive tendency of human beings towards "structured ways of interpreting experience" (Fillmore, 1976, p. 20), this theory proposes the idea that a given linguistic item can only be fully understood against a knowledge/cognitive structure that specifies "characteristic features, attributes, and functions of a denotatum, and its characteristic interactions with things necessarily or typically associated with it" (Allen, 2016, p. 55). Such knowledge structures are called semantic frames and frame semantics studies "how linguistic forms *evoke* [emphasis in original] or activate frame knowledge, and how the frames thus activated can be integrated into an understanding of the passages that contain these forms" (Fillmore & Baker, 2010, p. 317). The linguistic forms that evoke frames are frame-evoking lexical units (LUs) and other frame-related semantic components are termed frame elements (FEs). For example, according to Fillmore (2006), to understand verbs like *buy*, *sell*, *spend* and *cost*, one needs to have all adequate knowledge of a semantic frame called Commercial_transaction, which contains such FEs as Buyer, Seller, Goods, and Money. Note that these FEs should be understood as abstract concepts that exist in our "folk theory" understanding of a particular event, not scientific knowledge. In other words, one does not need accurate knowledge of every types of goods to have access to the concept of Goods. However, when focus is on a single semantic frame, further classification of FEs and analyses of their distribution may result in revealing findings. Previous research based on frame semantics has not explored this finer-grained territory, but it is precisely what we are going to do in the present study.

According to annotations in FrameNet[i], there are core FEs and non-core FEs. A core FE "instantiates a conceptually necessary component of a frame" and thus makes "the frame unique and different from other frames" (Ruppenhofer, Ellsworth, Petruck, Johnson & Scheffczyk, 2016, p. 19) while a non-core FE does "not uniquely characterize a frame, and can be instantiated in any semantically appropriate frame" (Ruppenhofer et al., 2016, p. 20). Let's take an annotated sentence in FrameNet (henceforth FN) as an example, in which the typical surprise marker *surprising* evokes the Frame of Stimulus_focus:

1) [Stimulus The upbeat account of the firm's local fortunes] may be regarded as [Degree somewhat] SURPRISING[Target] [Circumstances given the fact that IBM Eastern Europe has not announced the conclusion of any major contracts so far this year]. [ii]

In this example, according to the annotating convention of FN, "the upbeat account of the firm's local fortunes" is the Stimulus, which indicates what triggers surprise; "somewhat" tells us the Degree of surprise; and "given the fact … this year" gives the Circumstances under which the Stimulus is considered as SURPRISING. For this frame, FN only considers the Stimulus as the core FE while annotates all others as non-core FEs. However, while this classification may be justified for Stimulus_focus, it

seems questionable for surprise. According to studies of surprise as an emotion, violation of expectation is a key feature of surprise (Casti, 1994; Teigen & Keren, 2003). Therefore, a presupposed expectation should be a necessary component in any frame where surprise is involved. This means, if there is a frame dedicated to surprise, Circumstances should be a core FE and it may deserve another more transparent name. It is perhaps because FN does not pay specific attention to surprise, there is inconsistency of annotation in the examples for *surprising*. For example, in the following two annotated sentences from FN:

2) Since all body cells contain the same genes, it might seem SURPRISING^{Target} [$_{Stimulus}$that all body cells aren't the same as each other].

3) Given that there are so few women in powerful positions [$_{Stimulus}$this] may seem SURPRISING^{Target}.

both subclauses provide explanations for why the Stimulus is surprising because they both give the source of expectation. "Since all body cells contain the same genes", one would expect that all body cells are the same as each other, but this expectation is violated. Similarly, given "than there are so few women in powerful positions", we would expect a certain scenario, but the contrary happens. As you can see, these two subclauses are conceptually identical in terms of their contribution to a complete understanding of surprise to the "given the fact …" part in Example 1. However, while "given the fact …" in Example 1 was annotated as Circumstances, these two subclauses in Example 2 and 3 were not annotated at all, which means they were not considered as an FE of Stimulus_focus, not even a non-core FE.

Further analyses of the original FN show that inconsistency of annotation does not only exist among the example sentences for one LU, but also exist at the frame level. In FN, *surprise* and *surprised* can evoke two semantic frames: Stimulate_emotion (which could be evoked by *surprise* as a noun and as a verb) and Just_found_out (which could be evoked by *surprise* as a noun and *surprised* as an adjective). Both these two semantic frames contain an FE called Explanation. For Stimulate_emotion, the definition of Explanation is the "reason why the Stimulus causes the emotion in the Experiencer", and in Just_found_out, the definition of Explanation is "the explanation for why the Stimulus evokes a certain emotional response". Clearly, as discussed above, for surprise, these two definitions should point to the same thing—the source of expectation. It is also obvious that, as far as surprise is concerned, the Explanation in Stimulate_emotion and Just_found_out overlaps conceptually with Circumstances in Stimulus_focus.

This conceptual overlapping between FEs across different semantic frames that could be evoked by surprise markers makes it possible to generate a Surprise frame, which could characterize how human being express this feeling linguistically. This frame, in turn, could serve as an analytical framework for detailed analyses of surprise markers in RAs. As a detailed study requires us to look at how surprise markers are used, our analyses should not stop at the FE level but should further categorize each FE according to its semantic property. Taking all these into consideration, the present study set out to answer the following questions:

1. Is it possible to generate a Surprise frame?

2. If it is, what are the FEs for this frame and what categories are there under each frame?

3. Are there differences in the use of surprise markers between RAs in applied linguistics and those in clinical psychology in terms of the distribution of each FE category?

4. Are there differences in the use of surprise markers between RAs adopting qualitative paradigm and those adopting quantitative paradigm in terms of the distribution of each FE category?

Methods

The corpus

To answer the research questions listed above, we compiled a two-million-word corpus, consisting of 320 full-length RAs randomly collected from four prestigious journals in each discipline. The journals are *Applied Linguistics*, *Modern Language Journal*, *TESOL Quarterly* and *Language Learning* for applied linguistics, and *Journal of Counseling Psychology*, *Journal of Clinical and Consulting Psychology*, *Journal of Abnormal Psychology* and *Professional Psychology: Research and Practice* for clinical psychology. Those journals were chosen by taking into consideration both their Impact Factor and expert advice. For our corpus to be representative of the two disciplines and the two research paradigms, we adopted stratified random sampling. All empirical RAs in those eight journals that are published during the same period were first coded with research paradigm according to the types of data they collected and the methods they adopted to analyze their data. After excluding RAs that adopted mixed method, we then had four "pools" of RAs, which are qualitative RAs from applied linguistics, quantitative RAs from applied linguistics, qualitative RAs from clinical psychology, and quantitative RAs from clinical psychology. Then, 80 RAs were randomly selected from each pool to form our corpus. Each article was labelled in the format of "Discipline-Paradigm-number". For example, "AL-Qual-20" refers to the 20th article in applied linguistics that reported on a qualitative study, and "CP-Quan-14" means an article in clinical psychology that presented a quantitative study. Table 1 gives descriptive statistics on the corpus.

Table 1 Descriptive information on the corpus

| Measure | Applied Linguistics | | Counseling Psychology | | |
	Quantitative	Qualitative	Quantitative	Qualitative	Total
No. of RAs	80	80	80	80	320
Total words	464,769	524,415	427,746	443,372	1,860,302
Average words/RA	5,810	6,555	5,347	5,542	5,813

The surprise markers

To identify all possible surprise markers in our corpus, we compiled a list of headwords that are either synonyms or antonyms of *surprise* and all its derivative forms. This list was informed by Roget's International Thesaurus (Kipfer, 2010) and WordNet (Fellbaum, 1998). We considered both the synonyms and antonyms of *surprise* and all its derivative forms because many phrases that contained their antonyms could also express surprise, for example, *contrary to our expectation* and *against expectation*. See Appendix A for a complete list of the headwords.

The Surprise frame, its FEs and FE sub-categories

This section will answer the first two research questions. After searching the corpus with all the headwords, we identified 83 surprise markers, which occurred 439 times in our corpus. According to annotation in FN, these 83 surprise markers could evoke seven frames, which are Expectation, Stimulus_focus, Typicality, Just_found_out, Emotion_directed, Desirability, and Stimulate_emotion. The next step was to determine whether we could collapse these seven frames into one single frame for the expression of surprise. This is possible if we could find conceptual overlapping between FEs when the frames are evoked by surprise markers. Further analyses confirmed our expectation.

First, the four frames, Stimulus_focus, Just_found_out, Emotion_directed and Stimulate_emotion, are conceptually connected as manifested by their shared FE of Stimulus. In fact, these four frames were interconnected by various frame-frame relations as defined in FN (Ruppenhofer et al., 2016). We are not going into details of this aspect of FN because it is not relevant to our current pursuit. As mentioned earlier, they also share the FE of Explanation although in Stimulus_focus it is termed Circumstances. As for the FE of Experiencer, though it is not considered as a relevant FE in Stimulus_focus, it is a necessary component of a conceptual understanding of surprise as any emotion must be experienced by someone. Furthermore, the presence or absence of an experiencer could have the potential to distinguish one discipline from another, or one paradigm from the other.

Second, although there is no Stimulus and Experiencer in Expectation frame (which could be evoked by surprise markers like *unexpected*), we found Phenomenon and Cognizer in this frame perfectly correspond to Stimulus and Experiencer in the four frames discussed above respectively when surprise is concerned. An obvious piece of evidence is a sentence from our corpus in which *surprising* and *unexpected* are juxtaposed as being interchangeable:

4) Although the results of the current study support our predictions regarding response modulation and primary psychopathy, we recognize that aspects of these findings may seem **surprising** or **unexpected**. (CP-Quan-62)

In this example, if we follow the convention of FN, *surprising* would evoke Stimulus_focus frame and *unexpected* would evoke Expectation frame. Correspondingly,

aspects of these findings would be both Stimulus and Phenomenon at the same time. However, *surprising* and *unexpected* are obviously meant to be synonyms in this example and thus they should evoke the same frame. The only way out of this dilemma is to formulate a frame that can be evoked by both words. In fact, the juxtaposition of *surprising* and *unexpected* can also be found in Corpus of Contemporary American English (COCA) (Davies, 2017). Not surprisingly, *unusual* and *surprising* are also sometimes used together as synonyms in the same sentence. Therefore, it is also justifiable to consider *unusual* as an LU that evokes the same frame as *surprising* does, though in FN the former evokes Typicality frame.

Third, in our corpus, Desirability frame was evoked by such words as *amazingly* and *remarkable*. Note that *remarkable* is not listed as an LU for any frame in FN. Here we consider it as a surprise marker that evokes Desirability because, for one thing, being surprising is part of its definition, and for another, it describes a desirable quality as it means surprisingly good, which conforms to the definition of Desirability in FN. Although these words add an evaluative quality to being surprising, the being surprising part is still a core component of their semantic property. Therefore, it makes sense to treat them as LUs that evoke a Surprise frame. In this case, the Evaluee in Desirability corresponds to Stimulus and Phenomenon in other frames.

Finally, according to the surprise routine described in Tutin (2015), resolution, which refers to how a surprising finding could be resolved, is an integral part of expression of surprise in RAs. This means Resolution should be a necessary component for expressing surprise, at least in the genre of RAs. However, none of the seven frames discussed above contains this FE.

According to the analyses above, it is not only feasible but also necessary to formulate a Surprise frame, of which the most frequently occurring FEs could constitute the basis of our analytical framework for surprise markers in RAs. At this point, we could list at least five FEs for our Surprise frame: Experiencer, which tells who is surprised; Trigger, which combines Evaluee from Desirability, Phenomenon from Expectation and Stimulus from the other four frames, and was so named to avoid confusion; Source_of_expectation, which combines Circumstances from Stimulus and Explanation from other frames, and was so named to reflect the special property of surprise, i.e., the violation of expectation (Casti, 1994); Degree, which describes how surprising a Trigger is; and Resolution, which shows how a surprising Trigger is resolved.

With this understanding, we then went on to code all the sentences where surprise markers occur. It is worth noting that when coding Resolution, we necessarily go beyond the intra-sentential level and expand our scope to the entire article because Resolutions of expressed surprises sometimes may not appear in the same sentence where surprise markers occur. For example, there are cases where surprises are expressed in the result part, but their resolutions appear in the discussion part. We consider this discourse level annotation of FEs as a beneficial extension to the annotation convention of FN, especially when applying frame semantics to detailed discourse analyses. When other FEs than those five were identified in the context, we coded them with those FEs defined in the original FN.

As expected, the five FEs we defined accounted for 94.46% (784 out of 830) of all the FE instantiations while all other possible FEs occur less than five times (see Table 2 for the distribution of those five FEs). Therefore, they could adequately characterize the use of surprise markers in the corpus.

Table 2 Distribution of the most frequent five FEs of Surprise frame

FEs	Raw frequency	% of frame instances[a]
Trigger	439	100.00
Resolution	117	26.65
Source_of_expectation	104	23.69
Degree	77	17.54
Experiencer	56	12.76

[a] Percentage of all instances of the five identified FEs

For a finer-grained analysis, we then went on to classify each FE with subcategories that could adequately characterize how surprise markers are used in RAs. This was done by manual and iterative coding of each surprise markers. The result is an analytical framework with five FEs and 20 categories as shown in Figure 1. To test the robustness of this analytical framework, a graduate student of applied linguistics was invited to independently code 10% of the sentences after three hours of training. The interrater reliability as measured by Cohen's kappa reached substantial agreement (Landis & Koch, 1977) for all the FEs. Table 3 lists the Cohen's kappa values for each FE ($p < .05$). With this framework, we coded all the surprise markers identified in the corpus. Below we give a brief explanation of the five FEs and their sub-categories in our analytical framework. Relevant examples will be provided in the Results section.

Table 3 Levels of interrater reliability

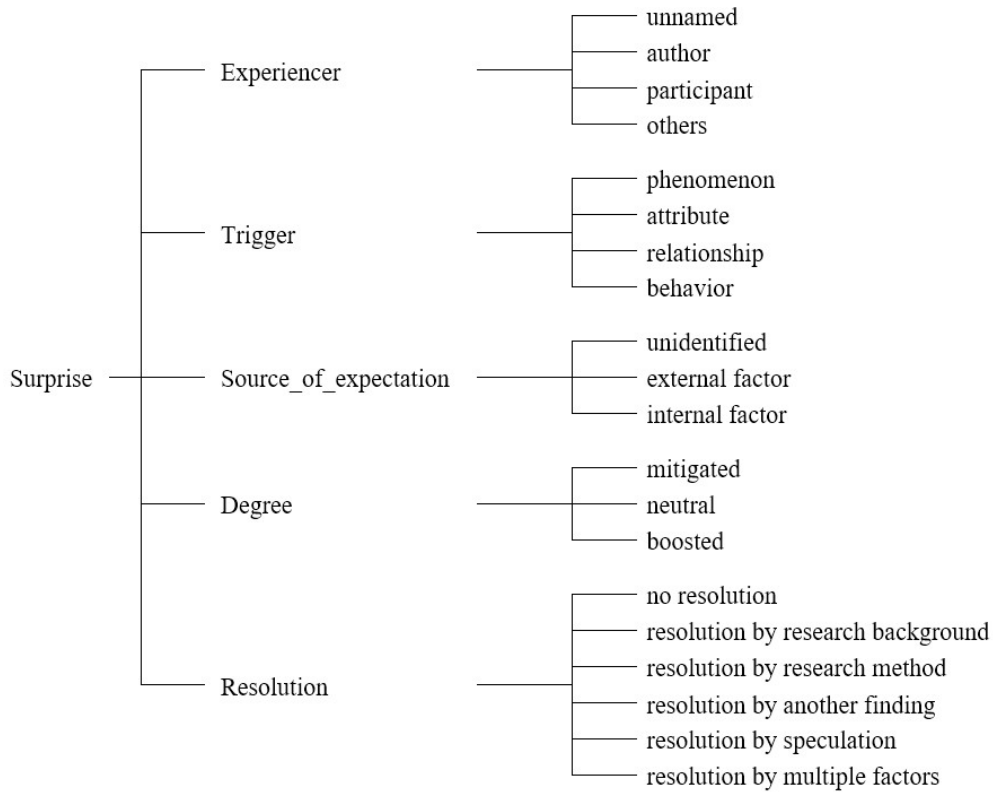| Dimension	Cohen's Kappa
Experiencer	0.702
Trigger	0.652
Source_of_expectation	0.645
Degree	0.852
Resolution	0.714

Figure 1 The Surprise Frame

Experiencer refers to who feels surprised. As pointed out by Tutin (2015), most surprise markers in academic writings are source oriented, meaning there is no explicit mention of the experiencer. This is also true of our data. Among those cases where Experiencers were given, three categories were identified. They are "author", "participant", and "others" (meaning experiencers other than the author and participants). In the present study, experiencers expressed by "one" and inclusive "we" were included in the category of "author".

Trigger concerns what is surprising, which are classified into four categories according to their semantic property: "relationship", which refers to the relationship between research variables, between findings in the same study or between findings in different studies; "attribute", which includes the characteristic of a participant, research method or any other research object depicted in RAs; "behavior", which describes what people do (physical behavior), say (verbal behavior) or think (mental behavior); and "phenomenon", which deals with any events that do not belong to the first three categories.

While a Trigger tells us what is surprising, Source_of_expectation reveals why something is surprising. In other words, what makes one feel surprised is not the trigger itself, but the conceptual incongruence that results from the gap between one's expectation and what actually happens (Casti, 1994). When the expectation was based on findings of previous studies or the research contexts, the source of expectation is "external factor" for the simple reason that this information is known prior to the study.

Contrarily, if the expectation was based on things within the research at hand, for example, other findings in the same study or attributes of research objects, the source of expectation is "internal factor" because this information is only available after the study begins. Although surprise is inevitably the result of conceptual incongruence, the authors may choose not to present the source of expectation explicitly. Considering this, an additional category, "unidentified" is added. To sum up, Source_of_Expectation consists of three categories, namely "unidentified", "external factor" and "internal factor".

Degree reflects the intensity of surprisingness. It has three categories: "mitigated", "neutral" and "boosted". As their name suggests, mitigated surprises are those hedged in some ways (e.g., *it is somewhat surprising*) while boosted surprise are those intensified in some ways (e.g., *it is very surprising*). When a surprise marker is not modified (e.g., *it is surprising*), the expression is considered as a case of "neutral" surprise.

Finally, Resolution deals with how authors explain the surprisingness, thereby resolving it. In our data, it was found that researchers explained or resolved surprising scientific facts mainly in five ways. First, a surprising finding in a study may be found consistent with findings of some previous studies or may not be so surprising if the situational characteristics had been taken into consideration. Because the previous research and current situation precede the reported study and form its background, this category is called "resolved by research background". Second, sometimes the surprisingness of a finding could be attributed to the instruments used in the study, hence the category "resolved by research method". Third, there are also cases in which a surprising finding was explained by another finding in the same study, giving rise to the category of "resolved by another finding". Fourth, when no resolution could be found from the above three sources, the researchers would offer their tentative explanation or put forward a new hypothesis. This constitutes the category of "resolved by speculation". This category is probably the most important one in the sense that resolution by speculation is none other than new knowledge. Thus, we expect most surprises to be resolved in this way in a knowledge-making genre if they are resolved at all. Fifth, occasionally, more than one explanation would be given to resolve the surprisingness. In such a case, the surprisingness was considered as "resolved by multiple factors". It was found that not every surprise expressed in an RA was resolved. Therefore, a category called "no resolution" was added.

The statistical analyses

We conducted binary logistic regression analyses on every category to see whether our two predictor variables (i.e., discipline and research paradigm) can reliably predict the presence or absence of those categories. There were only two values for the dependent variables, "absence" if a category did not appear in an RA and "presence" if it appeared at least once. This choice of statistical method was based on two characteristics of our data. One, the distribution of surprise markers is somewhat dispersing. In fact, those 439 occurrences only appeared in 187 RAs, while the other 133 texts did not contain any surprise makers as defined in this study. No systematic differences were observed between those RAs that contained surprise markers and those that did not. Second, multiple occurrences of surprise markers sometimes pointed to the same trigger. For all

the analyses, the value of "absence" was set as the reference value. Bonferroni adjustment was applied when determining the alpha value, which was set at 0.0025 because the same statistical methods were conducted 20 times. The cutting point for classification rate was set as 0.5. For the predictor variables in all the analyses, applied linguistics was set as the reference variable for the discipline and the qualitative paradigm was set as the reference for the research paradigm.

Results

The binary logistic regression analysis for the general distribution of surprise markers in the corpus did not return a significant chi-square value (χ^2 (2, n = 320) = 2.973, p = .226), indicating that the independent variables as a group could not reliably predict the absence or presence of surprise markers in each RA. In other words, there was no significant difference between the two disciplines and the two research paradigms in terms of using or not using surprise markers in the RAs. However, binary logistic regression analyses on the FE categories reveal significant paradigmatic differences. The following subsections report the statistical results by FEs.

Experiencer

Consistent with the finding in Tutin (2015), among all the surprises expressed in our corpus, only a small fraction included experiencers and most of them were source oriented (see Figure 2). This indicates that academic writing in general prefers an objective tone by reducing human presence (see example 5 and 6). All the examples below are from our corpus, and in the examples the surprise markers are in boldface and the target FEs are underlined. In the cases where experiencers were explicitly given, "participant" accounted for the largest proportion (see examples 7 and 8), suggesting a scholarly interest in the attitudes of research subjects in social sciences. Following "participant" is the category of "author", which is realized in various ways (see example 9 and 10 for illustration). There are also some rare cases in which the surprise of other people, for instance, other researchers, was described as shown in example 11.
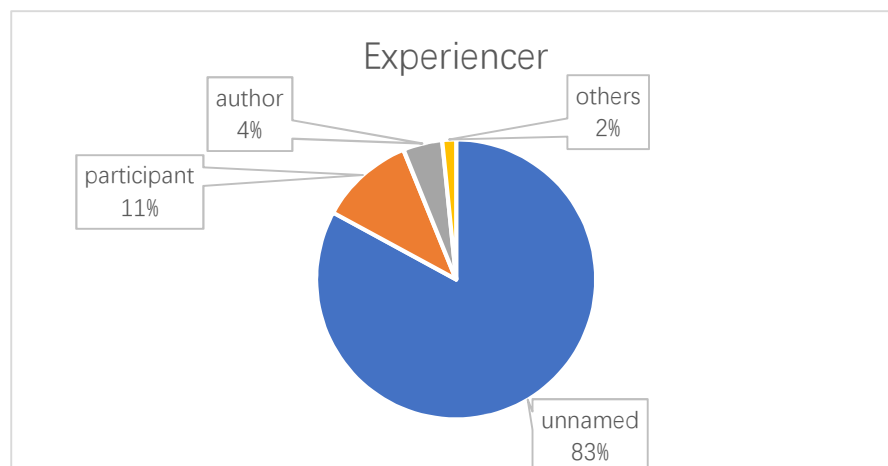

Figure 2 Distribution of Experiencer

5) Furthermore, recent studies have found **surprising** results indicating that L1 reading skills have less impact on L2 reading performance than expected. (AL-Quan-68)

6) It is **strange** that in the OED, influenced in its notational system by the work of Ellis, no attempt is made to define a standard. (AL-Qual-04)

7) <u>His practice colleagues</u> were **shocked** when they received word that Charles had died of pancreatic cancer. (CP-Qual-53)

8) <u>One of the plant managers who was in attendance</u> noted his **surprise** that the individual in question would stand up and speak in front of a group. (AL-Quan-46)

9) What pleased and **surprised** <u>me</u>, though, was that the write-before-you-read exercise seemed to result in more than simple comprehension. (AL-Qual-40)

10) As a final factor predicting L2 reading attitudes, <u>we</u> were somewhat **surprised** to find that L1 reading frequency was not a significant contributor to any L2 reading attitudes. (AL-Quan-73)

11) <u>Schlesinger</u> (1968) was **surprised** to learn that his subjects performed about as well on sentences that were structurally complex as on those that were less complex. (AL-Qual-06)

Statistical analysis shows that the two predictor factors can only reliably predict the absence or presence of "participant" (χ^2 (2, $N = 320$) = 22.631, $p < .001$). Odds Ratio (OR) value indicates the biggest difference between the two research paradigms: qualitative RAs are 13.33 times, which was calculated by dividing 1 with the Odds Ratio 0.075, more likely to describe the surprises of research participants. Table 4 gives the detailed statistics for Experiencer.

Table 4 Results of binary logistic regression analyses for Experiencer

Dependent variable	Independent variable	B	SE	$Wald$	p	Odds ratio	95% CI for OR	
							Lower	Upper
unnamed	Discipline	0.076	0.225	0.113	0.736	1.079	0.694	1.675
	Paradigm	-0.025	0.225	0.013	0.911	0.975	0.628	1.514
	Constant	0.163	0.194	0.702	0.402	1.177		
	$R^2 < .001$ (Cox & Snell); $R^2 = .001$ (Nagelkerke); Model $\chi^2(2) = .126$, $p = .939$							
author	Discipline	-1.020	0.541	3.549	0.060	0.361	0.125	1.042
	Paradigm	-1.020	0.541	3.549	0.060	0.361	0.125	1.042
	Constant	-2.022	0.333	36.851	0.000	0.132		
	$R^2 = .024$ (Cox & Snell); $R^2 = .069$ (Nagelkerke); Model $\chi^2(2) = 7.835$, $p = .020$							
participant	Discipline	0.278	0.432	0.414	0.520	1.321	0.566	3.082
	Paradigm	-2.587	0.747	12.011	0.001	0.075	0.017	0.325
	Constant	-1.930	0.328	34.716	0.000	0.145		
	$R^2 = .068$ (Cox & Snell); $R^2 = .162$ (Nagelkerke); Model $\chi^2(2) = 22.631$, $p < .001$							
others	Discipline	-0.712	0.877	0.659	0.417	0.491	0.088	2.736
	Paradigm	-1.638	1.102	2.209	0.137	0.194	0.022	1.685
	Constant	-3.136	0.543	33.304	0.000	0.043		
	$R^2 = .011$ (Cox & Snell); $R^2 = .067$ (Nagelkerke); Model $\chi^2(2) = 3.660$, $p = .160$							

Trigger

Across the corpus, the four types of triggers had a roughly balanced distribution (see Figure 3). Overall, observed "phenomenon" (see example 12) as a trigger occurred most frequently in the RAs followed by "attribute" trigger (see example 13). A close

examination revealed that "relationship" surprises usually occurred in the presentation of findings involving two or more research variables (see example 14), which had a higher visibility in quantitative RAs. On the other hand, "behavior" surprises were most often triggered by the behaviors of participants (see example 15), which were more frequent in qualitative RAs.
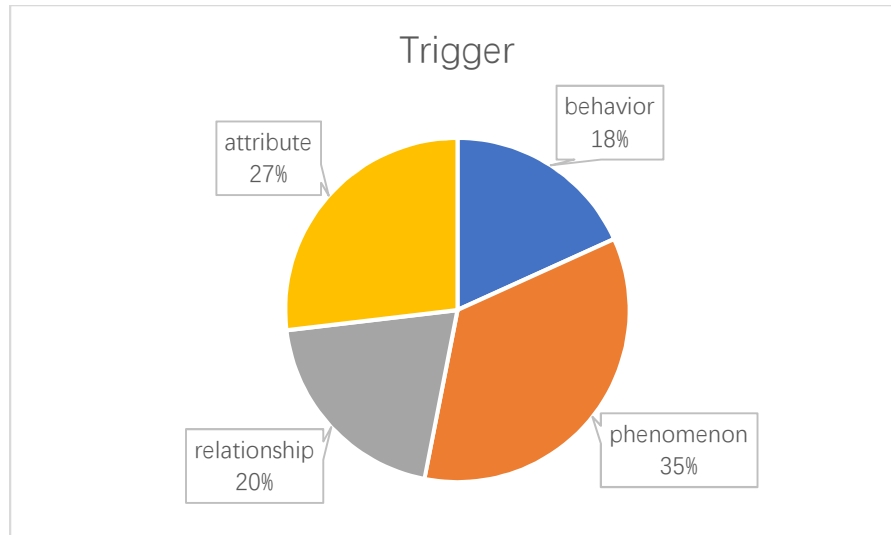

Figure 3 Distribution of Triggers

12) We should notice, however, that what is **unexpected** here is not so much the fact that the choice of CS is not constrained in the orientation, but rather that lexical borrowing is the main strategy employed by the subject in the episodic component. (AL02)

13) Perception of speech sound contrasts by infants is **remarkably good** before any productive language emerges and seems relatively independent of the infant's linguistic environment (Eimas et al. 1971; Kuhl 1978), while perception of phoneme contrasts by listeners who have mastered a particular language is constrained by that language. (AT24)

14) **Unexpectedly**, social-relevant behavior related negatively and significantly to solitary relevant behavior (r = -.52). (PT25)

15) What is **remarkable** about the teacher's three contrasting gestures is that they are almost a replica of one another despite the fact that they have been formulated across considerable stretches of talk. (AL58)

Tests for the full model (with both discipline and paradigm in the model) versus the constant-only model were statistically significant only for "relationship" and "behavior", and for both categories, only research paradigm was the reliable predictor (see Table 5 for the statistic results).

Table 5 Results of binary logistic regression analyses for Trigger

Dependent variable	Independent variable	B	SE	Wald	p	Odds ratio	95% CI for OR Lower	Upper
phenomenon	Discipline	-0.118	0.243	0.236	0.627	0.889	0.552	1.431
	Paradigm	-0.813	0.246	10.963	0.001	0.443	0.274	0.718
	Constant	-0.295	0.201	2.158	0.142	0.745		
	R^2 = .035 (Cox & Snell); R^2 = .049 (Nagelkerke); Model $\chi^2(2)$ = 11.508, p = .003							
attribute	Discipline	0.108	0.268	0.161	0.688	1.114	0.659	1.883
	Paradigm	-0.822	0.275	8.973	0.003	0.439	0.257	0.753
	Constant	-0.872	0.219	15.796	0.000	0.418		
	R^2 = .029 (Cox & Snell); R^2 = .044 (Nagelkerke); Model $\chi^2(2)$ = 9.492, p = .009							
relationship	Discipline	-0.503	0.306	2.707	0.100	0.605	0.332	1.101
	Paradigm	1.355	0.333	16.531	0.000	3.876	2.017	7.448
	Constant	-2.119	0.306	48.067	0.000	0.120		
	R^2 = .065 (Cox & Snell); R^2 = .106 (Nagelkerke); Model $\chi^2(2)$ = 21.423, p < .001							
behavior	Discipline	-0.824	0.330	6.229	0.013	0.439	0.230	0.838
	Paradigm	-1.854	0.390	22.565	0.000	0.157	0.073	0.337
	Constant	-0.628	0.225	7.801	0.005	0.533		
	R^2 = .104 (Cox & Snell); R^2 = .176 (Nagelkerke); Model $\chi^2(2)$ = 34.996, p < .001							

Source_of_expectation

In most cases, researchers chose not to explicitly give the incongruence source as can be seen in example 16. When the sources of expectation were given, three types prevailed. For external factors, the expectations were usually based on previous research (example 17) or current situation (example 18) and for internal factors, it was other findings (example 19 and 20) that caused researchers to have a certain expectation. The most

frequent phrases used to introduce the source of an expectation were "given that" and "considering that". Figure 4 gives the distribution of the Explanations.
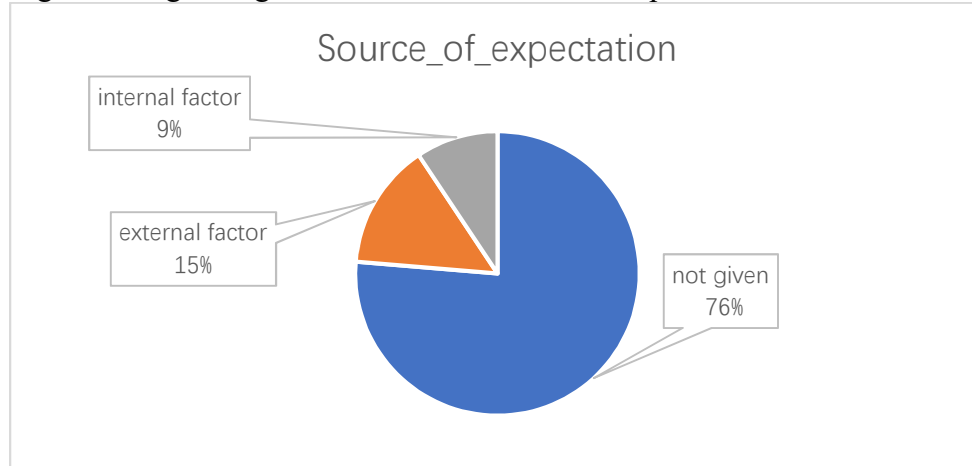

Figure 4 Distribution of Source_of_expectation

16) The **remarkable** fact about these utterances is that uninverted question structures were never explicitly introduced by the teacher. (AL-Qual-18)

17) The finding that neither the focus of students, self-statements nor their subjective anxiety reactions contributed to test performance is **surprising** given previous evidence that high-compared to low-test-anxious individuals evidence differences on these two variables. (CP-Quan-11)

18) Considering the large numbers of adult L2 learners and the variety of programs organized for them, it is **surprising** that the possibility of age constraints upon L2 achievement has received so little attention. (AL-Quan-40)

19) The inclusion of such wording in the proposed three-paragraph narrative description is all the more **striking** as the words 'dependency', 'submissive', or 'subservient' do not appear anywhere within the DSM-IV-TR diagnostic criteria or in the text description of BPD, which spans five pages. (CP-Quan-55)

20) Somewhat **surprisingly**, while observations indicate that Maria received explicit writing instruction from teachers in high school-though not frequently-she did not mention their pedagogical guidance as a resource she drew upon to write. (AL-Qual-69)

Tests of the full model versus the constant-only model were statistically significant only for "not given", of which only the research paradigm is a reliable predictor. The detailed results are presented in Table 6 below.

Table 6 Results of binary logistic regression analyses for Source_of_Expectation

Dependent variable	Independent variable	B	SE	Wald	p	Odds ratio	95% CI for OR Lower	Upper
unidentified	Discipline	-0.078	0.229	0.118	0.732	0.925	0.591	1.447
	Paradigm	-0.838	0.229	13.424	0.000	0.433	0.276	0.677
	Constant	0.497	0.199	6.236	0.013	1.644		
	$R^2 = .042$ (Cox & Snell); $R^2 = .056$ (Nagelkerke); Model $\chi^2(2) = 13.833$, $p = .001$							
internal	Discipline	0.366	0.352	1.082	0.298	0.694	0.723	2.874
factor	Paradigm	0.742	0.363	4.180	0.041	0.476	1.031	4.274
	Constant	-2.622	0.354	55.005	0.000	0.073		
	$R^2 = .017$ (Cox & Snell); $R^2 = .033$ (Nagelkerke); Model $\chi^2(2) = 5.460$, $p = .065$							
external	Discipline	-0.491	0.317	2.400	0.121	0.612	0.329	1.139
factor	Paradigm	0.894	0.328	7.427	0.006	2.445	1.285	4.650
	Constant	-1.976	0.293	45.318	0.000	0.139		
	$R^2 = .032$ (Cox & Snell); $R^2 = .054$ (Nagelkerke); Model $\chi^2(2) = 10.274$, $p = .006$							

Degree

As shown in Figure 5, most expressed surprises were neither mitigated nor boosted (see example 21 for an illustration). When they were mitigated, *somewhat* is the most frequently employed hedge (see example 22) followed by some epistemic modality such as *may* or *might* (example 23). In the case of boosted surprises, the choices were more diverse (see example 24 and 25).
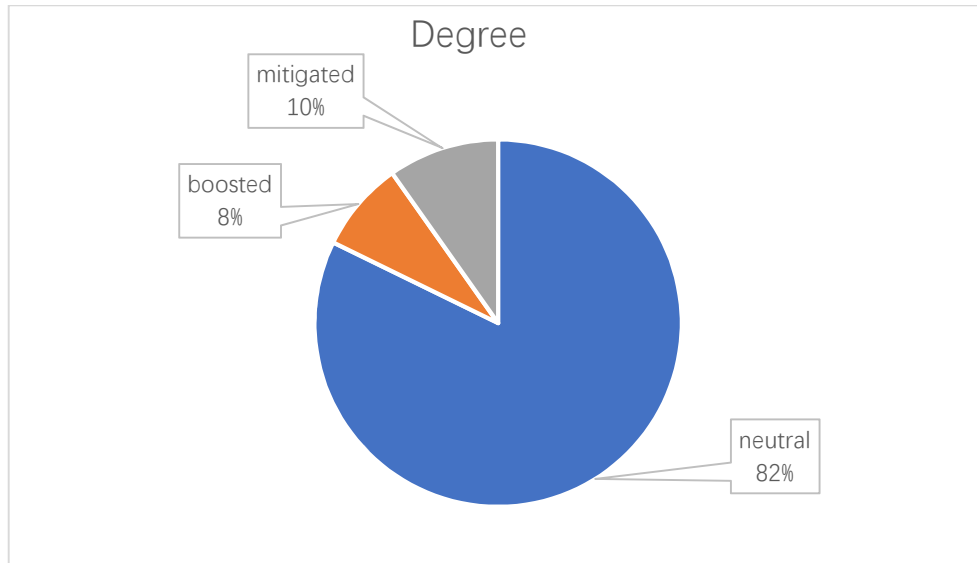
Figure 5 Distribution of Degree

21) **Surprisingly**, beginning-level learners of Japanese whose L1 was alphabetic (BA) did better than intermediate learners of Japanese whose L1 was alphabetic (IA), while IA had the lowest scores across all Chinese character types. (AL-Quan-68)

22) Somewhat **unexpectedly**, they were not unusually low on GATB aptitude scores. (CP-Quan-31)

23) These findings may seem **surprising**, as overachievers excelled in rote learning, which was generally considered mechanic, but not in tasks that demanded higher level processes, such as integrating multiple sources of information in a sentence or in a discourse context. (AL-Quan-77)

24) I certainly **did not expect** to be dealing with a suicidal person over a computer terminal. (CP-Qual-05)

25) And very **striking** were the revisions that took place almost simultaneously with the rereading and that were recorded with such a sense of urgency that the first version was not crossed out until a later rereading. (AL-Qual-38)

Binary logistic analyses on this FE show that the predictor variables combined cannot reliably predict the absence or presence of any of the three categories. The results are presented in the table below.

Table 7 Results of binary logistic regression analyses for Degree

Dependent variable	Independent variable	*B*	*SE*	*Wald*	*p*	Odds ratio	95% CI for OR	
							Lower	Upper
neutral	Discipline	0.026	0.226	0.013	0.910	1.026	0.659	1.597
	Paradigm	-0.380	0.226	2.836	0.092	0.684	0.439	1.064
	Constant	0.393	0.197	3.982	0.046	1.481		
	$R^2 = .009$ (Cox & Snell); $R^2 = .012$ (Nagelkerke); Model $\chi^2(2) = 2.862, p = .239$							
mitigated	Discipline	-0.739	0.362	4.167	0.041	0.477	0.235	0.971
	Paradigm	0.121	0.348	0.121	0.728	1.129	0.570	2.234
	Constant	-1.748	0.283	38.037	0.000	0.174		
	$R^2 = .014$ (Cox & Snell); $R^2 = .027$ (Nagelkerke); Model $\chi^2(2) = 4.487, p = .106$							
boosted	Discipline	-0.073	0.382	0.036	0.849	0.930	0.440	1.965
	Paradigm	-0.984	0.413	5.680	0.017	0.374	0.166	0.840
	Constant	-1.800	0.295	37.146	0.000	0.165		
	$R^2 = .019$ (Cox & Snell); $R^2 = .041$ (Nagelkerke); Model $\chi^2(2) = 6.245, p = .044$							

Resolution

We found that only slightly over a quarter of expressed surprises were resolved (see Figure 6 for the distribution). Closer examination revealed that all the resolved surprises concerned research findings, which suggested a connection between function and the FE configuration of surprise markers—a research direction not covered in this report but worth further exploration. Among the resolved surprises, "resolved by speculation" (see example 26) was found to be the most frequently used. As speculation is a sign of new knowledge, the close association of surprise with knowledge was confirmed. Both "resolved by research method" (see example 27) and "resolved by another finding" (see example 28) account for 4% of the expressed surprises respectively. The least used category for resolving surprises are "research background" (example 29) and "multiple factors" (example 30).
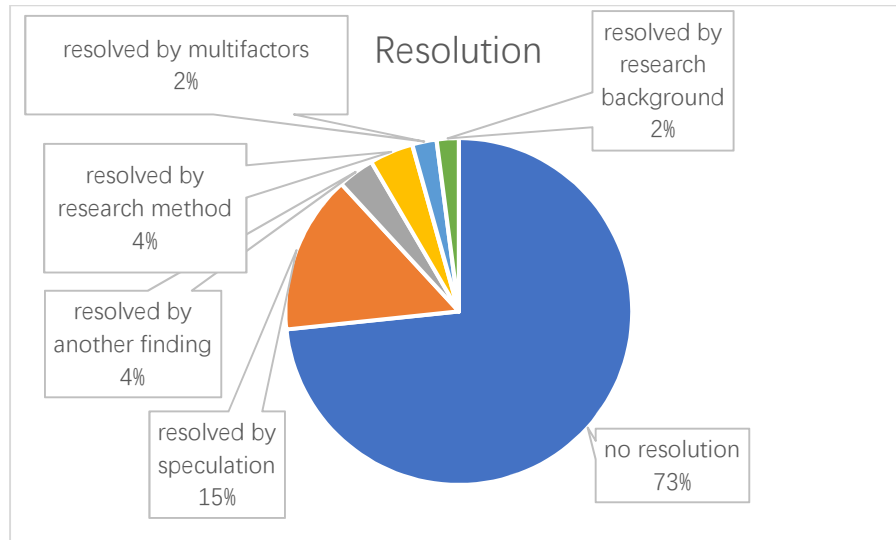
Figure 6 Distribution of Resolution

26) Judging from the data presented in the preceding sections the number of structural parallels between utterances of tutored and naturalistic LI learners is in fact quite **striking**. … <u>This observation suggests that at least some of the principles that govern naturalistic language acquisition also determine the processes by which students learn a foreign language under classroom conditions</u>. (AL-Qual-18)

27) Changes in ADHD severity on the other hand were not associated with subsequent changes in neuropsychological functioning for either TD or H/I children between the ages of 4–5 and 5–6 years, but **surprisingly**, changes in ADHD severity at the age of 5–6 years was negatively associated with changes in neuropsychological functioning at 6–7 years among both TD and H/I children. … <u>NEPSY has poorer reliability in very young children and some subtests differ between preschoolers and school-age children (Korkman, Kirk, & Kemp, 1998); these differences may have impacted our results in unmeasured ways</u>. (CP-Quan-56)

28) **This reversal of what was anticipated** can be explained <u>by noting that the low-level students experienced comprehension difficulties even when sandhi-variation was "absent"</u>. (AL-Quan-29)

29) **Contrary to expectations**, Chinese orientation was not significantly related to parenting efficacy … <u>These findings are consistent with other research that has suggested that orientation toward the new culture may have stronger implications for adjustment than orientation toward the ethnic culture (Abbott et al., 2003; Hwang & Ting, 2008; Ryder et al., 2000)</u>. (CP-Quan-41)

30) In relation to the degree to which mothers anticipated being in control of their child's responses, anxious mothers who had

children who went on to struggle in the task anticipated that they would have higher levels of control than those whose children did not struggle, and the reverse was found for nonanxious mothers (i.e., nonanxious mothers anticipated being more in control of children who did not go on to struggle in the task). These findings were **unexpected**, and they warrant further empirical examination… <u>It has certainly been established that highly anxious adults experience high levels of perceived responsibility (e.g., Salkovskis et al., 2000), and it might be anticipated that this will extend to perceived parental responsibility. This is consistent with the suggestion that anxious parents may view their child's environment in accordance with their own negative perspective of the world due to the systematic activation of hypervalent schemata, which guide cognitive processing toward negative aspects of the self and environment (Lester, Field, Oliver, & Cartwright-Hatton, 2009). (CP-Quan-59)</u>

For this FE, binary logistic analyses indicate that the presence or absence of "no resolution" and "resolved by research method" can be reliably predicted by the paradigm predictor. See Table 8 below for detailed results.

Table 8 Results of binary logistic regression analyses for Resolution

Dependent variable	Independent variable	B	SE	Wald	p	Odds ratio	95% CI for OR Lower	Upper
no	Discipline	0.291	0.231	1.595	0.207	1.338	0.852	2.103
resolution	Paradigm	-0.947	0.231	16.876	0.000	0.388	0.247	0.609
	Constant	0.288	0.197	2.141	0.143	1.334		
	$R^2 = .057$ (Cox & Snell); $R^2 = .076$ (Nagelkerke); Model $\chi^2(2) = 18.875, p < .001$							
resolved by	Discipline	-0.866	0.322	7.242	0.007	0.421	0.224	0.790
speculation	Paradigm	0.573	0.314	3.332	0.068	1.774	0.959	3.281
	Constant	-1.583	0.265	35.630	0.000	0.205		
	$R^2 = .034$ (Cox & Snell); $R^2 = .057$ (Nagelkerke); Model $\chi^2(2) = 10.983, p = .004$							
resolved by	Discipline	-1.381	0.667	4.285	0.038	0.251	0.068	0.929
research	Paradigm	1.881	0.776	5.880	0.015	6.557	1.434	29.980

method	Constant	-3.896	0.726	28.780	0.000	0.020		

$R^2 = .041$ (Cox & Snell); $R^2 = .136$ (Nagelkerke);
Model $\chi^2(2) = 13.436$, $p = .001$

resolved by	Discipline	-0.713	0.716	0.990	0.320	0.490	0.120	1.996
research	Paradigm	0.230	0.681	0.114	0.735	1.259	0.331	4.787
background	Constant	-3.366	0.561	35.946	0.000	0.035		

$R^2 = .004$ (Cox & Snell); $R^2 = .016$ (Nagelkerke);
Model $\chi^2(2) = 1.163$, $p = .559$

resolved by	Discipline	-0.845	0.612	1.906	0.167	0.429	0.129	1.426
another	Paradigm	-0.493	0.583	0.714	0.398	0.611	0.195	1.916
finding	Constant	-2.600	0.413	39.684	0.000	0.074		

$R^2 = .009$ (Cox & Snell); $R^2 = .030$ (Nagelkerke);
Model $\chi^2(2) = 2.787$, $p = .248$

resolved by	Discipline	0.939	0.846	1.233	0.267	2.558	0.487	13.419
multiple	Paradigm	0.939	0.846	1.233	0.267	2.558	0.487	13.419
factors	Constant	-4.943	0.934	27.999	0.000	0.007		

$R^2 = .008$ (Cox & Snell); $R^2 = .045$ (Nagelkerke);
Model $\chi^2(2) = 2.720$, $p = .257$

Summary of results

For the sake of discussion, we reorganize the statistical results by the independent variables and summarize the results of the binary logistic regression analyses in Table 9.

Table 9 Summary of statistical results

	Discipline	Research paradigm
Experiencer		
unnamed	No	No
author	No	No
participant	No	Qual > Quan
others	No	No
Trigger		
phenomenon	No	No
attribute	No	No
relationship	No	Qual < Quan
behavior	No	Qual > Quan
Source_of_expectation		
not given	No	Qual > Quan
internal factors	No	No
external factors	No	No
Degree		
neutral	No	No
mitigated	No	No
boosted	No	No
Resolution		
no resolution	No	Qual > Quan
resolved by speculation	No	No
resolved by research method	No	No
resolved by research background	No	No
resolved by another finding	No	No
resolved by multiple factors	No	No

Note. Qual = Qualitative; Quan = quantitative; No = no significant difference; > = more likely; < = less likely

Discussion

The statistical analyses revealed a noteworthy discrepancy: while neither of the two predictor variables can reliably predict the absence or presence of surprise markers, research paradigm can reliably predict the absence or presence of five FE sub-categories when Surprise frame is presented in RAs. This finding confirmed our initial concern that

distribution pattern at a finer grained level may be obscured by that at a higher level, which is part of the motivation for the present study. This result points to the necessity of conducting finer grained analyses of genre-specific language.

At first sight, the finding that discipline variable cannot reliably predict the absence or presence of any of the 20 FE sub-categories seems to be in line with the finding in Hu and Cao (2015) that there is no significant difference in the use of attitude markers between applied linguistics and clinical psychology. However, a closer examination would indicate the importance of considering the interaction between different types of metadiscourse. Hu and Cao (2015) found that applied linguistics used more boosters than clinical psychology, but the present study shows that applied linguistics is neither more likely to boost nor hedge expressed surprise than clinical psychology. Similarly, while Hu and Cao (2015) identified more self-mention in clinical psychology than in applied linguistics, the present study found no significant difference between the two disciplines when surprise markers are concerned. This discrepancy between metadiscoursal level pattern in Hu and Cao (2015) and the FE category level pattern in the present study can also be found between the two research paradigms. For example, while Hu and Cao (2015) found quantitative RAs used more hedges and boosters than qualitative RAs, in the present study, research paradigm is not a reliable predictor for whether one paradigm is more likely to hedge or to boost expressed surprises. Based on these comparisons, it is safe to say that frame-based analyses, as a cognitive approach to genre-specific language that center around frame-evoking LUs and take into consideration all contextual information, have the potential to provide us with deeper and finer grained understanding of academic writing.

The above being said, the fact that the discipline is not a reliable predictor for the absence or presence of all categories whereas the research paradigm can reliably predict the absence or presence of five categories across four FEs in the Surprise frame indicates that epistemological assumptions may be a major factor influencing the use of surprise markers in RAs. This claim is based on existing recognition that disciplines within soft fields tend to have similar epistemological assumptions whereas qualitative paradigm and quantitative paradigm subscribe to very different epistemological assumptions (Dörnyei, 2007). However, one should be very careful to generate this explanation to other attitude markers because the findings in the present study may merely reflect the special property of surprise as knowledge emotion (Silvia, 2009) which makes it sensitive to epistemological assumptions. Other attitudes such as importance and obligation may show different pattern. It is also an empirical question to ask whether interest and confusion show similar pattern to surprise since they are also knowledge emotions (Silvia, 2009). Perhaps it is because surprise is intrinsically related to knowledge and knowledge making, we found all the differences between the two paradigms at the five categories can be adequately explained by the distinctive epistemological assumptions between them and the unique property of surprise as knowledge emotion.

On the FE of Experiencer, qualitative RAs were found 13.33 times more likely to describe the surprises of participants—the largest difference in our data. This finding perhaps best reflects the epistemological differences between the two research paradigms. While quantitative research in social science "reduces human behaviors, attitudes,

performances, demographics, and other attributes to numerical information and mathematically modeled relationships" (Cao & Hu, 2014, p. 27), thereby largely eliminating the necessity of giving detailed descriptions of behaviors and attitudes of individuals, qualitative research values more a "qualitative/subjective description, empathetic understanding, and exploration" (Johnson & Christensen, 2012, p. 34) of the thoughts, behaviors and experiences of individuals involved in the study. The latter's epistemological preference can conceivably provide more opportunities to describe the attitude of participants, including their surprises.

In terms of Trigger, quantitative RAs were 3.88 times (the result of dividing 1 by the Odd Ratio) more likely to express surprises towards "relationship" proposition whereas qualitative RAs were more likely to express surprises towards "behavior". This possibly reflects the different focuses between these two research paradigms in knowledge-making. Quantitative RAs are much more likely to express surprise towards "relationship" simply because knowledge making in quantitative studies typically revolve around "formulation and testing of hypotheses about the relationship between variables" (Hu & Cao, 2015, p. 22). By contrast, qualitative RAs are more likely to show surprises toward "behavior" because in social sciences human subjects are frequently involved and qualitative studies tend to directly describe the observed behavior of those human subjects.

Now let's turn to Source_of_expectation and Resolution, two of the essential components of the surprise routine (Tutin, 2015). It was found that qualitative RAs were less likely to present either Source_of_expectation or Resolution than quantitative RAs (Odd Ratio being 0.433 and 0.388 respectively for "unidentified" and "no resolution). We would like to argue that these two findings conform to our hypothesis that distinctive epistemological assumptions underpinning different research paradigm may affect the expression of a specific attitude, particularly when this attitude is related to knowledge. As pointed out by Kövecses (2015), surprise is the result of the conceptual incongruence between one's expectation and reality. Considering this, that qualitative RAs are less likely to present expectations may be attributed to the fact that qualitative researchers usually do not begin their academic inquiry with specific expectations but instead keep an open mind to whatever may emerge from the data (Maxwell, 2012). Similarly, the observation that qualitative RAs are less compelled to resolve the expressed surprises may result from the tendency of qualitative studies to generate themes for further exploration instead of testing hypotheses. By contrast, quantitative studies usually set out to test a hypothesis that is often based on existing theories, previous research or prevailing contexts, which constitute the sources of expectation for quantitative enquiries. Consequently, when expectations are not borne out, quantitative RAs may be more obliged to provide an explanation. Interestingly, although the two research paradigms differ in the likelihood of presenting Source_of_expectation and Resolution, there is no significant difference between them in terms of the type of expectations and resolutions when these two FEs are presented. Given the results that more Source_of_expectation are "external factors" and "resolved by speculation" is the most frequent category for Resolution, this finding suggests a typical knowledge-making routine, in which researchers first formulate hypotheses based on previous research, then collect and analyze data to verify the hypotheses, then decide whether the results confirm to or

violate expectations, and finally either provide further evidence for existing theory if the results confirm to expectations or form new hypotheses if the results violate expectations. In this sense, the surprise routine as described in Tutin (2015) is only part of this knowledge-making routine because there are inevitably expectations that are indeed borne out. From a frame semantics perspective, this means we may be able to generate an umbrella frame that could characterize this knowledge-making routine. It would also be interesting and potentially revealing to look at the ratio between confirmation of expectations and violation of expectations in RAs from different disciplines or published at different times to have a deeper understanding of disciplinary difference in or diachronic change of knowledge making, though it falls out of the scope of the present study.

Conclusions and contributions

The present study investigates the use of surprise markers across two social disciplines and two research paradigms with an analytical framework generated from seven semantic frames that could be evoked by surprise markers. Statistical analyses revealed significant paradigmatic differences along five FEs of Surprsie frame, and all those differences may be attributed to distinctive epistemological assumptions between the two research paradigms. As one of the initial attempts that take a cognitive approach to genre-specific language, this study has the potential to contribute to both the theory of frame semantics and studies of English for Academic Purpose (EAP) in at least the following two ways:

Theoretically, the present study extends the application of frame semantics to discourse analysis. It has the potential to create a new research direction for genre analysis because it tries to account for the genre-specific linguistic features from a cognitive perspective, which has not been systematically practiced before. This new direction is semantically oriented and cognitively informed, which could enrich our understanding of academic writing by adding new perspective to the current functionally heavy paradigms. For example, future studies in this direction could investigate other types of attitude markers in a similar way, for example, how "importance" or "interest" are expressed in academic writing.

Methodologically, since frame-based analyses are inherently dimensional because frame semantics offers a structured understanding of linguistic meaning, this study will provide a possible solution to two drawbacks of functionally oriented studies mentioned in the beginning of this paper. They are, 1) pattern found at a functional level (e.g., attitude markers) may not be applicable to a semantic level (e.g., surprise markers), and 2) the interaction between different types of linguistic markers are overlooked. These problems are almost inevitable for studies that analyze linguistic markers only by pre-determined categories because such analyses are based on a questionable assumption that all lexical items that belong to the same metadiscoursal category must assume the same discourse function, no matter what context they are used in. By contrast, frame-based analyses are fully contextualized, bringing in different types of linguistic markers for characterizing the use of one frame-evoking LU. In this way, the two problems mentioned above can be avoided.

To conclude, although this kind of analysis is inescapably time-consuming as it requires strenuous manual coding, we believe it is a fresh research line that merits more scholarly attention. We sincerely encourage more frame-based studies on academic discourse, complementing the currently socio-functionally focused EAP study with a cognitive perspective.

References

Abdi, R. (2002). Interpersonal metadiscourse: an indicator of interaction and identity. *Discourse Studies*, *4*(2), 139–145.

Abdollahzadeh, E. (2011). Poring over the findings: Interpersonal authorial engagement in applied linguistics papers. Journal of Pragmatics, 43, 288–297.

Allan, K. (2016). 3 A history of semantics. In N. Riemer (Ed.), *Routledge handbooks in linguistics. The Routledge handbook of semantics* (pp. 48–68). Milton Park, Abingdon, Oxon, New York, NY: Routledge.

Becher, T. (1987). Disciplinary discourse. *Studies in Higher Education*, *12*(3), 261–274.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Buckingham: The Society for Research into Higher Education & Open University Press.

Bernstein, B. (1999). Vertical and Horizontal Discourse: An essay. *British Journal of Sociology of Education*, *20*(2), 157–173.

Cao, F., & Hu, G. (2014). Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics*, *66*, 15–31.

Casti, J. L. (1994). *Complexification: Explaining a paradoxical world through the science of surprise*. New York, NY: Harper Collins.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London, New York: Routledge Falmer.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, Calif., London: Sage.

Davies, Mark (2017). Corpus of Contemporary American English (1990-2017). Available online at https://www.english-corpora.org/coca/, accessed on 8/15/2019.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies. Oxford applied linguistics*. Oxford: Oxford University Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3-4), 169–200.

Ellsworth, P. C., & Scherer, K. R. (2009). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Series in affective science. Handbook of affective sciences.* New York: Oxford University Press.

Fellbaum, C. (1998). *WordNet: An electronic lexical database. Language, speech and communication*. London: M.I.T. Press.

Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, *280*(1), 20–32.

Fillmore, C. J. (1977). Scenes-and-frames semantics. In A. Zampolli (Ed.), *Fundamental Studies in Computer Science: Vol. 5. Linguistic Structures Processing* (pp. 55–81). Amsterdam: North Holland Publishing.

Fillmore, C. J. (1982). Frame semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm* (pp. 111–137). Seoul, Korea: Hanshin Publishing Company.

Fillmore, C. J. (2006). Frame semantics. In D. Geeraerts (Ed.), *Cognitive linguistics research: Vol. 34. Cognitive linguistics: Basic readings* (pp. 373–400). Berlin, New York: Mouton de Gruyter.

Fillmore, Charles J.; Baker, Collin (2010). Frame approaches to semantic analysis. In Bernd Heine,

Heiko Narrog (Eds.): *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press (Oxford handbooks in linguistics), pp. 313–340.s

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, *18*(12), 1050–1057.

Foster, M. I., & Keane, M. T. (2015). Why some surprises are more surprising than others: Surprise as a metacognitive sense of explanatory difficulty. *Cognitive Psychology, 81*, 74–116.

Frijda, N. H. (1986). *The emotions*. *Studies in emotion and social interaction*. Cambridge: Cambridge University Press.

Gillaerts, P., & van de Velde, F. (2010). Interactional metadiscourse in research article abstracts. *Journal of English for Academic Purposes*, *9*(2), 128–139.

Goddard, C. (2015). The complex, language-specific semantics of "surprise". *Review of Cognitive Linguistics, 13*(2), 291–313.

Gross, M. (2010). *Ignorance and surprise: Science, society, and ecological design*. *Inside technology*. Cambridge, Mass., London: MIT Press.

Guba, E. G. (Ed.). (1990). *The Paradigm dialog*. Thousand Oaks, CA, US: Sage Publications, Inc.

Harper, D. (2008). Clinical psychology. In C. Willig & W. Stainton-Rogers (Eds.), *The SAGE Handbook of Qualitative Research in Psychology* (pp. 430–454). London, UK: Sage.

Harwood, N. (2005). 'Nowhere has anyone attempted … In this article I aim to do just that': A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics*, *37*(8), 1207–1231.

Hood, S. (2012). Writing discipline: Comparing inscriptions of knowledge and knowers in academic writing. In F. Christie & K. Maton (Eds.), *Disciplinarity: Functional Linguistic and Sociological Perspectives* (pp. 106–128). New York: Continuum.

Hu, G., & Cao, F. (2015). Disciplinary and paradigmatic influences on interactional metadiscourse in research articles. *English for Specific Purposes*, *39*, 12–25.

Hyland, K. (1996). Writing Without Conviction? Hedging in Science Research Articles. *Applied Linguistics*, *17*(4), 433–454.

Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. *Continuum discourse series: Vol. 39*. London: Continuum.

Hyland, K. (2005b). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173–192.

Hyland, K. (2008). Genre and academic writing in the disciplines. *Language Teaching, 41*(04), 543–562.

Hyland, K. (2014). Disciplinary discourses: Writer stance in research articles. In C. N. Candlin (Ed.), *Writing: Texts, Processes and Practices* (pp. 99–121). Oxon: Routledge.

Hyland, K., & Jiang, F. (2017). 'We Believe That … ': Changes in an Academic Stance Marker. *Australian Journal of Linguistics*, *38*(2), 1-23.

Hyland, K., & Jiang, F. (2018). "In this paper we suggest": Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, *51*, 18–30.

Johnson, B., & Christensen, L. B. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). Thousand Oaks, Calif.: Sage.

Khedri, M., Chan, S. H., & Ebrahimi, S. F. (2013). An exploration of interactive metadiscourse markers in academic research article abstracts in two disciplines. *Discourse Studies, 15*, 319–331.

Kipfer, B. A. (2010). *Roget's international thesaurus* (7th ed.). New York: Collins.

Kövecses, Z. (2015). Surprise as a conceptual category. *Review of Cognitive Linguistics*, *13*(2), 270–290.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data.

Biometrics, *33*, 159-174.

Macedo, L., Reisenzein, R., & Cardoso, A. (2004). Modeling forms of surprise in artificial agents: Empirical and theoretical study of surprise functions. In K. Forbus, D. Gentner, & T. Rigier (Eds.), *Proceedings of the 26th annual conference of the Cognitive Science Society* (pp. 873–878). Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.

Maton, K. (2000). Languages of Legitimation: The structuring significance for intellectual fields of strategic knowledge claims. *British Journal of Sociology of Education*, *21*(2), 147–167.

Maton, K. (2014). *Knowledge and knowers: Towards a realist sociology of education*. Milton Park, Abingdon, Oxon: Routledge.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (3rd ed.). *Applied social research methods series: Vol. 41*. Thousand Oaks, Calif., London: Sage.

Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, *21*(3), 251–274.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, *70*(2), 151–167.

Mur-Dueñas, P. (2010). Attitude markers in business management research articles: a cross-cultural corpus-driven approach. *International Journal of Applied Linguistics*, *20*(1), 50–72.

Mur-Dueñas, P. (2011). An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics*, 43(12), 3068–3079.

Mu, C., Zhang, L. J., Ehrich, J., & Hong, H. (2015). The use of metadiscourse for knowledge construction in Chinese and English research articles. *Journal of English for Academic Purposes*, *20*, 135–148.

Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions* (1. paperback ed.). Cambridge: Cambridge Univ. Press.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology: Emotions, relationships and health* (5th ed., Vol. 7). Beverly Hill, CA: Sage.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., & Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice* (v. 21): Institut für Deutsche Sprache, Bibliothek.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion*. Hillsdale, N.J.: L. Erlbaum Associates.

Schneirla, T. C. (1959). An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation* (pp. 1-42). Lincolnm, NE: University of Nebraska Press.

Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(1), 48–51.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838.

Soriano, C., Fontaine, J. R. J., & Scherer, K. R. (2015). Surprise in the GRID. *Review of Cognitive Linguistics*, *13*(2), 436–460.

Teigen, K. H., & Keren, G. (2003). Surprises: low probabilities or high contrasts? *Cognition*, *87*(2), 55–71.

Tutin, A. (2015). Surprise routines in scientific writing. *Review of Cognitive Linguistics*, *13*(2), 415–435.

Wignell, P. (2007). Vertical and horizontal discourse and the social sciences. In F. Christie & J. R. Martin (Eds.), *Language, knowledge and pedagogy: Functional linguistic and sociological perspectives* (pp. 184–204). London, New York: Continuum.

Appendix: The complete list of search words used to identify surprise markers

Category	Part of speech	Search words
Unsurprisingness / expectation	Verb	*expect, anticipate, predict, assume, suppose, presume, take for granted, hypothesize, theorize, reckon, await*
	Noun	*expectation, expectancy, anticipation, prospect*
	Adjective	*expected, unsurprising, anticipated, awaited, hoped-for, expectable, foreseen, foretold, predicted, matter-of-course, supposed, assumed, usual, common, predictable*
	Adverb	*expectedly, unsurprisingly, expectably, supposedly, usually, commonly, expectably, predictably*
Surprisingness / inexpectation	Verb	*surprise, amaze, astonish, shock, strike, astonish, impress, astound, flabbergast, stun, floor, explode a bombshell, ball over, blow out, take aback, marvel*
	Noun	*surprise, astonishment, amazement, bombshell, shock, wonder, wonderment, admiration, stupefaction, surprisingness, unexpectedness, inexpectation*
	Adjective	*surprising, startling, astonishing, astounding, remarkable, alarming, shocking, amazing, disconcerting, disturbing, surprised, startled, astonished, astounded, shocked, amazed, unexpected, amused, unusual, uncommon, intriguing, exceptional, incredible, extraordinary, stunning, unforeseen, unpredictable, sudden, unannounced, unheralded, unpredicted, unanticipated, unlooked-for, unhoped, unhoped for, unthought, unthought of, unprovided for, jolting, uncharacteristic*
	Adverb	*surprisingly, amazingly, astonishingly, remarkably, unusually, unexpectedly, extraordinarily, incredibly, uncommonly, uncharacteristically, exceptionally, alarmingly, suddenly*

---

[i] FrameNet is an ongoing project to build a lexical database of English that is both human- and machine-readable. This project is created by Charles Fillmore and his colleagues based on the theory of frame semantics and is has been in operation at the International Computer Science Institute in Berkeley, California since 1977. It is now an important tool for lexical research with a semantic orientation (Ruppenhofer et al., 2016).

[ii] The annotation convention adopted in FrameNet is used for Example 1 to 3 for the sake of discussion because these examples are taken from FrameNet. The rest of the examples, which are from our corpus, are presented in a simpler format, with focus only on the FE under discussion.