

**Macro Discourse Markers in TED Talks:
How Ideas are Signaled to Listeners**

Kanokrat Uicheng

Faculty of Arts, Chulalongkorn University, Thailand
E-mail: kanokrat.uc@gmail.com

Michael Crabtree

Faculty of Arts, Chulalongkorn University, Thailand
E-mail: crabtree@chula.ac.th

Abstract

This study investigated the macro discourse markers used by TED speakers to signal ideas to listeners. 150 TED transcripts from six talk genres: technology, entertainment, design, business, science, and global issues were compiled to identify the frequency and distribution of macro discourse markers. This investigation was conducted to confirm the hypothesis that genres of talks did not affect the distribution of discourse markers. The study also hypothesized that a number of macro discourse markers which explicitly containing the word 'idea' or 'ideas' in them could be found. The results confirmed that the occurrences of macro discourse markers were

not subject to talk genres and that 51 markers containing the word 'idea' and 'ideas' were found.

Keywords: discourse marker, macro discourse marker, idea signalers corpus-based study, public talk, talk genres

Introduction

When listening to a public speech, one may notice that there are factors contributing to or even preventing us from understanding the message that the speaker is trying to deliver. Sometimes it is the speed or pace of speaking which might be too fast to follow. Sometimes it is the complexity of concept which is not clearly explained by the speaker. As a result, most listeners need to count on linguistic devices like discourse markers to help them comprehend the speech better. While some discourse markers work as pause fillers, others work as sequence, fact, or idea signalers.

Several studies agree that discourse markers help listeners to understand academic spoken discourse like lectures, more easily (Chaudron & Richard, 1986; Flowerdew & Tauroza, 1995; Rido, 2015). However, only a small number of studies have investigated specifically how speakers use discourse markers to aid a successful talk. This is possibly because most researchers have not identified a group of speakers who represent a successful model that they can study for the way discourse markers are used. This study, however, argues that TED Talk speakers represent such a model and that one can learn from them the role that discourse markers play in their proficiency as public speakers.

TED is a non-profit organization which offers short powerful talks on various topics to audiences around the world. People can either attend their live conferences or watch TED talk videos online. The topics of TED talks vary, although most can be grouped under one or more of the following: technology, entertainment, design, business, science, and global issues. There are approximately 2,700 talks available on the internet and their popularity has been increasing. One talk by Sir Ken Robinson about education and

creativity, for example, has been viewed more than 50 million times and has been translated into more than 60 languages. TED speakers have also been recognized in a number of studies, which have attempted to identify features of TED talks which distinguish them from other types of spoken discourse.

For instance, Caliendo and Compagnone (2014), who defined TED Talks as speech events where speakers from different fields share their knowledge with a large audience, compared TED Talk speakers to university lecturers in their use of ELVs, or epistemic lexical verbs. The four most recurrent ELVs: *see*, *show*, *know*, *think*, were investigated using two corpora of spoken discourse: TED_ac and MICASE. The results showed that unlike in lectures, the pronoun 'we' as used in TED talks, typically included the speakers and their team but excluded the audience. Their use of 'we', as in *we see*, *we know*, and *we think*, helped TED speakers build an image of themselves as experts who delivered their stories or discoveries confidently. This particular use of epistemic lexical verbs with the pronoun 'we' was identified as one element that distinguishes TED Talk speakers from university lecturers.

Tsai (2015) also compared TED speakers with university professors, but in terms of prosodic voice characteristics. His main aim was to identify the characteristics that distinguished TED speakers from other public speakers. He collected 5-minute audio samples from TED talks and lectures, then developed a discriminative classifier to predict whether the samples were from TED talks or from lectures. Predictions were based on a set of features derived from pitch and energy. It was found that TED talks tended to be more 'dense' in that there was less time in silence and more time in high-energy speech. TED speakers were also consistent in delivering lengthy messages while lecturers were found to have more varied chunks of silence and speech. The study also suggested that TED speakers had deeper voices than university professors. Tsai pointed out that these results were possibly due to the fact that TED talks were shorter than the lectures. Furthermore, he suggested that one element that made TED Talk speakers successful public speakers is their ability to deliver a consistently

high-energy talk. In Tsai's opinion, TED talk speakers *'have something to say and know how to say it'* (p.4).

Chris Anderson, the Head of TED Talk, however, views TED speakers from another perspective. He believes that all TED speakers have one element in common. That element is their *'idea'* which they see as worth sharing with the audience. Despite the fact that TED speakers talk on different topics, they have the same goal, which is to successfully deliver their ideas to the listeners. This, Anderson suggests, is the key common feature shared by TED speakers.

Given the time limitation of approximately 20 minutes and the large audience from different backgrounds, TED speakers need to be clear when they deliver their *'ideas'*. They must ensure that all listeners are with them and do not miss any important concepts during the talk. This is where discourse markers are believed to play a role. A discourse marker is a linguistic feature that helps indicate relations between utterances in the discourse. With ideas as elements to deliver, it is assumed that TED speakers use discourse markers to signal their thoughts. This study aims to identify such discourse markers used by TED speakers, especially what Chaudron and Richards (1986) term the macro version, such as 'I believe' and 'the point is'.

Review of the Literature

Discourse Markers: Terminology and Definitions

The markers which are the focus of this study are macro discourse markers. However, when researchers use the term 'discourse marker' in general and more specifically, the term 'macro discourse marker', they do not always use them to refer to exactly the same elements.

For example, Chaudron and Richards (1986) proposed that there are four versions of lecture discourse where the use of discourse markers could be found: 1) baseline version, 2) micro version, 3) macro version, and 4) micro-macro version. The term 'version' here, as the researcher understands it, refers to the type of lecture, considering this explanation by Chaudron and Richards:

A second version of the lecture, the 'Micro' version, was then constructed, in which various markers of Intersentential relations, framing of segments, and pause fillers, were inserted.

Chaudron and Richards (1986:71)

From this excerpt, it is clear that the term 'version' refers to a type of lecture. The term 'markers' being used here also refers to 'discourse marker', as the title of the paper, *The Effect of Discourse Markers on the Comprehension of Lecture*, suggests. Chaudron and Richards divided lectures into four types and identified discourse markers found in each type of lecture. Below are examples of the discourse markers found in four different lecture versions.

The first version of lecture is called the baseline version. Chaudron and Richards claim that there are no markers needed for the baseline version of lecture. All of what appears in the discourse is necessary to convey the meaning. For example, every word in this sentence, *By 1803, the original 13 colonies had doubled in size...* as exemplified by Chaudron and Richard, is semantically required.

The second type of lecture discourse is called the micro version. Markers in this version signal intersentential relations, frame segments, and fill pauses. They can be divided into 5 sub-categories: which are temporal link, causal link, contrastive relationship, relative emphasis, and framing or segmentation. Following are examples of markers in each type as suggested by Chaudron and Richards.

Temporal link: *then, and, now, after this, at that time*

Causal link: *because, so*

Contrastive relationship: *but, actually*

Relative emphasis: *you see, unbelievably, of course*

Framing / segmentation: *well, ok, alright?*

The third type is called the macro version. Markers in this lecture type can be signalers or metastatements that indicate major propositions or important transition points in the lecture. Examples of these markers are:

*To begin with, what we've come to by now was that,
let's go back to the beginning, What I'm going to talk about today
That/this is why As you may have heard*

The last type is called the micro-macro version. The name suggests that this is a combination between the micro version lecture and macro version lecture. The markers from both types are used together as in these examples provided by Chaudron and Richards.

Well, *to begin with,*
And so, *what we've come to by now was that,*

However, similar to other researchers (Morell 2000, Bellés 2004, Rido 2010) who called the discourse markers in Chaudron and Richards' study 'Micro markers' and 'Macro markers' after the names of the lecture types, this present study calls the markers in micro version and macro version lectures 'micro discourse markers' and 'macro discourse markers', respectively.

Unlike Chaudron and Richards who examined markers in the discourse of lectures, Schiffrin (1987) studied the use and distribution of markers in everyday discourse. She defined discourse markers as '*sequentially dependent elements that bracket units of talk*' (p.57). Markers as defined by Schiffrin are non-obligatory and syntactically detachable. They usually occur as utterance initials that correlate with ongoing talk and text. These discourse markers appear in several word classes, such as, conjunctions, interjections, adverbs, and lexicalized phrases. Following are Schiffrin's examples of discourse markers from each class.

Conjunctions: *and, but, or*
Interjections: *Oh*
Adverbs: *now, then*
Lexicalized phrases: *y'know, I mean*

Without looking at any particular types of discourse, Fraser (1996), classified '*discourse markers*' as a type of pragmatic marker.

In his definition, a discourse marker refers to an expression that signals the relationship between a basic message and the foregoing discourse. This marker does not influence the representative sentence meaning, but only helps to mark the procedural meaning, as in this example: *Mary went home. After all, she was sick.* This marker 'After all' does not change the representative meanings of the two sentences. It only signals to the readers that there is a relationship between these two incidents: *Mary went home* and *She was sick.* This marker 'after all', as well as items such as *thus, moreover, however, and in other words* are Fraser's examples of discourse markers.

Another researcher who studied discourse markers in lecture discourse is Morell (2000). She investigated the role of discourse markers in two types of lectures: interactive and non-interactive. Morell adopted Chaudron and Richards (1986)'s classification of micro and macro markers but made some addition to the sub-category of micro markers. Following are examples of discourse markers from her classification.

Non-interactive Micro-markers

Segmentation	Temporal	Causal	Contrast	Emphasis	Elicitation
Ok, and, or, now, well	And then, after that	So, that means, which means, because, so (that), therefore	But, although	In fact, of course, such as, note	Ready?

Interactive Micro-markers

Segmentation	Temporal	Causal	Contrast	Emphasis	Elicitation
Ok, and, or, now, anyway	Then	So	But, unless	Of course, as you know	Anything else? What? Why not? Louder please

Non-interactive Macro-markers

Starter	Elicitation	Accept	Attitudinal	Metastatement	Conclusion
Today I'm going to talk about... To begin... The reasons...	What do we mean by... Remember... Any questions...	-	I would dare to say... I believe that... I wouldn't doubt it.	I have a quote for you... I'd like to read this to you...	What you have seen in this lecture first was...

Interactive Macro-markers

Starter	Elicitation	Accept	Attitudinal	Metastatement	Conclusion
We are going to get started Can I have your attention? We will begin now	I have a question for you. Do you think...? Do you agree...?	That's right. That'	I think... It's a difficult question to answer.	To back up that statement	To finish today's lecture, We'll continue with this tomorrow

A more recent definition of discourse marker is that proposed by Carter and McCarthy (2006). Similar to Fraser (1996), Carter and McCarthy defined 'discourse marker' as a sub-type of pragmatic marker. The two researchers divided pragmatic markers into 1) discourse markers, 2) stance markers, 3) hedges, and 4) interjections. The discourse markers in Carter and McCarthy's schema include expressions like *so*, *well*, and *anyway*.

It can be seen from all the definitions above that researchers have approached discourse markers differently. Some specified discourse markers in particular types of discourse such as lectures, while others defined 'discourse marker' in more general term. Moreover, since their definitions of discourse markers differ, the functions of markers are not the same. Discourse markers as defined by Chaudron and Richards, for example, can perform various functions from linking, signaling relationships, signaling transition points, or filling pauses, while in Fraser, the function of most discourse markers is to mark the procedural meaning. It is necessary to consider both discourse types and functions when one would like to adopt a particular definition of discourse markers in their study.

Discourse Markers in this Study

This study finds the classification of discourse markers proposed by Chaudron and Richards best fits its purposes. While Schiffrin focuses on discourse markers in everyday discourse, which may include conversations and written texts, Chaudron and Richard focus only on lectures. TED talks may be considered as similar to lectures in that both are delivered by a speaker who talks in a highly monologic manner, in academic spoken discourse, and in front of a large audience. It is assumed, therefore, that the classification of discourse markers in lectures is applicable to the classification of discourse markers occurring in TED presentations, although the distinction between markers found in the two discourses might be a topic for later investigation.

While the present study adopts the taxonomy of discourse markers posited by Chaudron and Richard, it also uses the list of markers in Morell as a guideline for the manual investigation of discourse markers in the first part of the data collection. However, this study does not intend to explore discourse markers of any functions. Rather, it aims to look particularly for markers that TED speakers use to signal ideas, and therefore macro discourse markers seem most relevant. Another reason for choosing macro discourse markers is that a large number of previous studies investigated the use of micro discourse markers (e.g. *well, really, you know?*) while relatively few have investigated macro discourse markers. Most macro DMs are, therefore, still unexplored. This study aims to fill this gap by studying the use of macro discourse markers in the spoken discourse of TED talks, especially when those markers signal ideas to the listeners.

Previous Studies

Discourse markers have been the subject of investigation in numerous studies and in a variety of languages (Roggia 2012, Landone 2012, Rhee 2013, Tanghe 2015). Some studies investigated the use of markers in written discourse (Zhang 2000, DÜlger 2007, Jalilifa 2008, Rezanova & Kogut 2015,) but in others,

including the present study, the focus has been on discourse markers used in spoken discourse.

Studies that Investigated Discourse Markers in Spoken Discourse

Lee-Goldman (2010) studied different senses of ‘no’ via two speech corpora: the ICSI Meeting corpus and the Fisher English Training Corpus. He found that, in addition to the normal sense of ‘no’ as negation or rejection, there are three more senses of ‘no’ as discourse markers. ‘No’ as a discourse marker involves: 1) topic shifting, 2) misunderstanding mitigation, and 3) turn-taking management. Following is an example of ‘no’ used as a discourse marker to shift from a joke and then return to the conversation.

1 B: i did and i started thinking about it and i contributed
2 it all to my mother
3 [ha ha it's all on her side of the family] ha ha ha ha
4 A: [ha ha ha ha ha]
5 A: maternal heart attack
6 B: yes
7 A: ((laugh))
8 B: yes
9 B: i blame her
10 B: --> *but no* um
11 B: --> so that's how i kind of got involved cause i like doing this
12 B: stuff you know and uh
13 B: it's interesting and you know and i enjoy talking to other 14 people
and...

Lee-Goldman (2010: 9)

By providing evidence of ‘no’ functioning as a discourse marker, Lee-Goldman also pointed out that ‘no’ does not always occur alone but can be accompanied by another discourse marker, such as ‘*but*’ as in the excerpt above.

In a later study, Tay (2011) examined the co-occurrences of discourse markers and metaphoric utterances in English psychotherapeutic talks. He analyzed two extracts obtained from the Counselling and Psychotherapy Transcripts published online by Alexander Street Press. In the first extract, Tay investigated the patient’s use of discourse markers which were followed by metaphoric utterances. He discovered that the patient repeatedly

used the discourse markers, ‘*I mean*’ and ‘*you know*’ (or *you know what*) to discuss a situation to which the therapist could relate. In the second extract, Tay investigated the discourse markers used by the therapist to explain an abstract concept to the patient. He found that the therapist also used the discourse markers ‘*you know*’, ‘*and*’, and ‘*right?*’ frequently with the patient. These three markers preceded metaphoric utterances as can be noticed below in this part of the excerpt.

“If you have this huge – ***you know***, it’s kind of like even an example – and we have to begin wrapping up. But like ***-you know***, like a person – the people who have to disbomb bombs? ***You know*** – I forget what they call them, but if a building even called, the bomb squad comes in, ***right?***”

Tay then concluded that metaphors and discourse markers are co-occurring features found in psychotherapeutic talks. The use of discourse markers, both by patient and the therapist, was motivated by the metaphoric expressions that followed.

In response to a perceived lack of research into discourse markers in languages other than English, Lai and Lin (2012) designed a study that looked specifically into the use of Chinese discourse markers by Chinese-speaking seniors in descriptive and narrative discourses. Thirty Chinese-speaking seniors with Alzheimer’s disease (AD) and another thirty seniors without AD participated in this study. The findings revealed that both groups of participants used more discourse markers in the narrative task than in the descriptive task. However, the group with AD used the discourse markers with less frequency and variation. The discourse markers frequently used by the participants with AD were, for example, 应该 (should), 好像 (seem), and 大概 (probably).

Studies that Specifically Investigated Macro Discourse Markers

Bellés (2004) investigated the similarities and differences of discourse markers used in American and British lectures in the Social Sciences. She compiled a corpus which comprised twenty

transcripts, ten American English lecture transcripts from the University of Michigan and ten British English lecture transcripts from the University of Birmingham. The result of her corpus investigation showed that there were similarities in the use of macro-markers between American and British English lectures. Both tended to use the same macro-markers, especially in these following types: metastatements, attitudinal, and starter. However, Bellés found a distinction between the use of two metastatement macro-markers: *Let me (lemme)* and *let's*. American lecturers used these two markers more frequently, especially *Let me* or *lemme*, which did not occur in the British lectures at all. This finding implied, as suggested by Bellés, that American lecture discourse uses less formal speech features than British lecture discourse.

Rido (2015) investigated the use of discourse markers in science lectures by a professor of physics in a university in Malaysia. The investigation was done by observing and recording two of her lecture sections to identify discourse markers used during the lessons. Two types of discourse markers were found. The first type was macro markers which functioned as rephrasers and topic shifters. To rephrase what had been said, the lecturer in this study used *that mean*, *I mean*, and *which means to say that*. She was also found using 'so', 'now', and 'so anyway' as topic shifters. The second type of marker was the micro marker. Rido identified two sub-types of markers in this group: additional makers and causal markers. Both were used to signal internal or ideational relations within the sentences. The additional markers found were 'and' and 'or' while the causal marker found was 'because'. Rido believes that the use of these discourse markers can help the students to comprehend the lectures.

The abovementioned studies provide an overview of discourse markers which have been investigated in different contexts and different types of discourse. One of the most frequently examined types of discourse among researchers is the lecture. However, the discourse markers examined in most studies on lectures were often one of these: *so*, *well*, *right*, *now*, *really*, *I mean*, and *you know*. This present study seeks to identify a wider range of

discourse markers, especially the macro version, which TED speakers use to efficiently deliver ideas. Specifically, the study attempts to answer the following research questions:

Research Questions

1. What are the macro-discourse markers used to signal ideas in TED Talk presentations?
2. Are the macro discourse markers found in TED talks specific to the topics of the talks, or are they equally distributed across different genres?

The hypotheses which the study sets out to test are:

Hypotheses

1. There will be a wide range of macro discourse markers used to signal ideas in TED Talk presentations.
2. Many macro-discourse markers contain the word 'idea' or 'ideas'.
3. There are no significant differences between the distributions of macro-discourse markers across TED Talk genres.

Research Methodology

Corpora and Data Collection

The data in this study came from two corpora. One corpus (hereinafter TED-dm) is a compilation of 150 TED talk transcripts, 25 each from six popular topics: technology, entertainment, design, business, science, and global issues. This corpus is used to confirm the occurrences and frequency of macro discourse markers found in each type of talk. Another corpus is called TCSE, or the TED Corpus Search Engine, developed by Yoichiro Hasebe from Japan. This corpus has a large database of 2,589 TED talk transcripts with over 6,245,134 elements. The number of transcripts in this corpus is close to the number of talks available on the official TED site (around 2,700+ videos). For this reason, the corpus is considered as having data representative of TED talks generally and therefore is used in this study as a reference corpus.

Data Preparation

In order to compile the TED-dm, 25 transcripts from each of six talk genres were specified. This was done by using the browsing tools of the official TED site. The duration of talks was set for 12 to 18 minutes and the topics selected were from the six most popular genres. Once the duration and topics were set (Ex. *Technology, 12-18 minutes*), TED displayed a long list of videos starting from the present year (2018) to the first year in which TED talks were given (2006). Initially, the researcher intended to select the 25 most current videos from each genre list. However, this proved problematic as some videos appeared under more than one topic and therefore, it was not clear to which topic such videos should be assigned. For example, the talk *'What is it like to be a robot'* given by Leila Takayama, appeared under both technology and design. If this talk were assigned to either genre, we would not be able to generalize the distribution of macro discourse markers in these two different genres. Hence, any videos appearing with more than one of these six topics tags: technology, entertainment, design, business, science, and global issues, were not included in the study. Only the videos with a single tag from these categories were selected, despite the different year in which they were given.

Data Analysis

After all 25 transcripts from talks on the six topics were specified and collected, the two most recent transcripts from each topic were manually analyzed to identify the macro-discourse markers. Within one transcript, a speaker can use several macro discourse makers, but this study only looked for markers that introduced, expressed, or signaled ideas. For example, in the two excerpts below, the marker *'But I want to be clear'* was chosen for the study while the marker *'let me give you an example'* was not, because the latter told *'information'* rather than signaled an *'idea'*.

***"But I want to be clear:* postpartum hemorrhaging – it's not a Rwanda problem, it's not a developing-world problem – this is a global problem."**

Keller Rinaudo, 2017

“Let me give you an example. A couple months ago, a 24-year-old mother came into one of the hospitals that we serve, and she gave birth via C-section. But that led to complications, and she started to bleed.”

Keller Rinaudo, 2017

Following are the macro-discourse markers found in the two transcripts taken from each talk genre by manual investigation.

Technology	I think / if you want to / one thing that / the reason is / But I want to be clear / So picture that for a moment
Entertainment	I want you to picture this / I start to wonder / it turns out / I learned that / what if / I mean / And remember this.
Design	I believe / I'd like for you to just think about / Here's the thing
Business	So, three things that I have learned / I learned that / So I realized that / I've also discovered that / This is because / That's because / But my point is that
Science	The reason that...is / So the first thing that we discovered / you'll notice / that's why / Imagine if / what if / So in this case / Try to think about / So now we know that / So suppose now / But the question is / But the reason why ... is / we discovered that / so as you can imagine / If you think about
Global Issues	I'm going to focus on / We realized / Here's what I've learned

To ensure that these markers are in fact signaling ideas, one native speaker of English was asked to read the concordance lines in which these markers occurred and see whether he agreed with the list. Once the list was confirmed, the corpus TED-dm was used to identify the occurrences and frequency of these discourse markers, genre by genre.

One can notice, however, that some of the macro-discourse markers in the list above co-occurred with micro markers in front of them, such as, **and** *remember this*, **but** *my point is that*, and **so** *I realized that*. These micro markers are considered part of the macro

markers' variations and therefore were not included when searching the corpus. The complimentizer 'that' and possessive pronouns like 'my' were likewise excluded due to the fact that they too, were optional and subjective to the speakers. The example below illustrates when the marker 'but my point is' was investigated in the TED-dm topic-specific sub-corpora.

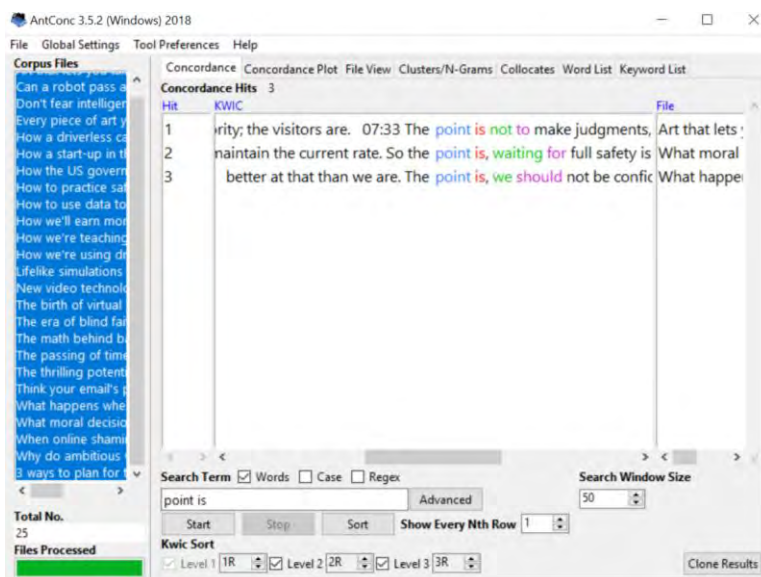


Figure 1: The result when searching 'point is' from the Technology sub-corpus of TED-dm using Antconcord program

The result shows that from the 25 TED transcripts under the topic 'technology', there were three occurrences of the marker 'The point is', one with the micro marker 'so' and two without it. The next step in the analysis was then to read each concordance line and see if the marker really introduced or signaled an idea. The frequency of instances where the markers introduced ideas was counted. This process was applied to every marker in the list and with every TED-dm sub-corpus ranging from technology to entertainment, design, business, science, and global issues. The TCSE corpus was also used to confirm the frequency of each marker in TED talks in general. This step marked the end of the first part of data collection.

In the second part, data collection no longer investigated the discourse markers mentioned above. As the study hypothesized that macro-discourse markers specifically containing the word ‘idea’ or ‘ideas’ must also occur, each TED-dm sub-corpus was searched using these two keywords. Once again, there were concordance lines to analyze and those with macro discourse markers signaling ideas were counted. For example, the data below shows instances with the word ‘idea’ in them but only numbers 12, 22, 26, and 29 were counted. This step in the investigation marked the end of all data collection.

Concordance Results 4:
 Concordance Hits 32

Hit	KWIC	File
1	in 1990 came to the Global Fund with a big idea -- a woman from Mexico by the name of Lucero	To solve the world's bigg
2	lly fit our 21st century values, not some archaic idea about bringing home the bacon, is long overdu	The new American Dream
3	nd Kerr not only listened, but he implemented the idea and then afterwards , gave U'Ren all the credi	5 ways to lead in an era o
4	ge. The second stage: remember, in isirika, every idea counts . Bridges have big posters and they hav	To solve the world's bigg
5	idges have big posters and they have nails. Every idea counts -- small or big counts. And third, isi	To solve the world's bigg
6	-doubt is paralyzing. It leads you to freeze. But idea doubt is energizing . It motivates you to test	The surprising habits of o
7	different kinds of doubt. There's self-doubt and idea doubt . Self-doubt is paralyzing. It leads you	The surprising habits of o
8	ld that go? 00:35 Bob, I've got this incredible idea for a completely new type of personal health	How to build a business tl
9	ership style, U'Ren felt comfortable bringing the idea forward . And Kerr not only listened, but he i	5 ways to lead in an era o
10	-Central. The late mayor of Jackson had a similar idea . He wanted to turn his entire city into a	When workers own comp
11	that day in his wheelchair. I just had no idea he was going to swim. I mean, where is	How to find work you lov
12	ll be the subject of future research. 12:49 The idea I want to leave you with is it's	What really motivates pec
13	how good ideas turn into great ideas, because no idea is born fully formed. It emerges a little bit	Forget the pecking order
14	12:09 Interestingly, it wasn't actually Kerr's idea . It was the idea of his 28-year-old assistant	5 ways to lead in an era o
15	ad. After all, America is deeply invested in this idea of economic transcendence , that every generat	The new American Dream
16	wasn't actually Kerr's idea. It was the idea of his 28-year-old assistant, Nick U'Ren. Bec	5 ways to lead in an era o
17	that turning yourself up is just not some perfect idea of how to be great, it's a way	A pro wrestler's guide to
18	's much easier to improve on somebody else's idea than it is to create something new from scrat	The surprising habits of o
19	organizational psychologist, this is the kind of idea that I test . So I challenged her to get	The surprising habits of o
20	that means cutting the demand and changing this idea that it's a victimless crime. Let's all	How fake handbags fund
21	ally hundreds or thousands. So I committed to the idea that my character wasn't going to be as	A pro wrestler's guide to
22	research, we look at this. We look at the idea that people have so-called protected values.	What really motivates pec
23	oss industries and ask people about their biggest idea , their most important suggestion, 85 percent	The surprising habits of o
24	one cup helps save the planet." I had no idea this plastic cup was so powerful. 04:09 (La	How to make a profit whi
25	that she doesn't think it's a good idea to play the woman card. It's just the	What it's like to be a worr
26	h, founding father of modern economics. His basic idea was that if everybody behaves in their own se	What really motivates pec
27	so many of us, when we have an important idea , we don't bother to try. But I have	The surprising habits of o
28	. But then my sister-in-law had a brilliant idea . What if I gave my daughter a choice? This	How to speak up for your
29	raging feminist. I mean, just to give you some idea: when I was five or six years old and	What it's like to be a worr
30	get their stock back. Even now, we have no idea where all that money went, to who it went	How fake handbags fund
31	so the customs officials would literally have no idea who had sent the products in the first place	How fake handbags fund
32	ys-on" transformation. 02:09 When I shared this idea with my wife Nicola, she said, "Always-on tra	5 ways to lead in an era o

Figure 2: The result when searching the keyword ‘idea’ from the Business sub-corpus of TED-dm using Antconc program

Findings and Discussion

Range of Macro Discourse Markers

The data from the study confirmed the first hypothesis which assumed that there is a wide range of macro discourse markers used by TED speakers to signal ideas. The first part of data collection found 31 macro discourse markers in 12 TED transcripts while the second part of the data collection discovered another 51 macro discourse markers containing the word ‘idea’ or ‘ideas’. The number of macro discourse markers identified in this study, then, was 82 in total.

When looking closely at all 31 macro discourse markers, some recurrent patterns can be found as illustrated in the table below.

Subject + Verb	Verb	If Hypotheses	Noun + Verb to be	Other constructions
I think	Remember,	If you want to...	The reason is	One thing
I realized	Imagine (if, that)	What if...	The point is	Here’s the thing
I learned	Think about it.	If you think about...	The question is	This is because
I mean	Suppose (now, that)			That is because
I believe	Picture (that)			I want to be clear
I start to wonder				I want you to think
I discovered				I want to focus
I’ve learned				It turned out
We know				You’ll notice
We discovered				
We realized				

Table 1: Common patterns of 31 macro discourse markers found in the first part of data collection

Some epistemic lexical verbs (ELVs) found frequently in the TED_ac corpus of Caliendo and Compagnone (2014) occurred in this data as well. Those ELVs are, for example, *think*, *know*, *believe*, *imagine*, *notice*, and *suppose*. In their study, Caliendo and Compagnone investigated the co-occurrences of these verbs with the pronoun ‘You’, ‘I’ and ‘We’ and then retrieved lists of possible

clusters in which these ELVs occurred. This study, however, selected only clusters that function as discourse markers. Examples of such clusters are: *I think, I believe, we know, and you'll notice.*

However, two of their most frequent ELVs: *see* and *show*, were missing from this study. This is possibly because the two verbs, when used as discourse markers, tend to deliver information rather than ideas. For example;

***"I want to show you** some brave souls who've had the courage to embrace this advanced potty-training approach."*

***Now, I want to show you one last thing.** I call this a Life Calendar. **As you can see,** the system is very complex and vast. **And what you see here** in this methylation mapping, a dramatic separation at birth of the animals that had a high social status versus the animals that did not have a high status.*

These four examples support the idea that the macro discourse markers in which the verbs 'show' and 'see' occurred, normally deliver information rather than ideas to the listeners. Nevertheless, this is not to imply that the epistemic lexical verbs found in discourse markers in this study *always* signal ideas. Some markers have more than one function and they too can signal information as well as ideas. They will be discussed further in a later part of the study.

Macro Discourse Markers Across Genres of Talks

Once the list of 31 macro discourse markers had been obtained, the TED-dm sub-corpora were investigated to examine the distribution of markers across genres of talks. The results were as follows:

No	Macro Discourse Markers	Technology	Entertainment	Design	Business	Science	Global Issues	Total
<i>Macro DMs from 2 Technology Transcripts</i>								
1	I think (that)	51/57	29/32	51/59	24/27	38/44	35/39	228
2	If you want to If you wanted to	7/7	1/1	3/3	7/7	4/4	2/2	24
3	one thing (that) There's one thing	1/1	2/2	2/2	1/2	1/1	1/1	8
4	(And) the reason is (that) the reason... is (that) the reason... is this.	10/11	5/5	4/4	3/3	2/3	4/4	28
5	(But) I want to be clear Now to be clear	1/1	1/1	-	1/1	-	-	3
6	(So) picture that for a moment I want you to picture	1/1	1/1	-	-	-	-	3
<i>Macro DMs from 2 Entertainment transcripts</i>								
7	I start to wonder I started to wonder	-	1/1	-	1/1	-	-	2
8	It turns out	5/11	2/3	2/2	1/9	6/10	1/2	17
9	I learned that	-	1/1	-	1/1	2/2	2/2	6
10	What if	19/19	9/9	11/11	5/5	15/17	1/1	60
11	I mean	3/10	13/23	1/5	4/12	4/13	3/3	28
12	And remember this (But) remember...	1/1	2/3	1/2	1/3	1/3	4/6	10
<i>Macro DMs from 2 Design transcripts</i>								
13	I believe (that)	11/11	3/3	12/12	7/7	3/3	17/17	53
14	I'd like you to think I want you to think	-	-	1/1	-	2/2	-	3
15	Here's the thing	-	2/2	1/1	-	1/1	1/2	5
<i>Macro DMs from 2 Business transcripts</i>								
16	I have learned Here's what I have learned	0/2	-	1/1	2/2	1/1	1/1	5
17	I realized (that)	1/1	8/14	1/1	4/8	2/8	1/8	17
18	I've discovered (that) I discovered (that) What I discovered is that	1/1	3/3	0/1	3/3	-	0/1	7
19	This is because	-	-	1/3	1/1	0/1	1/1	3
20	That's because	-	1/1	1/1	4/4	1/3	0/1	7
21	(But) my point is (that) The point is,	2/2	-	-	4/4	2/2	2/2	10
<i>Macro DMs from 2 Science Transcripts</i>								
22	(So, the first thing that) we discovered What we discovered	-	-	-	-	1/2	2/4	3
23	(And the second thing that) you'll notice You'll notice here (that)	0/1	-	0/1	1/1	1/3	0/1	2

24	(but) imagine if Imagine that Imagine this Let's imagine Just imagine that You can imagine I just want you to imagine I want you to first imagine	12/12	8/9	17/23	6/9	13/14	6/6	62
25	Think about it. Think about it this way.	2/2	-	1/1	2/2	3/3	1/2	9
26	If you think about (it)	2/2	1/1	2/3	2/2	2/2	-	9
27	So suppose now Suppose	3/3	-	-	-	1/1	-	4
28	(Now) we know that	3/3	-	2/3	1/2	3/3	1/2	10
29	(But) the question is	4/4	1/1	3/3	1/1	6/6	6/6	21
<i>Macro DMs from 2 Global Issues Transcripts</i>								
30	I'm going to focus I want to focus	-	-	0/1	-	1/1	1/1	2
31	We realized (that) We realized something...	1/2	-	-	-	1/1	2/2	4

Table 2: Distribution of 31 macro discourse markers across different genres of talks

The data from Table 2 shows that there are occurrences of most macro discourse markers in each type of talk but not all of the occurrences introduced ideas. The number 4/8, for example, illustrates that there were eight occurrences of the marker '*I realized*' found in Business talks, but only four of them signaled ideas.

In sum, 12 of the 31 markers occurred across six TED talk genres: *technology*, *entertainment*, *design*, *business*, *science*, and *global issues*. Another 19 markers occurred in 5 genres, 4 genres, 3 genres, and 2 genres respectively. No marker occurred in only one genre.

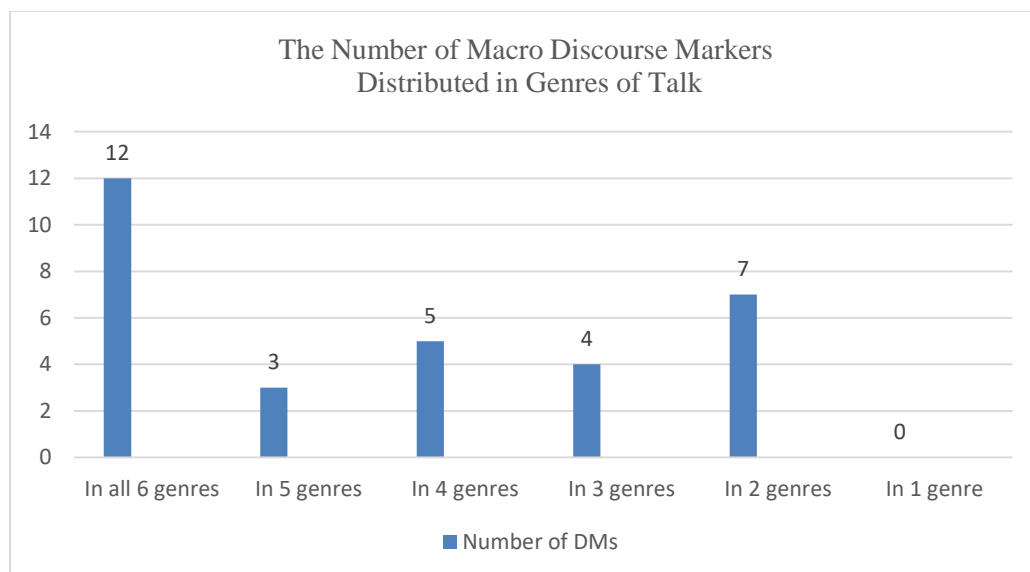


Figure 3: Number of macro discourse markers distributed across six TED talk genres

For more details, these are the explicit lists of markers and the number of TED talk genres in which they occurred.

Distribution	Macro Discourse Markers	Number
In 6 Genres	I think / if you want to / one thing / the reason...is / it turns out / what if / I mean / remember / I believe / I realized / imagine / the question is	12
In 5 Genres	Think about it / if you think about / now we know that	3
In 4 Genres	I learned that/ here's the thing / that's because / I have learned / the point is	5
In 3 Genres	I want to be clear / I discovered / this is because / we realized /	4
In 2 Genres	I want you to picture / I started to wonder / I want you to think about / we discovered / you'll notice / so suppose now / I want to focus	7
In 1 Genre	-	-

Table 3: The list of macro DM categorized by the number of talk genres in which they occurred

It can be noticed from Table 3 that 20 macro discourse markers or 64 % of all markers occurred in 4 to 6 genres of talk. The researcher views this number as representing a wide distribution of macro discourse markers across genres. However, 11 markers occurred in only 2 or 3 talk genres, and further explanation for their low rate of distribution is needed. Is it because these macro discourse markers are specific to those talk genres, or

is there another underlying reason? To investigate, the reference corpus, TCSE or TED Corpus Search Engine, was used to identify the frequency of markers in TED talks more generally. The results appear in the tables below:

6 Genres		3 Genres		2 Genres	
I think	5,377	I want to be clear	32	I want you to picture	17
If you want to	350	I discovered	170	I started to wonder	12
One thing	618	This is because	49	I want you to think about	39
The reason is	880	We realized	123	We discovered	98
It turns out	734			You'll notice	74
What if	725			So suppose now	29
I mean	1,735			I want to focus	30
Remember	61				
I believe	781				
I realized	490				
Imagine	642				
The question is	243				

Table 4: The distribution of macro discourse markers on TCSE

The data implies that the low distribution rate of macro discourse markers in this study is subject to their infrequency in TED talks in general. The marker '*I started to wonder*', for example, occurred only 12 times in the TCSE and therefore it is not unexpected that it is not distributed across genres in the corpus of this study. This then confirms that the low distribution rate of certain discourse markers is not because they are specific to talk topics, but rather because they themselves occur infrequently in TED talks generally.

Another reason that may explain why some macro discourse markers are distributed across only two or three genres is that they performed functions other than only signaling ideas. The marker '*You'll notice*', as illustrated in Table 2, for example, occurred in five different genres of talks, but it functioned as an idea signaler in only two genres. That was why it was categorized in the low distribution group.

As 64 % of the macro discourse markers in this study were distributed across genres and the remaining 36 % of the markers

were not specific to talk topics, the hypothesis was confirmed that there are no significant differences between the distributions of macro-discourse markers across TED Talk genres.

Macro Discourse Markers Containing the Words ‘idea’ or ‘ideas’

The second part of the data collection was done by searching the keywords ‘idea’ and ‘ideas’ in the TED-dm sub-corpora to identify macro discourse markers that explicitly signal ideas to the listeners. The markers found in each genre of talk are as follows:

No.	Macro discourse markers	Frequency
1	Let’s shift our ideas about...	1
2	And today the idea behind this speech...	1
3	One of our recent ideas...	1
4	Here’s the idea	1
5	This next idea comes from...	1
6	This is based on an idea...	1
7	And that’s the idea	1
8	And that is the idea I’m trying to illustrate	1
9	And so my first idea is	1
10	And the idea is quite simple	1
11	And the basic idea is that	1
12	Which actually is, the idea of my talk	1
13	In hindsight, this idea	1
14	This is a bit of a strange idea	1
15	It’s an idea that	2
16	It’s based on the idea that	1

Table 5: Macro discourse markers found in 25 Technology talks

No.	Macro discourse markers	Frequency
1	And I had this idea:	1
2	The idea is...	1
3	But my today idea is	1
4	My idea today is to tell you that...	1
5	The idea was to...	2
6	That would be an idea you might take with you.	1
7	Thank you for letting me presents my idea about...	1
8	My idea today is to tell you that	1

Table 6: Macro discourse markers found in 25 Entertainment Talks

No.	Macro discourse markers	Frequency
1	I really love this idea.	1
2	and so the idea is that	1
3	And the idea is to	1
4	The idea is very, very simple.	1
5	But there's even another idea	1
6	all based on this one idea	1
7	I think we need to question this idea that	1
8	They are challenging the idea that	1
9	There's an idea that	1
10	So the idea was	1
11	So the basic idea was	1
12	The second idea worth questioning is	1

Table 7: Macro discourse markers found in 25 Design talks

No.	Macro discourse markers	Frequency
1	The idea I want to leave you with is...	1
2	We look at the idea that...	1
3	His basic idea was that...	1
4	just to give you some ideas:	1

Table 8: Macro discourse markers in 25 Business Talks

No.	Macro discourse markers	Frequency
1	Think about that idea for a second.	1
2	The idea is,	2
3	And the idea is that,	1
4	And this is the whole idea of	1
5	because there's this idea that	1

Table 9: Macro discourse markers found in 25 Science talks

No.	Macro discourse markers	Frequency
1	I want to give you four ideas as a starting point.	1
2	and one of the ideas	1
3	So here's an idea.	1
4	The idea behind this law is that	1
5	The fourth and final idea I want to put forward is	1
6	We have a research idea that we think is perfect for you.	1

Table 10: Macro discourse markers found in 25 Global Issues talks

Obviously, all of these macro discourse markers highlighted TED speakers' intention to deliver ideas. Distinctions between the 51 markers were the word choices of the speakers themselves which varied in terms of tenses, lexical verbs, adjectives, and co-occurring micro markers.

Most of the markers were used more commonly in present tense than in past tense to signal ideas that are ‘new’ to the listeners. The lexical verbs that co-occurred with the word ‘idea’ or ‘ideas’ were *shift, based on, illustrate, tell, give, leave, look at, come from, question, present, and put forward*. The markers also varied in the choice of adjectives which co-occurred with the keyword ‘*idea (s)*’, for example, *final, basic, whole, simple, strange, recent, and worth sharing*. Some TED speakers used micro markers with their macro markers as well, for instance, *and, and so, so, and but*. These four main features: tenses, lexical verbs, adjectives and macro markers, contributed to a wide variety of macro discourse markers that signal ideas.

Another hypothesis was therefore confirmed that there are many macro-discourse markers that contain the word ‘idea’ or ‘ideas’ found in the spoken discourse of TED Talks.

Conclusion

The results of this corpus-based study identified two sets of discourse markers used by TED speakers to signal ideas. The first set comprises 31 macro discourse markers of which 64% occurred across all six popular TED talk genres. Although the other markers were not distributed likewise across every genre, they were not specific to a certain topic of talk. Therefore, it can be concluded that these macro discourse markers are available for use in all six talk topics: technology, entertainment, design, business, science, and global issues.

The second set of discourse markers found 51 various ways that speakers can express ideas to listeners. With markers made explicit by the addition of the word ‘*idea*’ or ‘*ideas*’, listeners cannot fail to recognize that an ‘idea’ is being passed to them from the speaker. The variations in discourse markers in this group are subject to the speakers’ choice of tense, lexical verb, adjective, and co-occurring micro-marker. One speaker, for example, might use ‘*so the basic idea was*’ while another speaker may use ‘*the next idea comes from*’, but both have the same underlying goal; that is, to deliver an idea to the listener.

The results of this study highlight the fact that there is a wide range of discourse markers used in academic spoken discourse to signal ideas. Some macro discourse markers appearing in TED talks might resemble the discourse markers found in lectures (e.g. *I believe that, remember, the reason is*) by Morell (2000), and Chaudron and Richards (1986), but several are new contributions to the list. One more observation is that, the large number of markers that contain the word 'idea' or 'ideas' is particularly noteworthy in TED Talk presentations since the main goal of TED speakers is to deliver ideas to listeners.

Pedagogical Implications

The list of macro discourse markers identified in this study can be beneficial in a pedagogical context. Teachers may encourage their students in a public speaking course to express ideas more clearly and effectively using these discourse markers. Students who need to deliver a speech in public may use some of these markers to sound more natural when discussing ideas. The marker 'So, here is an idea.', for example, can be used to signal that a talk has reached a significant point or that the speaker has a solution to offer. There will be less overproduction of the marker 'I think' as well when the students learn to share the role of thinking with their audiences by engaging them with discourse markers like 'if you think about', 'imagine', and 'picture that for a moment.'

Although the list of macro discourse markers in this study was derived from the highly monologic talks of TED speakers, it is noteworthy that the list is applicable to other interactive activities like group discussions and debates, as long as they involve the act of expressing or exchanging ideas. For example, the marker 'I think we need to question the idea that... ' can help students to raise their argument in a debate and the marker 'the point is' can be helpful in group discussions.

Nevertheless, to my knowledge, there are no explicit lessons or activities in Thai schools that specifically focus on macro discourse markers in spoken discourses. Students often learn the cohesive expressions, such as, *however, on the other hand,* and

what is more, in essay writing but are seldom taught to recognize the role of discourse markers in speaking. It is therefore important for teachers to start emphasizing the use of discourse markers as they contribute to the naturalness of students' language use. The macro discourse markers identified in this study can serve as an initial list. Teachers could consciously use these markers themselves in the classroom when their goal is to effectively deliver ideas to their students. They can also include these markers explicitly in lessons and activities like group discussions, debates, and presentations.

Limitations and Suggestions for Future Research

There are three principal limitations to this present study. First, it investigated a preliminary set of discourse markers found in only 150 TED talk transcripts. A larger number of TED Talk transcripts would provide more representative data and allow for more concrete generalizations.

Second, the frequency number of discourse markers found in the TCSE as shown in Table 4 is an estimate. The researcher cannot claim with certainty that the discourse markers of this study signaled ideas in all those occurrences. While the marker '*One thing*', for example, occurred 618 times in TCSE, some of these might as well introduce information. Therefore, the only assumption the researcher could make was: the higher the number of DMs occurring in TCSE, the greater the tendency for them to function as idea signalers. Future researchers might want to limit the number of discourse markers they wish to study so all occurrences in TSCE could be reasonably analyzed.

Finally, the decision to consider a marker as an idea signaler was based on the judgment of the researcher alone. The result would be more reliable if raters helped to confirm the data.

The Authors

Kanokrat Uicheng (kanokrat.uc@gmail.com) is an English teacher at Deebukphangnga Wittayayon School, Phang-nga, Thailand. She graduated with a B.A. in English from Silpakorn

University and is now pursuing her master's degree in English at the Faculty of Arts, Chulalongkorn University. Her field of study is in corpus linguistics and discourse analysis.

Michael Crabtree (crabtree@chula.ac.th) is a member of the English Department at the Faculty of Arts, Chulalongkorn University, where he teaches Thai-English translation, research methodology, advanced composition, and contemporary fiction.

References

- Bellés, B. (2004) *The Spoken Academic Discourse of Social Sciences: Discourse Markers within the University Lecture Genre*. Unpublished Master Thesis. Castellón (Spain): Universitat Jaume I
- Caliendo, G., & Compagnone, A. (2014). Expressing epistemic stance in university lectures and TED talks: a contrastive corpus-based analysis. *Lingue E Linguaggi*, 11, 105-122. doi:10.1285/i22390359v11p105
- Carter, R., & McCarthy, M. (2006). *The Cambridge grammar of English: A comprehensive guide*. Cambridge: Cambridge University Press.
- Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, 7(2), 113-127. doi:10.1093/applin/7.2.113
- Dülger, O. (2007) Discourse markers in writing. *Selçuk University, Sosyal Bilimler Enstitüsü Dergisi*, 18 (2007), pp. 257-270
- Flowerdew, J., & Tauroza, S. (1995). The effect of discourse markers on second language lecture comprehension. *Studies in Second Language Acquisition*, 17(04), 435. doi:10.1017/s0272263100014406
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6(2), 167-190.
- Jalilifar, A. (2008). *A Comparative Study of Message Functions in Written and Spoken Reported Speech Produced by Iranian EFL Students*. *Systemic Functional Linguistics in Use*.
- Lai, Y., & Lin, Y. (2012). Discourse markers produced by Chinese-speaking seniors with and without Alzheimers disease.

- Journal of Pragmatics*, 44(14), 1982-2003.
doi:10.1016/j.pragma.2012.09.002
- Landone, E. (2012). Discourse markers and politeness in a digital forum in Spanish. *Journal of Pragmatics*, 44(13), 1799-1820. doi:10.1016/j.pragma.2012.09.001
- Lee-Goldman, R. (2010). No as a discourse marker. *Journal of Pragmatics*, 1-46.
- Morell, T. (2000). EFL content lectures: A discourse analysis of an interactive and a non-interactive style. Working papers 7. Alicante, Spain: Departamento de Filología Inglesa, Universidad de Alicante.
- Rezanova, Z. I., & Kogut, S. V. (2015). Types of discourse markers: their ethnocultural diversity in scientific text. *Procedia - Social and Behavioral Sciences*, 215, 266-272. doi:10.1016/j.sbspro.2015.11.633
- Rhee, S. (2013). "I Know I'm Shameless to Say this": Grammaticalization of the Mitigating Discourse Marker Makilay in Korean. *Procedia - Social and Behavioral Sciences*, 97, 480-486, doi.org/10.1016/j.sbspro.2013.10.262.
- Rido, A. (2015). The use of discourse markers as an interactive feature in science lecture discourse in L2 setting. *TEFLIN Journal*, 21(1).
- Roggia, A. (2012). Eh as a polyfunctional discourse marker in Dominican Spanish. *Journal of Pragmatics*, 44 (13), 1783-1798. doi.org/10.1016/j.pragma.2012.08.010.
- Schiffrin, D. (2001). Discourse markers: Language, meaning and context. In: Schiffrin, D., Tannen, D. & Hamilton, H. E. (Eds.), *The Handbook of Discourse Analysis*, Blackwell, Malden, MA, pp.54-75.
- Schiffrin, D. 1987a *Discourse Markers*. Cambridge: Cambridge University Press.
- Tanghe, S. (2016). Position and polyfunctionality of discourse markers: The case of Spanish markers derived from motion verbs. *Journal of Pragmatics*, 93, 16-31 doi:10.1016/j.pragma.2015.12.002

- Tay, D. (2011). Discourse markers as metaphor signaling devices in psychotherapeutic talk. *Language & Communication*, 31(4), 310-317. doi:10.1016/j.langcom.2011.02.001
- Tsai, T.J. (2015). Are you TED talk material? comparing prosody in professors and TED speakers. *INTERSPEECH*.
- Zhang, M. (2000). Cohesive features in the expository writing of undergraduates in two Chinese universities. *RELC Journal*, 31 (1), 61-95