Journal of Modern Research in English Language Studies
Print ISSN: 2676-5357
Online ISSN: 2676-5985

Unveiling the Argumentative Nature of Meta-Analysis in Applied Linguistics: An Argument-Mining Approach

Arsalan Yaghoubi[1], Mohammad Amini Farsani[2*], Behrouz Minaei-Bidgoli[3], Ali Asghar Taghizadeh[4]

[1]MS Student of Computational Linguistics at Brandeis University, USA, arsalanyaghoobi@brandeis.edu

[2*]Assistant Professor in Applied Linguistics at Iran University of Science and Technology, Tehran, Iran, mohammad_farsani@iust.ac.ir

[3]School of Computer Engineering, Iran University of Science & Technology, Hengam Ave, Narmak, Tehran, Iran,b_minaei@iust.ac.ir,

[4]School of Computer Engineering, Iran University of Science & Technology, Hengam Ave, Narmak, Tehran, Iran, taghizade@iust.ac.ir

Article Info

Article Type:

Research Article

Received:

31/05/2023

Accepted:

06/08/2023

ABSTRACT

Despite paradigmatic research advancements and movements in applied linguistics, the issue of rhetoric, which serves as one of the fundamental pillars of each paradigm, remains largely unaccounted for. Considering the commensurability of argumentation and meta-analysis, coupled with the increasing rate of meta-analytic studies in the field of applied linguistics, there arises a need to examine the argumentation behavior of applied linguistics' meta-analysts. As such, following research synthesis techniques and an argument-mining approach, we examined the academic argumentation genre of meta-analysis published in leading applied linguistics journals through argument-mining techniques in light of the modified Toulmin framework proposed by Qin and Karabacak (2010). The current study, employing the modified Toulmin framework, examined the argumentative writing components represented in the introduction section of 54 meta-analytic studies published in leading journals of applied linguistics through argument-mining techniques. Our findings highlight the complexity and argumentativeness of the meta-analysis genre. We further found that the Modified Toulmin Model is implementable for the task of argument mining, which can have a great impact on argumentation, meta-analysis, and argumentative academic writing. Implications and recommendations for academic argumentative writers and meta-analyzers are discussed.

Cite this article: Yaghoubi. A., Farsani, M. A., Minaei-Bidgoli, B., & Taghizadeh, A. S. (2023). Unveiling the Argumentative Nature of Meta-Analysis in Applied Linguistics: An Argument-Mining Approach. *Journal of Modern Research in English Language Studies, 10*(4), 135-162.

DOI: 10.30479/jmrels.2023.18873.2222

1. Introduction

The field of applied linguistics (AL) has recently witnessed growing attention to the strand of research methodology and its related issues in many different avenues and paradigms (Amini Farsani & Abdollahzadeh, 2019; Byrnes, 2013). Such paradigmatic research advancements are more highlighted given the function of meta-researchism, which addresses the quality of quantitative, qualitative, mixed-methods, and research synthetic studies (Amini Farsani et al., 2021). Such movement is further warranted because applied linguists are living in "a golden age of applied linguistics research" (McKinley, 2020, p. 1). As Plonsky and Oswald (2015) put it, progress in applied linguistics research is contingent on "sound research methods, principled data analysis, and transparent reporting practices" (p. 325). Despite such appealing movements in the field, what is almost unaccounted for is the issue of rhetoric as one of the fundamental pillars of each paradigm.

Furthermore, the multidisciplinary field of applied linguistics is in lively dialogue with different hard and soft disciplines. Such dialogue and interaction, as Amini Farsani et al. (2021) assert, are more established in *Educational Studies* (38.35%), *Life Sciences* (33.44%), *Physical Sciences* (13.88%), *Social Sciences* (4.1%), and *Arts & Humanities* (1.64%), respectively. As such, the recent empirical document represents the need for communication through sound arguments to support or refute the ideas. Accordingly, there is a strong demand for employing "sophisticated rhetorical moves" in academic rhetoric (Graff & Birkenstein, 2010, p. 3) and cultivating sound argumentative skills. In McKinley's (2020) terms, we should move beyond "types of research" towards "highly impactful research" (p. 2).

One such strategy to adhere to the untaken path projected by McKinley (2020) is to orient applied linguists towards the skill of argumentation, which is characterized as:

> "a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge'', has been used as a guideline to plenty of studies." (van Eemeren et al., 1996, p.5)

Argumentative writing, as the lion's share of academia, has notably grabbed the attention of different scholars studying argumentation skills. In applied linguistics, most of the research projects have centered their claims on argumentative rubrics proposed by Toulmin (2003) and others (Qin & Karabacak, 2010; Stapleton & Wu, 2015). However, the unit of analysis was argumentative essays written by different students in the above studies. For example, Abdollahzadeh, Amini Farsani, and Beikmohammadi (2017), following a modified Toulmin framework, examined Master of Arts (MA) students' argumentative behavior through examining students' writing tasks in their advanced academic writing courses.

1.1. Argumentation and Meta-analysis

One of the registers that exhibits a high level of argumentation in academic discourse is related to meta-analysis. Meta-analysis narrowly includes "the averaging of effect sizes across a set of primary studies in a given research area". Broadly, it consists of "an entire set of procedures designed to yield a view of the domain in question that is more objective, transparent, and systematic than traditional literature review" (Loewen & Plonsky, 2016, p. 112). This genre needs a strong mode of reasoning and persuasiveness when it comes to arguing the (in)effectiveness of L2 problems in the literature (meta-analyzing pragmatics; meta-analyzing L2 corpus; meta-analyzing oral and written feedback; see also Anani Sarab and Amini Farsani (2023) for a detailed topic covered in meta-analyses in applied linguistics (AL). Within meta-analyses, the introduction and discussion subgenres need a highly competent researcher to (counter)claim, support, and refute the assertions (Plonsky, 2013, 2017). Accordingly, meta-analysts should put themselves into a community of others in which argumentation is key (Hoey & Winter, 1983).

Given the commensurability of argumentation and meta-analysis (see Melendez-Torres et al., 2017), coupled with the growing prevalence of meta-analytic studies in AL (Amini Farsani & Babaii, 2018; Amini Farsani et al., 2021; Plonsky, 2017), it seems that examining the argumentation behavior of applied linguistics' meta-analysts is needed. This is more notable in the introduction section, in which meta-analysts need to support or refute their assertions with sound reasons. Such alignment is more warranted given the recent call for studies on big data and data mining in the field of applied linguistics (Warschauer et al., 2019).

Unlike the previous studies that primarily focused on academic writings or persuasive essays produced by different writers in different contexts (Qin & Karabacak, 2010; Stapleton & Wu, 2015), we employed a secondary study dataset, namely meta-analyses. According to Rapanta, Garcia-Milla, and Gilabert (2013), argumentative competence refers to a "group of skills mainly investigated in both students (and especially adolescents) and teachers". This set of skills can be represented in "discourse forms, in the use of specific strategies, or as the fulfillment of an argumentation goal in a particular context" (p. 512). Meta-analysis can be considered a discourse form to help applied linguists shape and enhance their argumentative competence. It is to be noted that the meta-analysts are professionals in the field who have already internalized argumentative competence. Because of the big data and voluminous information presented in the introduction section—as in most argumentative genres of meta-analysis—we followed the argument mining techniques to address the argumentative behavior of meta-analysts in light of the recently-developed modified Toulmin framework (see Qin & Karabacak, 2010). This modified theoretical framework consists of six argument discourse units (ADUs): claim, data claim, counterclaim, data counterclaim, rebuttal, and data rebuttal.

Surveying argumentative skills through the lens of computational linguistics brings about the nascent field of argumentation mining, a strand that has received recent attention in various disciplines. Studies conducted through argument-mining techniques in different disciplines with different units of analysis (e.g., persuasive essays, microblogs, and online product reviews) could be categorized into three groups: (a) those studies that examined arguments versus non-arguments (e.g., Ajjour et al., 2017); (b) those studies that investigated claim, premise, and non-arguments (e.g., Aker et al., 2017); and (c) those studies that broaden the issues to consider major claims, minor claims, and premises (e.g., Al-Khatib et al., 2016). Aside from the non-argument-mining literature, the argument-mining literature highlights the inadequacies of the Toulmin model in different studies. Although some researchers argued for the feasibility of the model in everyday and usual argumentation (Qin & Karabacak, 2010), some others asserted that the Toulmin model might not be useful for describing real-life argumentative texts or complex ones (Ball, 1994). Concomitantly, the issue of quality has been documented as a hotly debated line of research in the argumentative mining field. That is, researchers utilize argument mining "to be able to extract, assess, and even produce argumentation quality; however, what is missing is what model of argumentative structure proves most suitable" for operationalizing quality (Wachsmuth et al., 2018, p. 1689).

Nevertheless, no empirical studies have been conducted in the literature to simultaneously examine the academic argumentation genre of meta-analysis published in AL journals through argument-mining techniques in conjunction with the modified Toulmin framework proposed by Qin and Karabacak (2010). Accordingly, the current study, employing the modified Toulmin framework, examined the argumentative writing components represented in the introduction section of 50 meta-analytic studies published in leading journals of applied linguistics through argument-mining techniques. The following research questions are addressed:

1. Is the modified Toulmin framework well represented by the argumentative components in meta-analytic studies through the argument-mining technique? If yes, what is the direction and strength of such a model?

2. What is the overall argumentative structure of the introduction section of the meta-analysis written by applied linguists?

The first research question, i.e., the main question of the study, will be presented based on the interface of artificial intelligence, data mining, and L2 learning. The next research question is more concerned with the contribution of new datasets (AL meta-analyses) and their contributions to the fields of computational linguistics, computer sciences, and artificial intelligence.

1.2. Theoretical Underpinnings

This study is prompted by the three theoretical frameworks of argumentation, argument mining, and meta-analysis. From the argumentation perspective, we applied the modified Toulmin model developed by Qin and Karabacak (2010). We used the modified version of the Toulmin model for two reasons: First, the traditional model of Toulmin does not present a satisfactory framework for different genres in science. Second, this model does not consider other contextual features like mode of presentation and soundness (see Stapleton & Wu, 2015).

From the perspective of argument mining, we leveraged Artificial Neural Networks, specifically employing transformers (Bert and Roberta) and the Bi-LSTM model, for the sequence labeling task of argument mining. Bert (Roberta), or in a more general sense, transformers, are models that use self-attention that makes training on long sequences possible and takes the context for each vocabulary into account as well. The whole training procedure was summarized in the black box of the transformer.[1]

From a meta-analysis perspective, we adhered to Luke Plonsky's ideas (Plonsky, 2013, 2014, & 2017). It is an approach that focuses on primary research findings and has the purpose of integrating past research through generalization from many separate investigations that address the same hypotheses (Loewen & Plonsky, 2016). This genre is considered the most argumentative genre, which makes it the best potential source for argumentation and argument mining.

2. Literature Review

In the following paragraphs, we review the related studies in light of two grand themes: (a) those studies that examined argumentative writing in applied linguistics (see Table 1), and (b) those related studies inspired by argument-mining techniques in different disciplines [2](see Table 2). The related studies section covers these two lines of research in the following sections.

---

[1] The black box of the transformers are the hidden layers that handles the calculations for the language model automatically.

[2] Applied linguistics, computer science, computational linguistics, and research synthesis

2.1. Related Studies on Argumentative Writing in Applied Linguistics

Qin and Karabacak (2010) projected a context-specific modified Toulmin model to account for the limitations of the Toulmin model. The focus of the study was to analyze one hundred and thirty-three argumentative writing essays produced by English-major Chinese students. The essays were analyzed in terms of surface structure and quality or soundness of written arguments. The results revealed that although Chinese students were almost successful in producing structurally sound argumentation, their performance was far from complete in creating persuasive arguments.

Employing the modified Toulmin framework, Stapleton and Wu (2015) examined the soundness of arguments crafted by high school writers in Hong Kong. In this study, 125 students were asked to produce argumentative essays applying the modified Toulmin model. Forty-six Ph.D. students were called on to assess the six outstanding essays based on their surface argumentative structure. The results, lending support to Qin and Karabacak's (2010) assertions, revealed that students failed to provide adequately sound data or reasons for the claims.

Given that L2 argumentative writing is highly context-specific, Abdollahzadeh et al. (2017) examined 150 argumentative essays produced by graduate L2 learners in an Iranian EFL context. They adhered to the modified Toulmin model to examine the surface structure and quality of written arguments produced by Iranian MA students. The results revealed that the primary elements of argumentation (i.e., claim and data) were more represented than the secondary components (i.e., counter-arguments; rebuttals). The results further supported the above studies, signifying a lack of adequate attention to the persuasiveness and quality of written argumentation.

Besides EFL contexts, Osman and Januin (2021) surveyed the argumentative behaviors of ESL writers in Malaysia. Unlike the previous studies, they examined the structure of persuasive writings by ESL writers using the Toulmin model. The results, consistent with the above-mentioned studies, revealed that the learners applied all different elements of the Toulmin model except for rebuttal and data rebuttals (two quality criteria). Overall, the studies reviewed above highlight two important phases of argumentation, including surface and soundness layers, in different contexts with different samples (see Table 1). As such, besides providing profound insights for L2 argumentative writers, they emphasize the flexibility and dynamicity of the modified argumentative model developed by Qin and Karabacak (2010). However, labeling and annotation procedures in all the aforementioned studies were handled manually. Their focus was also on students' argumentative essays produced during or out of class (i.e., timed vs. untimed argumentative writing). Furthermore, the data set of the studies consisted solely of all written essays produced by L2 learners with different language competencies, overlooking other corpora such as academic articles and secondary studies.

Table 1.

A Profile of Studies Following the Modified Toulmin Framework

Year	Author	Topic	Data-driven Analysis	Theoretical Framework	Implications
2010	Qin & Karabacak	Surveying argumentative writing in an EFL context	Persuasive essays of academia	Modified Toulmin	Teaching Learners how to write argumentative text by means of both primary and secondary Toulmin elements
2015	Stapleton & Wu	Poring over the soundness of arguments in students' persuasive writing	Persuasive essays of high school students + Persuasive essays of the academic context	Modified Toulmin	The centrality of examination beyond the generic features of argumentative structure and the significance of the substance and quality
2017	Abdollahzadeh et al.	Surveying argumentative writing in an EFL context	Argumentative essays of graduate learners	Modified Toulmin	Even at advanced levels, the argument structure does not guarantee argumentative substance.
2021	Osman & Januin	Investigating ESL persuasive essays employing Toulmin’s model of argument	15 persuasive essays written by tertiary learners	Toulmin	Although 15 ESL writers found out the appropriacy of Toulmin’s model, researchers observed that the rebuttal element was not visible in the dataset.

2.2. Related Studies on Argument Mining

We found no related studies in AL. Accordingly, we reviewed some of the most relevant research projects disseminated in different disciplines:

Adopting argument mining techniques, Stab and Gurevych (2017) explored 402 persuasive English essays retrieved from a database named ‘essayforum.com’. The students produced their argumentative essays in response to the topic ‘Competition or Cooperation?’ The final corpus consisted of 77, 116 sentences with 147, 271 tokens. Their purpose was to project a new model for parsing argumentation structures in persuasive essays. As such, they concentrated on two macro levels: (1) the level approach, encompassing minor claims, major claims, and premises; and (2) The claim approach, comprising two labels: claim(s) and premises. An end-to-end method in which all the processes of training and testing are handled automatically using a deep learning approach was followed. The findings revealed that their identification model yields good accuracy for argument component extraction; however, this reported accuracy varies from text to text in light of different levels of argumentation.

Simultaneously, Eger et al. (2017), taking an argument-mining approach, explored neural techniques for end-to-end argumentation mining. They used the same dataset reported in Stab and Gurevych’s (2017) study. They framed argument mining as a dependency parsing and token-based sequence-tagging problem. The findings revealed that the dependency parsing approach outperformed the token-based sequence tagging approach. Furthermore, an end-to-end computational argument-mining model outperforms the previous models of argument mining.

Nguyen and Litman (2018) examined two persuasive-essay datasets: (1) the ASAP dataset, consisting of eight essay sets produced by students in grades 7-10; and (2) the TOEFL dataset, comprising more than 8,000 essays written by non-native test takers. Initially, they distinguished between argumentative and non-argumentative essays. Subsequently, they classified the argument components into major claims, claims, and premises, employing the level approach. The interrelationship between and among these elements in light of support and attack was also considered. The findings were in favor of using end-to-end argument mining, which can be used to evaluate argumentative essays.

Quite recently, Cocarascu et al. (2020) conducted a study to identify the best generalizable model of argument mining. They utilized various datasets that had different approaches to argumentation, encompassing the annotation of persuasive essays based on the level approach, which incorporated major claims, minor claims, and premises. This study considered different argumentation approaches and employed different argument-mining methods, with Bert being one of the most significant models used. The main difference between Bert and other models applied in this experiment lies in its ability, as a transformer, to consider contextual embeddings of words. The results of this study revealed a model of argument mining that performed almost in the same way in different argumentative contexts. This achievement is significant from the lens of generalizability, which is of utmost importance in both the fields of argumentation and argument mining.

All of these studies show that argument mining can help researchers and text producers shape their argumentative attitudes, competence, and behavior. All these studies concentrated on essays written by students from different disciplines. The researchers concentrated on examining the two argumentative components of claim and premises and, more notably, on major claims, minor claims, and premises (three components). However, studies applying multi-label argumentation approaches are sparse. What is notably missing in the literature is related to how argumentative mining is implemented in the argumentative discourse represented in academic contexts.

Table 2.

A Review of the Studies Conducted Based on the Argument Mining Approaches

Year	Authors	Approach	Framework	Data-driven Analysis
2015	Peldszus & Stede	Gradient Descent as an optimization technique in the machine and deep Learning,	Argument Extraction Relation prediction	Persuasive essay (Argument-micro text corpus as a parallel German / English corpus of 112 short texts)
2017	Stab & Gurevych	Deep Learning (End to End Argument Mining)	Argument Extraction Relation prediction	Persuasive essay (402 English essays from essayforum.com)
2017	Eger, Daxenberger, Gurevych	End-to-End Argument Mining (Bi-LSTM LSTM-ER)	Argument Extraction Relation Prediction	Persuasive essay (dataset of persuasive essays- PE – from Stab & Gurevych, 2017)

2018	Nguyen & Litman	End-to-End Argument Mining	Argument Extraction Relation Prediction	Persuasive essay (two corpora of holistically scored persuasive essays (first, the essays written for prompt; second includes over 8,000 essays from the TOEFL)
2019	Chernodub, et al.	End-to-End Argument-Mining	Argument Extraction Relation Prediction	Persuasive Essays; Microblogs and Web debating Platforms from (Gurevych, 2017)
2020	Cocarascu, Cabrio, Villata, & Toni	End-to-End Argument Mining	Relation prediction	Variety of datasets including persuasive essays

3. Method

This section is divided into two main parts. Initially, adopting the research synthetic approach (Plonsky & Oswald, 2015), we explained the steps taken to identify and select the representative meta-analyses disseminated in applied linguistics (i.e., meta-analysis identification and retrieval). Then, we adhered to argument mining techniques as a guidepost to identify and analyze the introduction section of meta-analyses.

3.1. Synthetic Approach: Meta-analysis Identification

Systematically locating the domain of meta-analyses in AL is the first step in the research synthesis approach. We operationally defined the domain via a three-pronged framework of location, time, and scope (see Plonsky, 2013, 2014). To representatively sample AL journals, we initially applied the two sampling criteria projected by Amini Farsani et al. (2021):

(a) The alignment of journals with AL strands identified by the American Association of Applied Linguistics (AAAL) and the British Association of Applied Linguistics (BAAL)

(b) The representation of AL journals in the SCOPUS

In this study, we further set two more benchmarks:

(c) AL journals need to be covered in the SSCI journal with a reported impact factor.

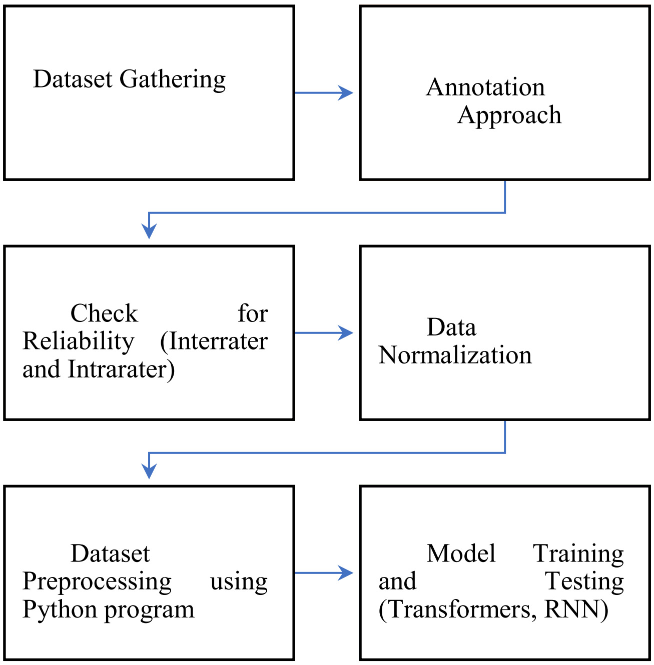
(d) Those meta-analyses were published in leading AL journals from 1998 to 2019.

Accordingly, we located 54 meta-analyses published in: *Language Learning, Computer Assisted Language Learning, Applied Linguistics, Language Teaching Research, Studies in Second Language Acquisition, Modern Language Journal, System, Language Learning & Technology, Applied Psycholinguistics, ReCALL, Language Testing, TESOL Quarterly, Second Language Research, and Canadian Modern Language Review.* All these journals are leading ones in the field of applied linguistics (see Amini Farsani et al., 2021) and have published meta-analyses within the time set.

3.2. Argument Mining Approach

As depicted in Figure 1, we adhered to argument-mining techniques reported in the literature for collecting and analyzing the data, which are all given in the following paragraphs:

Figure1.
A Flow Chart of Argument Mining Task for This Study



3.3. Data Gathering

Having selected the meta-analyses, we then extracted their introduction section, yielding 142,000 words. Comparing the newly developed dataset with the previous studies, we found three striking differences. The nature of the data in this study (i.e., published meta-analyses) was so different from the previous ones (i.e., students' argumentative essays). We used a newly developed argumentative framework projected by Qin and Karabacak (2010). Given its dynamicity and flexibility, this modified Toulmin model lends itself to argument mining techniques, notably the multi-label approach. As for the multi-label approach, we adopted the modified Toulmin model developed by Qin and Karabacak (2010). This newly developed argumentative framework comprises six components: (1) Claim; (2) Data claim; (3) Counterclaim; (4) Data counterclaim; (5) Rebuttal; and (6) Data rebuttal (see Table 3).

Table 3.

Argumentation Elements Used in Meta-Analysis Dataset

Number	Elements	Definitions and Examples
1	Claim	Definition: 'A declaration in response to a challenging and contentious subject.' Example: "The DDL approach is geared to making sense of language input but has several potential advantages that other input approaches do not." (Boulton & Cobb, 2017, p. 350)
2	Data claim	Definition: 'Data given as a support for the claim it refers to.' Example: "Core among these is that input assembly replaces input simplification, thus maintaining the authenticity of language. Another advantage lies in identifying which forms and meanings in a language (whether words, structures, pragmatic patterns, etc.) are most frequent and thus probably most worth knowing" (Boulton & Cobb, 2017, p. 350).
3	Counterargument claim	Definition: 'Probable opposing opinions that may challenge the core claim.' Example: "Shiotsu (2010) speculated that the uniquely low correlation found in the Japanese L1 group by Brown and Haynes may be attributable to the EFL teaching practices in Japan, which emphasize literacy skills over

		oral communication skills" (Jeon & Yamashita, 2014, p. 169).
4	Counterargument data	Definition: Data given as support for the counterargument it refers to. Example: "Informed by Brown and Haynes's findings, Shiotsu doubted that listening comprehension would have a high impact on L2 reading among his L1 Japanese participants and decided not to include this variable in his study" (Jeon & Yamashita, 2014, p. 169).
5	Rebuttal claim	Definition: 'Statements given by the writer or speaker who has proposed a claim as a respond to a counterclaim.' Example: "Nevertheless, the process of meta-analyzing the research domain taught us that the quality of reporting for moderator variables in the individual studies was often insufficient for our meta-analytical purposes" (Jeon & Yamashita, 2014, p. 163).
6	Rebuttal data	Definition: 'Data given as a support to the rebuttal which include clarification of probable deficiencies and weaknesses of the claim including invalid conjectures, fallacies, etc.' Example: "Thus, the final pool of moderator variables was limited to age, L1–L2 language distance, L1–L2 script distance, L2 proficiency, and three different types of measurement characteristics" (Jeon & Yamashita, 2014, p. 163).

Qin and Karabacak (2010) projected a context-specific modified Toulmin model to account for the limitations highlighted in the first version of the Toulmin model. Six elements or ADUs are represented in the second version of the Toulmin model. This representativeness includes either the presence of rare ADUs like 'rebuttals' (rebuttal claim and data) or the argumentation format like cataphoric[3] or anaphoric[4]. The last significant feature that differentiates this project from the others, specifically in the field of argument mining, is the number of words in each ADU that is significantly higher compared to previous corpora. One ADU in this dataset may include more than two paragraphs, which is not at all comparable to other datasets in previous research. Figure 2 illustrates one paragraph of our dataset, representing three different ADUs of claim, data claim, and rebuttal in an anaphoric argumentation format.

Figure 2.

An Illustrative Example of ADUs for Claims, Data-claims, and Rebuttal within Corpus

```
Previous reviews
claim-20- - Meta-analyses of instructional practices on language learning have shown
some interesting results. * dataclaim-18-claim-20 In a meta-analysis on literacy practices and
reading and writing instruction, Graham et al. (2017) found medium
effect sizes (ES=.47) on writing quality among other outcome variables.
In another meta-analysis, Graham, McKeown, Kiuhara, and Harris
(2012) examined studies that analyzed six different writing interventions.
They found several positive effects from these interventions, for example,
prewriting activities (ES=.54), peer assistance when writing (ES
=.89), product goals (ES=.76), and assessing writing (ES=.42) etc. These
meta-analyses provide a very succinct picture of reading and writing
instructions and their effectiveness. * rebuttal-1-claim-20 However, they do not examine the
use of education technology applications on reading and writing, especially
on ELLs literacy development. *
```

[3] The core claim appears before the referential ADUs, including that data claim, counter claim, etc.

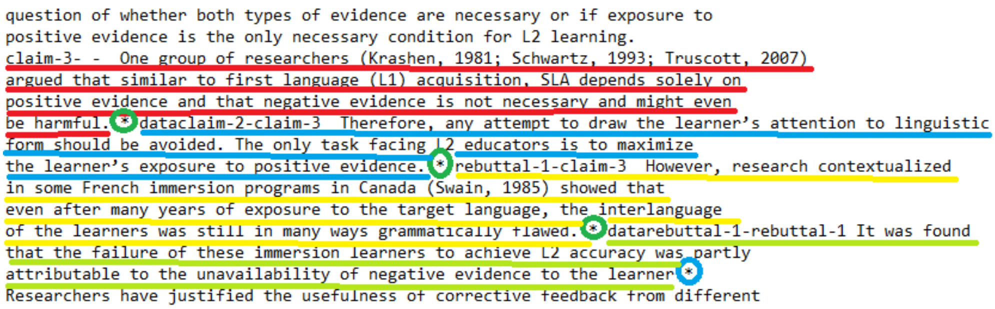
[4] The core claim appears after the referential ADUs, including that data claim, counter claim, etc.

3.4. Annotation

The second stage in the argument-mining approach is to annotate the collected data. Initially, we defined five statuses[5] regarding each ADU. As such, we dissected argumentative versus non-argumentative discourse units of the corpus based on the definitions given in the literature. Second, we distinguished the types of ADUs in light of the modified Toulmin model. Moreover, a non-argumentative discourse unit was added to the tags (labels), yielding seven tags (six modified Toulmin ADUs + non-argumentative units). Third, we defined the relationship between the ADUs and their references, which can be either supported or rejected. Fourth, by adding reference IDs to ADUs, we can track each ADU and its reference.

Figure 3.

An Example of Annotating with Labels and Codes


question of whether both types of evidence are necessary or if exposure to
positive evidence is the only necessary condition for L2 learning.
claim-3- - One group of researchers (Krashen, 1981; Schwartz, 1993; Truscott, 2007)
argued that similar to first language (L1) acquisition, SLA depends solely on
positive evidence and that negative evidence is not necessary and might even
be harmful.* dataclaim-2-claim-3 Therefore, any attempt to draw the learner's attention to linguistic
form should be avoided. The only task facing L2 educators is to maximize
the learner's exposure to positive evidence.* rebuttal-1-claim-3 However, research contextualized
in some French immersion programs in Canada (Swain, 1985) showed that
even after many years of exposure to the target language, the interlanguage
of the learners was still in many ways grammatically flawed.* datarebuttal-1-rebuttal-1 It was found
that the failure of these immersion learners to achieve L2 accuracy was partly
attributable to the unavailability of negative evidence to the learner*
Researchers have justified the usefulness of corrective feedback from different

To check the annotation's reliability, we examined both inter-rater and intra-rater reliability. In so doing, we divided the whole dataset into three sets of 15, 15, and 20 papers. Having completed the annotation procedure for each set, we considered both intra-rater annotation and inter-rater annotation as below. Five meta-analyses of each set were randomly chosen to be annotated by each author. A correlation of more than 80 percent was received, which was a cogent percentage to consider the annotation reliable. The inconsistencies in annotation were discussed until we reached an agreed-upon agreement.

3.5. Dataset Preprocessing

One of the most important phases of this research is the preparation of the annotated dataset for training the model. This process involves making changes and modifications to the original data, which is referred to as dataset preprocessing. In so doing, we followed the input requirements of the Bert and

[5] Argumentative versus non-argumentative, types of ADUs, ADUs relationships (support, attack, and non), tractable references, and finally the distance between each core ADU and its references.

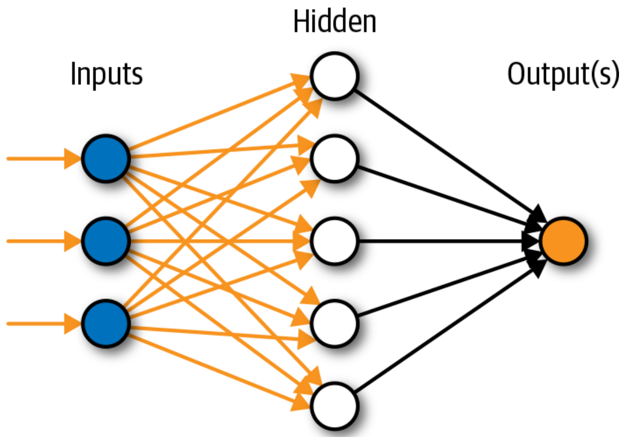
Bi-LSTM-CNN-CRF models, considering our ultimate purpose in language modeling. As the first step, we defined a tag structure that consisted of four distinct parts, including the Token position ID (TP ID), Argument Discourse Unit type (ADU type), Relationship type, and Reference ID, respectively. The TP ID signifies the position of the token in our intended sequence. This information is important since our model is a sequence-labeling model, which requires knowing where a sequence starts and ends. The TP ID includes three different characters: B (beginning of a sequence), I (Middle of a sequence), and * (End of a sequence). ADU type serves as a representation of the seven labels (tags) established in this research, including claim, data claim, rebuttal, data rebuttal, counterclaim, data counterclaim, and non-argumentative. These six components, along with the additional non-argumentative label, are defined based on the Modified Toulmin Model of Argumentation.

Relation types and Reference IDs in our Tag system complement each other. Relation types determine the relationship between the current ADU and the core ADU. The Claim ADU serves as the core ADU for the remaining five elements (e.g., data claim, rebuttal, data rebuttal, counterclaim, data counterclaim) of the modified Toulmin model. Relations types indicate whether the current ADU supports the core referenced ADU (e.g., data for the claim, rebuttal claim, data for rebuttal claim) or attacks it (e.g., counterclaim, data for counterclaim). Since referenced ADUs (claim, counterclaim, rebuttal claim) and their dependent ADUs (data for a claim, data for a counterclaim, data for a rebuttal claim, counterclaim for a claim, and rebuttal claim for a claim) may not appear consecutively or in close proximity, we used reference IDs to keep track of them. This matters because, contrary to previous studies in which sentences were the intended portions of language to be fed to the training model, the unit of analysis in this study was a paragraph at the pragmatic level rather than the semantic level. It is worth reminding readers that this procedure was repeated for all four approaches of the modified Toulmin: major–minor claim, level, and argumentative-non-argumentative approaches. Each time, our labeling procedure was adjusted based on the ADUs specific to the approach at hand.

3.6. Modelling and Training

To program and train the model, we adopted an end-to-end neural approach. The motive behind such a decision is that the neural network, with its unique structure and automaticity bonus, has enjoyed higher reliability and precision in many experiments compared to other machine learning algorithms. Figure 4 illustrates different parts of neural networks, with the input and output represented as the visible parts, and the hidden layers as the invisible parts of the neural network data processing stage, often referred to as the black box. After defining the model, the preprocessed input language, formatted to suit the model's input requirements, is fed into the model on one side. The model then tries to predict the label for each imported sequence of data. Considering whether the predicted tag and the gold standard (actual) tag for that sequence are the same or not, the model adjusts the weights and bias values. These weights and bias values help the model make 6 predictions for the input sequence.

Figure 4.
Different Parts of Neural Networks



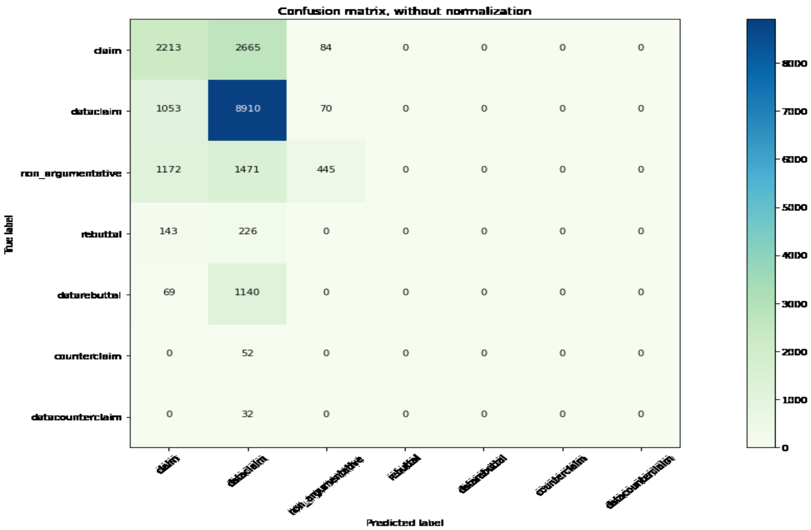
4. Results and Discussion

4.1. Results

Initially, our results were not what we expected. That is, the model could only be able to predict claims and data, resulting in a symmetrical representation in the recognition task (see Figure 5)

Figure 5.

Confusion Matrix for Modified Toulmin Argumentation Approach



We believe that there are three main solutions to improve the results in this case. First, we need to increase the dataset size so that the model has enough encounters with each argument discourse unit (ADU), including ADUs like rebuttals. Second, we should assign weights to different labels of the research to avoid bias in the system's behaviour. Unfortunately, we cannot normalize the dataset given the very low frequency of counterclaim, data counterclaim, rebuttal, and data rebuttal. Normalization of the data would lead to the deletion of rare labels like rebuttals, which are crucial for our dataset. Third, we must make sure no ADU is cut off at the end of paragraphs. This way, the system can gain a comprehensive understanding and holistic comprehension of the text at hand.

In order to ensure the accuracy of our resolution in regards to the modified Toulmin model and to address some of the shortcomings of the dataset, notably the low frequency of certain ADUs such as rebuttal, we implemented subsequent changes. Pouring from multiple classes (ADUs) into a single superclass, we tried to compensate for the deficiencies, which are presented in Table 4 based on the table below:

Table 4

Transformation

Approach	ADUs	Non-ADUs

Claim-premise (One-claim-approach)	Major-minor (Level-approach)	Modified Toulmin (multi-label-approach)
claim	Major claim	Counterclaim
	Minor claim	Counterclaim
		Rebuttal
Premises	Premises	Data rebuttal
		counterclaim
		Data counterclaim
Non-argumentative	Non-argumentative	Non-argumentative

This consideration provides us with a more symmetrical and balanced representation of the various ADU types, ensuring that our mechanism based on the modified Toulmin model remains reliable and accurate. As presented in Table 4, in the level approach, the model could not predict some of the labels as well. More specifically, the model failed to predict labels such as “premises” and “non-argumentative” labels. As such, we have decided to shift our concentration from the level approach to the one-claim approach to see how the new approach works in our prediction task.

As the modified Toulmin model and level approach did not work on our dataset, we have adopted the one-claim approach. In this approach, we distinguished three different discourse units from each other. First, we determine whether the token in question is argumentative or non-argumentative. Second, if the token is argumentative, we classify it as either a claim or a premise. Accordingly, in this approach, major claims are now referred to simply as claims, and all other ADUs such as minor claims, premises, rebuttals, and counterclaims are considered as premises for a claim. This does not mean that minor claims, including rebuttal and counterclaims, are ignored. Instead, they are all counted and embedded as premises for a claim (see Figures 6, 7).

Figure 6.

Confusion Matrix for Level Approach

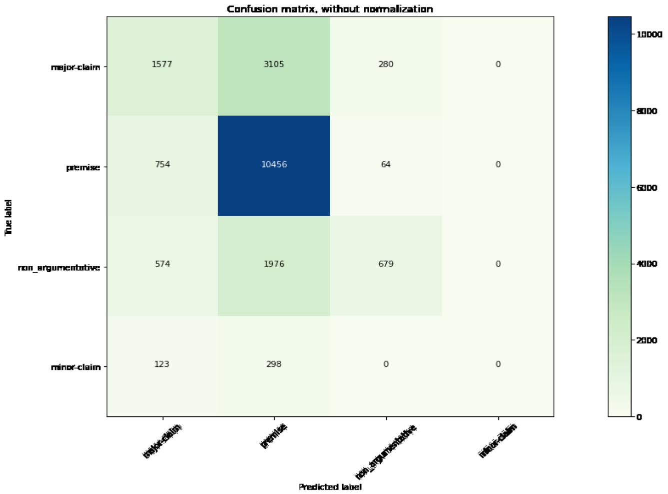


Figure 7.

Confusion Matrix for Main Claim Approach

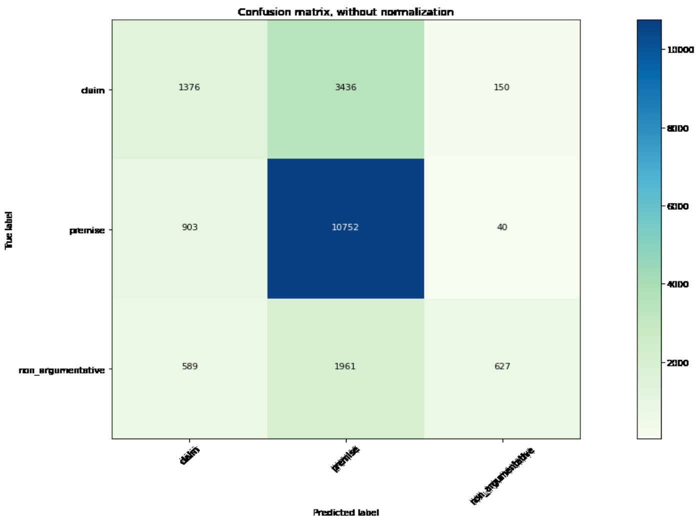


Table 5.

The result of the Main Claim Approach

	precision	recall	f1-score	support
1	0.4798	0.2773	0.3515	4962
2	0.6658	0.9194	0.7723	11695
3	0.7674	0.1974	0.3140	3177
accuracy			0.6431	19834
macro avg	0.6377	0.4647	0.4792	19834
weighted avg	0.6355	0.6431	0.5936	19834

As shown in Table 5, around 28 percent of the claim labels were extracted from the dataset, with a precision rate of around 48 percent and an *f1* score of 35 percent. Furthermore, around 92 percent of the total premise tokens are correctly identified, signifying a remarkably high precision rate of 66 percent. This notable performance brings about an impressive *f*1 score of 77 percent. Finally, around 20 percent of the whole tokens that are non-argumentative are dissected from other tokens with a high precision of 63 percent, which gives us an *f*1-score of 31 percent. On contrary to the previous approaches, this one works the best with the Bi-LSTM, and it does not get biased, as the results above exhibit. The above findings signify the argumentative nature of the meta-analysis, with over 80 percent of the text being argumentative. Furthermore, there are argumentative patterns in the meta-analysis genre that align with the Modified Toulmin argumentation approach. This suggests that argument-mining models can learn these patterns during the training procedure and use them to predict similar structures.

4.2. Discussion

The overall purpose of this research was to unveil the nature of argumentative writing presented in the meta-analyses published in the field of applied linguistics. As such, we examined the suitability and adaptability of the modified Toulmin model developed by Qin and Karabacak (2010) for analyzing argumentative texts of the academic genre, specifically targeting the introduction section of meta-analyses. We implemented different argument-mining approaches to explore an objective pattern for meta-analyses. The findings revealed that such a model is well-represented in the introduction section of meta-analyses. Moreover, the findings signify the fact that writing a well-developed argumentative introduction, coupled with incorporating the surface and quality of argumentative elements, to monitor author stances in supporting and refuting L2-related language problems is of utmost importance in producing a good-enough meta-analysis report. It also shows the persuasive nature of the introduction section of the meta-analysis in which researchers, here applied linguists, need to contextualize their projection in light of contrastive or inconclusive perspectives.

Given the cognitive complexity of producing good-enough claims, including claims, counterclaims, and rebuttal claims, and generating solid reasons for supporting the corresponding claims, it seems that there is no model or guidepost to help AL meta-analysts shape their argumentative competence, mindset, and behaviour. This lack of an argumentative map and landscape might affect meta-analysts' performance in producing improper argumentation. Such limited concentration on argumentation skills in formal academic training for applied linguists further exacerbates the issue. That is, AL meta-analysts have placed their primary emphasis on the research methodologies, content knowledge, and maturity of L2 problems, with less attention given to producing solid argumentative skills. The message echoes that applied linguists should boost their argumentative competence, which is highlighted by Rapanta et al.'s (2013) study. One way to enhance their argumentative performance, here for AL meta-analysts, is to follow the six-pronged argumentative model projected by Qin and Karabacak (2010) for producing solid argumentation skills. It can be applied explicitly in the introduction section, similar to the study the conducted in CALL journal in which Lin and Lin (2019) somewhat explicitly delineated claims, counterclaims, and rebuttals. They shaped their arguments on the effectiveness of mobile L2 vocabulary learning based on claims and counterclaims, on one hand, and refuted the documented proposition by explicit use of "The rebuttals to MALL" (Lin & Lin, 2019, p. 881), on the other hand. The point is, given our findings, we can delineate or move-analyze the introduction genre of meta-analysis in light of the argumentative model identified and substantiated in the study.

Our findings also shed light on the contributory role of the one-claim approach in substantiating the argumentative model. We implemented this approach for three reasons. Firstly, we encountered a lack of an adequate number of ADUs for certain labels within the dataset. Although the dataset contained more than 140,000 vocabularies, there was an inadequate quantity of long ADUs corresponding to the modified Toulmin model. Secondly, the distribution of labels within the dataset was unbalanced and asymmetrical, necessitating the adoption of the one-claim approach to address this issue. Finally, the third concern, somewhat related to the first issue, was the extensive length of tokens associated with each ADU. This posed a challenge for the model to learn the patterns perfectly, signifying the need to mitigate such concerns through the use of the one-claim approach.

5. Conclusion and Implications

Rhetoric, as one of the pillars of research paradigms, has not received adequate emphasis in the research-based strand of applied linguistics. This study focused on this untouched area of research, highlighting the argumentative nature of secondary research, notably meta-analysis, which is the most prevalent and typical genre (see Amini Farsani et al., 2021). Considering the commensurability of argumentation and meta-analysis (see Melendez-Torres et al., 2017) and given the increasing rate of meta-analytic studies in the field of applied linguistics (Plonsky, 2017), the findings of this study unveil the argumentative nature of the introduction section of meta-analyses with six argumentative elements: claim-data, counterargument claim-CA data, rebuttal-claim-RA data. Such findings carry significant implications for the fields of applied linguistics, L2 academic writing, meta-analysis and research synthesis, computational linguistics, and computer sciences.

First, in the field of applied linguistics, particularly in academic writing, the integration of the meta-analysis genre, the modified Toulmin approach of argumentation with 6+1 components, and the Bi-LSTM-CNN-CRF model of argument mining exemplify the multidisciplinary nature of applied linguistics in addressing L2 problems, especially in academic writing. This multidisciplinary orientation is warranted given the multidimensionality of academic writing issues, notably in those genres that require a higher cognitive load, such as meta-analysis. For applied linguists interested in working with the meta-analysis genre, such findings help them to write argumentatively in the introduction section in light of the model identified. Furthermore, the results of this study can be applied to academic writing courses and inform postgraduate students and researchers, particularly those who are passionate about research synthesis and meta-analysis. Such a guiding model can help them navigate an explicit model of argumentation when producing well-developed introductions.

From an academic writing perspective, this argumentative model can be considered a valuable tool for both learners and instructors. A modified version of the dataset can turn this model, or similar models, into a reliable and objective source for evaluating and generating different ADUs. Furthermore, researchers can boost their argumentative competence, enabling them to critically analyze primary studies' findings and effectively present their own stances in the introduction section of meta-analytic reviews. The incorporation of such an argumentative model can contribute to improving the overall quality of their academic writing, research output, and well-organized inferences.

In this study, we concentrated on the introduction section of meta-analyses as the unit of analysis. Therefore, the results of this study can be helpful for meta-analysts when writing the introduction section. By incorporating the insights provided by the modified Toulmin framework, meta-analysts can boost their argumentative writing skills and strategies and effectively structure their introductions. These pedagogical implications can be explicitly integrated into courses such as research methods and academic writing, in which rhetoric and its related aspects hold primary importance.

In the fields of computational linguistics and computer sciences, the results might have some methodological and pedagogical implications. Although these implications may not be directly tied to the field of applied linguistics, we argue that the multidisciplinary nature of applied linguistics can offer valuable insights to researchers in other related disciplines. By choosing meta-analysis, recognized as one of the most argumentative genres, as the dataset for argument mining and employing a comprehensive range of argumentation approaches, including state-of-the-art models such as Bi-LSTM-CNN-CRF and Bert, this pioneering study opens up avenues for further explorations and highlights the significance of addressing dataset characteristics in a machine learning approach.

As already mentioned, the challenges encountered with transformers (BERT and RoBERTa) in analyzing the dataset can be an interesting line of research. Examining the difficulties faced by pre-trained-of-the-art pre-trained models against different data mining and, in a more general sense, Information Extraction tasks can provide valuable insights into the limitations and potential areas for improvement for these models. The new datasets also allow computational linguists to consider secondary research genres such as meta-analysis. They can consider these datasets for future studies inspired by argumentative mining. For example, this research introduced meta-analysis as the new genre in argument mining science, which opens up a new line of research for computer sciences and computational linguistics. Remarkably, previous argument mining-based studies have not taken this genre into account, making it an untouched and unexplored landscape.

Future researchers should prepare a meta-analysis genre dataset of a larger size. The model should encompass a more comprehensive understanding of all labels, including counterclaim and data counterclaim, as well as ubiquitous labels like claim and data claim. By incorporating an ample number of labels, researchers can normalize the dataset and mitigate any potential bias stemming from the innate features of this genre. Prospective researchers should also replicate similar studies with a larger and more normalized number of samples and consider other methodological approaches, particularly mixed-methods research (MMR) in applied linguistics. Mixed-methods research, which integrates quantitative and qualitative inferences, provides a gestalt perspective on L2 issues. It is speculated that the argumentative landscape of MMR is different from mono-methodological approaches such as quantitative and qualitative research approaches (Amini Farsani & Mohammadi, 2020; Amini Farsani et al., 2022), thus warranting empirical investigation.

Finally, future researchers may use advanced generative models such as GPT-2 or GPT-3 (Generative Pretrained Transformers) and tune these models on our dataset in order to prompt the models to generate high-quality arguments. This line of research could be very useful for applied linguistics, specifically in the domain of L2 academic writing. Furthermore, the application of generative models in a comparative analysis between a primary argumentation dataset (e.g., argumentative essays) and a secondary argumentation dataset (e.g., meta-analysis) holds substantial potential.

References

Abdollahzadeh, E., Amini Farsani, M., & Beikmohammadi, M. (2017). Argumentative writing behavior of graduate EFL learners. *Argumentation*, *31*, 641-661.

Amini Farsani, M., & Babaii, E. (2018). Mapping past, current and future TEFL research trends in Iran: a synthetic review of topics and paradigms in three decades. *Journal of Modern Research in English Language Studies*, *5*(4), 81-108.

Amini Farsani, M., & Mohammadi, V. (2022). Mixed methods research in an EFL context: A quality assessment perspective. *Journal of Modern Research in English Language Studies*, *9*(2), 99-122.

Amini Farsani, M., Abdollahzadeh, E., & Beikmohammadi, M. (2019). Self-regulated learning, metacognitive awareness, and argumentative writing: A structural equation modeling approach. *Writing & Pedagogy*, *11*(2).

Amini Farsani, M., Babaii, E., Beikmohammadi, M., & Babaii Farsani, M. (2022). Mixed-methods research proficiency for applied linguists: a PLS-path modelling approach. *Quality & Quantity*, *56*(5), 3337-3362.

Amini Farsani, M., Jamali, H. R., Beikmohammadi, M., Ghorbani, B. D., & Soleimani, L. (2021). Methodological orientations, academic citations, and scientific collaboration in applied linguistics: What do research synthesis and bibliometrics indicate? *System*, *100*, 102547.

Anani Sarab, M. R., & Amini Farsani, M. (2023). Second-order synthesis of meta-analytic studies in applied linguistics (1998–2021). *Quality & Quantity*, 1-27.

Ajjour, Y., Chen, W. F., Kiesel, J., Wachsmuth, H., & Stein, B. (2017, September). Unit segmentation of argumentative texts. In *Proceedings of the 4th Workshop on Argument Mining* (pp.118-128). Copenhagen.

Aker, A., Sliwa, A., Ma, Y., Lui, R., Borad, N., Ziyaei, S., & Ghobadi, M. (2017, September). What works and what does not: Classifier and feature analysis for argument mining. In *Proceedings of the 4th Workshop on Argument Mining* (pp. 91-96).

Al Khatib, K., Wachsmuth, H., Hagen, M., Köhler, J., & Stein, B. (2016, June). Cross-domain mining of argumentative text through distant supervision. In *Proceedings of the 2016 conference of the North American chapter of the Association for Computational Linguistics: human language technologies* (pp. 1395-1404), Melbourne.

Ball, W. J. (1994). Using Virgil to analyze public policy arguments: A system based on Toulmin's informal logic. *Social Science Computer Review*, *12*(1), 26-37.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language learning*, *67*(2), 348-393.

Byrnes, H. (2013). Notes from the editor. *The Modern Language Journal*, 97, 825-827.

Cocarascu, O. Cabrio, E. Villata, S & Toni, F. (2020). A Dataset independent set of baselines for relation prediction in argument mining. Association for Computational Linguistics. https://doi.org/10.48550/arXiv.2003.04970

Chernodub, A., Oliynyk, O., Heidenreich, P., Bondarenko, A., Hagen, M., Biemann, C., & Panchenko, A. (2019, July). Targer: Neural argument mining at your fingertips. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 195-200). Association for Computational Linguistics.

Eger, S., Daxenberger, J., & Gurevych, I. (2017). Neural end-to-end learning for computational argumentation mining. In *Proceedings of ACL* (11-22)

Graff, G., & Birkenstein, C. (2010). *They say, I say. The moves that matter in Academic Writing*. W. W. Norton and Company.

Hoey, M., & Winter, E. O. (1983). *On the surface of discourse*. Allen & Unwin.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language learning*, *64*(1), 160-212.

Lin, J. J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878-919.

Loewen, S., & Plonsky, L. (2016). *An A–Z of Applied linguistics research methods*. Macmillan Education UK.

McKinley, J. (2020). Theorizing research methods in the golden age of applied linguistics research in J. McKinley and H. Rose (Eds.): The Routledge *Handbook of Research Methods in Applied Linguistics* (pp. 1-12). Routledge.

Melendez-Torres, G, J. O'Mara-Eves, A. Thomas, J. Brunton, G. Caird, J. Petticrew, M. (2017). Interpretive analysis of 85 systematic reviews suggests that narrative syntheses and meta-analyses are incommensurate in argumentation. *Research Synthesis Method*, *8*(1), 109–118.

Nguyen, H., & Litman, D. (2018, April). Argument mining for improving the automated scoring of persuasive essays. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1). Association for the Advancement of Artificial Intelligence.

Osman, W. H., & Januin, J. (2021). Analyzing ESL persuasive essay writing using Toulmin's model of argument. *Psychology and Educational Journal*, *58*(1), 1810-1821.

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*(4), 655–687.

Plonsky, L. (2014). Study quality in quantitative L2 research (1990-2010): A methodological synthesis and call for reform: Study quality in quantitative L2 research (1990-2010). *Modern Language Journal*, *98*(1), 450–470. https://doi.org/10.1111/j.1540-4781.2014.12058.x

Plosky, L. (2017). Quantitative research methods. In *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 505-521). Routledge.

Plonsky, L., & Oswald, F. L. (2015). Meta-analyzing second language research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 106-128). Routledge.

Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System*, *38*(3), 444-456. https://doi.org/10.1016/j.system.2010.06.012

Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research*, *83*(4), 483-520.

Stab, C., & Gurevych, I. (2017). Parsing argumentation structures in persuasive essays. *Computational Linguistics*, *43*(3), 619-659.