# A general-purpose semantic relationship taxonomy for the classification of logical relations and conceptual associations

***Stuart G. Towns***
*King Mongkut's University of Technology Thonburi*

**Abstract**
Recent advances in computer technology have given rise to text analysis research which investigates the linguistic features of text. One such feature is lexical cohesion, which includes referential cohesion, logical relations, and conceptual associations. Previous research has studied referential cohesion, but very few have analyzed the use of logical relations and conceptual associations in real world texts. In this study, three potential taxonomies are identified as potentially useful for this task. However, these taxonomies show both strengths and weaknesses for this purpose. Therefore, this study discusses the adaption of these taxonomies into a new easy-to-use, broad-based, semantic relationship taxonomy for the classification of logical relations and conceptual associations in a given text. A pilot study using the new taxonomy showed the taxonomy's application to differentiating between two texts.

## 1. Research into the linguistic features of text

The proliferation of computer-based language analysis tools in recent years has prompted investigations into the linguistic features that can be found in written texts. Many studies across various fields have attempted to use linguistic features as a way to distinguish between different categories of text. These differences might be between different levels in the development of students' language proficiency (Bulté & Housen, 2014; Mazgutova & Kormos, 2015), differences in writing quality found in professional writing (Lewoniewski, Węcel & Abramowicz, 2018; Louis & Nenkova, 2013), or differences between genres of text (Lieungnapar, Watson Todd, Trakulkasemsuk, 2017; Wolters & Kirsten, 1999). The linguistic features that are investigated in these types of studies are varied as well. Some focus on specific linguistic areas such as parts of speech and semantic similarity (Moschitti & Basili, 2004; Yannakoudakis & Briscoe, 2012), syntactic complexity (Kyle, 2016; Yang, Lu, & Weigle, 2015) lexical complexity (Higginbotham & Reid, 2019; Gregori-Signes & Clavel-Arroitia, 2015), or lexical cohesion (Flor & Klebanov, 2014; Yang & Sun, 2012). Others take a broader approach by investigating a wide range of linguistic features across all levels of language (Crossley & McNamara, 2011; Towns & Watson Todd, 2019). All of these studies into linguistic features of texts would be difficult if not impossible without the use of automated tools. One such tool that has been used in several studies is Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004). Coh-Metrix is able to automatically analyze texts on 108 linguistic features representing several different linguistic aspects such as syntactic complexity, lexical complexity, and referential cohesion. Other automated tools are more specialized, for example, only focusing on syntactic complexity (Lu, 2010) or lexical complexity (Lu, 2012).

With regards to the analysis of student writing proficiency, many studies have shown that qualitative measures of linguistic complexity often increase as proficiency improves. For example, as students become more proficient writers, their writing becomes more syntactically complex based on metrics such as sentence length and the number of clauses per sentence. It also becomes more lexically complex based on the diversity and sophistication of the words they use.

There appears, however, to be an inverse relationship between lexical complexity and referential cohesion. As students' writing improves and their vocabulary becomes larger and more sophisticated, the amount of referential cohesion created by repetition and references (e.g., with anaphors such as pronouns) decreases. In other words, as students improve their writing skills, they do not have to rely on referential cohesion to create coherence in the mind of the reader (Crossley, Weston, Sullivan, & McNamara, 2011).

**2. Research into logical relations and conceptual associations**

Even though studies have shown that referential cohesion plays a role in the judgement of writing proficiency as seen above, it is not the only type of lexical cohesion. There are two other main ways that concepts can be connected in a text: lexical relations such as pairs of synonyms or hyponyms, (Cruse, 2011) and semantic associations (Hoey, 2005). However, these two types of lexical cohesion are rarely investigated in actual texts, partly because there are no easy-to-use automated tools to aid in the research. For the studies on lexical cohesion which do exist, many of them use the Coh-Metrix software tool, or a more modern version of the same approach in a tool called TAACO: the Tool for the Automatic Analysis of Cohesion (Crossley, Kyle, & McNamara, 2016). While these tools are relatively straightforward and easy to use, the downside is that they are purely quantitative measurements of the average frequencies of connections across a text, mostly focusing on referential cohesion, that is, connections created through the use of repetition of words and by referencing previous concepts, for example by using a pronoun to refer back to a previously mentioned noun. These tools also claim to use some semantic association measurements such as Latent Semantic Analysis (LSA) or NLP algorithms such as word2vec. These semantic relationships may be logical relationships such as synonyms or hypernyms or some other kind of undefined semantic association, but the tools do not report what the actual word-pair connections are or what types of connections they might be. Therefore, while the quantitative metrics can shed some light on the differences between texts, it is impossible for researchers to extend the analysis to include a qualitative perspective in order to gain a deeper understanding of how cohesion is actually manifest in a specific text. If a researcher wants to identify and classify the semantic relationships in a text, the analysis must be done by using a time-consuming manual analysis.

A study by Towns and Watson Todd (2019) is the only known study that has reported full quantitative and qualitative results regarding the three different types of lexical cohesion given above on specific texts as a way of comparing the texts. In their study, lexical cohesion was measured by first manually identifying the lexical chains of related concepts and then by categorizing the connections into three main categories of referential cohesion (repetition and reference), logical relations (synonyms, antonyms, hypernyms, sister terms, and meronyms), and a catch-all category of what they called conceptual associations. It was this last category that showed the strongest results when comparing the texts in the study. Therefore, it seems that the category of conceptual associations is an important area to investigate further, as it appears to have a large effect on the perceived writing quality of the text, even surpassing that of syntactic complexity and lexical complexity based on the results of Towns and Watson Todd's study. Perhaps one way to gain deeper insights into the role of conceptual associations in a text is to not only identify them into one large category, but to also categorize them into smaller, fine-grained groups for further analysis.

## 3. Potential methods for future research on logical relations and conceptual associations

Following the methodology from Towns and Watson Todd (2019), there are two steps in analyzing texts for logical relations and conceptual associations: 1) identifying the words which are semantically related, and 2) categorizing the relationships. For the first of these two steps, there are some promising semi-automated tools that might be able to provide insight into the nature of conceptual associations. One study showed that some of these methods (i.e., look-ups in a digital thesaurus, WordNet, and a word association database, semantic tagging using the UCREL Semantic Analysis System, near neighbors scores using Latent Semantic Analysis, and Mutual Information (MI) scores from the Corpus of Contemporary American English (COCA)) could identify potential semantic connections in a text, but there was still a large amount of manual analysis needed because these tools were not built for the purpose of analyzing a given text in this way (Towns & Watson Todd, 2017).

After the semantically associated word-pairs in a text have been identified, a researcher might also want to know what kinds of connections these pairs are. For logical relations, these categories include synonyms, antonyms, hypernyms, and meronyms. These are relationships that typically might be found in a dictionary or thesaurus. The types of conceptual associations, however, are more difficult to define, as they are dependent on many variables such as the context where the words are found as well as the cultural and background knowledge that the reader brings to the text. These context-based semantic relationships have been researched in many different fields for different purposes. One example of a study of this kind is from Morris and Hirst (2006). In this study, participants were given a short text and asked to identity related words and to put them into categories in order to determine the type of relationship. However, as might be expected, the results showed that there was a large amount of disagreement about both the word-pairs chosen and the categories they were placed in. Perhaps a study like this would be more successful if the participants were coding the connections based on a pre-defined set of connection types, rather than making their own based on their own personal schemas and background knowledge.

Some potential taxonomies to use in classifying semantic relationships come from the field of cognitive psychology and are known as semantic feature norms. Research in this area investigates associated concepts which are collected through stimulus-response experiments where participants are given one concept and asked to respond with a related concept. The semantic features are then developed by independent coders using content analysis (Bolognesi, Pilgram, & van den Heerik, 2017). Many different taxonomies of feature types have been developed, with one of the largest being one by Wu and Barsalou (2009) which contains 37 categories. These categories are divided into five macro-categories. The first macro-category is taxonomic properties, which are logical relations. The remaining four macro-categories cover conceptual associations as entity properties, situation properties, introspective properties, and miscellaneous. The study by Bolognesi, et al. (2017) further adapted this taxonomy based on their investigation into its reliability and ended up with a similar taxonomy consisting of 20 categories in four macro-categories. They have similar macro-categories as Wu and Barsalou (2009) with one macro-category for logical relations and three macro-categories for conceptual associations. (The miscellaneous macro-category was removed.) These two taxonomies are discussed further in the next section.

A third potential taxonomy that could be used to classify the relationships between words in a text comes from the multidisciplinary field known as terminology. As the name suggests, this field is concerned with terms, or concepts, and the definitions and names given to those terms, especially in specialized (e.g., technical) fields (Sager, 1990). Concepts do not appear in a vacuum, of course, so one way to help define terms is to compare them to other terms. Terminology separates these semantic relations into three main macro-categories:

1. Generic relationships: These are hierarchical relationships which create a tree structure with general concepts at the top of the tree and successively more specific concepts at each level going down the tree. The tree has two kinds of relationships: vertical hypernym/hyponym relationships and horizontal co-hyponyms, or sister terms. These are also known as "type of" or "kind of" relationships, such as X is a kind of Y.
2. Partitive relationships: These are "part-whole" relationships, where one concept X is a constituent of another concept Y.
3. Complex relationships: This category was originally intended as a catch-all miscellaneous category for all other relationships. Sager (1990) provides a list of potential relationships in this category, which is discussed further in the next section.

As with the taxonomies from the semantic feature norms discussed above, it can be seen that these macro-categories cover both logical relations and conceptual associations. The first two categories of generic and partitive relationships are both logical relations, while the third macro-category of complex relations are made up of conceptual associations.

## 5. Research purpose

Although these taxonomies from cognitive psychology and terminology seem to have a strong potential to be appropriate classification systems for conceptual associations, it is not clear which one would fit the task better. There also does not seem to be any research which uses any of them in an actual text analysis for the purpose of identifying and classifying semantic relationships in a given text.

Since research into semantic associations might potentially provide clues into the progression of student writing proficiency and quality, or help to differentiate different types of text, and given that there are no clear tools or taxonomies to easily identify and classify these relationships, this paper aims to provide an easy to use, useful taxonomy for this task. The next section of this paper describes how and why the two previous taxonomies were adapted and combined to reach this goal. The paper then concludes with a discussion of a small pilot study using the new taxonomy.

## 6. Analysis of three taxonomies

### *6.1 Cognitive psychology: Semantic feature norms from word association studies*

As mentioned above, Wu and Barsalou (2009) created a semantic relation taxonomy with 37 categories across five macro-categories. However, a study by Bolognesi et al. (2017) tested this taxonomy for reliability with coders and found several issues such as ambiguous categories that were difficult to use. One result of their study was the creation of a new taxonomy with 20 relationship categories in four macro-categories. The current study analyzed these two taxonomies in an attempt to create an easy-to-use taxonomy for classifying conceptual associations in a general text. Adaptations to these taxonomies to achieve this goal can be divided into several types of changes, as follows.

1. Combining categories with directional relationships: The semantic network containing logical relations and conceptual associations that will be created by using the taxonomy created in this study is an undirected graph. In other words, there are no directions between concepts and the concepts can appear in any order. This means that several categories were able to be combined, such as the combination of the Superordinate and Subordinate categories into one hypernymic relationship.
2. Combining of specific relationships into more general ones: Several of the categories were too specific and could be combined into more general relationships. For example, in addition to the Superordinate and Subordinate categories described above, there was also an Ontological category which was a hypernymic relationship where one of the pairs was a very general concept such as thing or object. This category was also combined into the hypernymic category. Another example is in the macro-category of entity properties. Part-whole relationships were divided into separate Inside and Outside categories (such as watermelon-seed and watermelon-rind). This is too specific for the goals of this project and so were combined into one part-whole relationship category.
3. Removal of contextual relationships: Many of the relationships were only relationships in very specific contexts. Very often, these specific contexts were constructed of phrases rather than single words. The goal of this study is to create a general-purpose taxonomy with single-word pairs, so these multi-word contexts are not applicable. For example, the Quantity category which included relationships such as apple-very red and tree-lots of leaves was ignored. These contextual relationships were especially noticeable in the macro-category of Introspective Properties. These categories included people's reactions or feelings towards objects. These are too context-specific to be useful for the purposes of this study and were also ignored.
4. Accepting changes made by Bolognesi, et al.: There were several oversights in the original taxonomy by Wu and Barsalou (2009) that were corrected in Bolognesi et al. (2017). Three examples of these changes are as follows: 1) the logical relationship of Antonyms was added, 2) Time and Event categories were combined into one category (e.g., present-Christmas and breakfast-morning), and 3) individual taxonomic categories which were very context specific such as car-my car were ignored. In a full cohesion analysis, this latter category would be counted as referential cohesion.

After this analysis, the new taxonomy being developed for this study contained many of the categories from these previous taxonomies, with the exceptions as seen above. However, one problem with these categories is that almost all of them are based on concepts that might fulfill the role of nouns in a sentence. Only two of them were related to actions or processes. These two categories were Entity:Behavior and Situation:Action. Therefore, additional categories were needed, and so the second group of categories from the technical definitions of concepts found in terminology research was also used. The next section provides more details about this taxonomy.

### *6.2 Terminology: Conceptual relations*

As mentioned before, the field of terminology is concerned with defining terms, especially those in technical fields. Terminologists divide semantic associations into three categories: generic relationships (e.g., hypernyms), partitive relationships (e.g., meronyms) and a third catch-all category called complex relationships. However, Sager (1990) notes that the number of possible relationship types in this third category is potentially very large. Sager's list of possible complex relationship types and examples can be seen in Table 1.

**Table 1.** Relationships and word-pair examples from Sager (1990, p 35)

| Relationship category | Examples |
|---|---|
| cause – effect | explosion – fall-out |
| material – product | steel – girder |
| material – property | glass – brittle |
| process – product | weaving – cloth |
| process – instrument | incision – scalpel |
| process – method | storage – freeze-dry |
| process – patient | dying – textile |
| phenomenon – measurement | light – watt |
| object – counteragent | poison – antidote |
| object – container | tool – tool box |
| object – material | bridge – iron |
| object – quality | petrol – high-octane |
| object – operation | drill bit – drilling |
| object – characteristic | fuel – smokeless |
| object – form | book – paperback |
| activity – place | coalmining – coalmine |

By analyzing the examples in Table 1, it can be seen that many of these relationships would be useful for the purposes of creating a general list of relationships to use in our research. However, the field's roots in technical definitions perhaps causes this list to be too specific. For example, we probably do not need to differentiate between objects and materials or between processes and activities. This would allow us to combine some relationships, as our goal is to be as general yet as useful as possible. In addition, many of these relationships are already covered in the previous taxonomies discussed above. However, there were four categories that were especially useful. Three of them are concerned with actions or processes (activity-place, process-product, process-instrument), which was a shortcoming of the taxonomies discussed above, and the fourth is a technical measurement (phenomenon-measurement).

## 7. Development of a new taxonomy

By adapting and combining the three taxonomies shown above, a new taxonomy of 20 semantic relation categories was created. Since one of the aims of the current study is to create a taxonomy that is specific enough to produce a useful analysis, but not too specific so that there are too many categories to choose from, the 20 categories developed in the current study seem to be an appropriate number. This is backed up by previous research as well. In a review of the literature on taxonomies developed from word association tests, Bolognesi, Pilgram, and van den Heerik (2017) listed nine studies which created taxonomies for classifying semantic features. The number of categories ranged from four (Garrard, Lambon, Hodges, & Patterson, 2001) to 37 (Wu & Barsalou, 2009). Two other studies were in the middle of this range, with 19 categories each (Recchia & Jones, 2012; Lenci, Baroni, Cazzolli, & Marotta, 2013). Therefore, it seems that 20 categories is large enough to show useful distinctions, but small enough to be easy to use.

**Table 2.** The new taxonomy of conceptual association categories

| Category | Word 1 | Word 2 | Natural language phrase |
|---|---|---|---|
| **Logical relations** | | | |
| Synonym | bike | bicycle | X has the same meaning as Y |
| Antonym | love | hate | X is the opposite of Y |
| Parent-Child | car | vehicle | X is a type of Y |
| Sister terms | pepper | salt | X and Y have the same parent |
| Part-Whole | window | house | X is part of Y |
| **Activities** | | | |
| Agent | teacher | teach | X does Y (activity) |
| Instrument | writing | pencil | X (activity) is done using Y (obj) |
| Recipient | teach | student | X (activity) to Y (changed) |
| Theme | teach | subject | X (activity) Y (unchanged) |
| Location | teach | classroom | X (activity) occurs at Y (location) |
| Output | manufacture | product | X (activity) outputs/results in Y |
| **Entities** | | | |
| Property | table | wood | X has an property of Y |
| Origin | watermelon | ground | X comes from Y |
| Location | car | garage | X is contained/located in Y |
| Time/Event | watermelon | picnic | X is used in event Y |
| Measurement | light | Watt | X is measured in Y |
| Contingency | smoke | pollution | X causes/requires/depends on Y |
| Participants | child | toy | X (person) interacts with Y (obj) |
| Abstract | clothing | protection | X (person/obj) interacts w/abstract concept Y |
| Object | grass | sun | X (obj) interacts with Y (obj) |

Another goal of the taxonomy is that it should be able to be used in a linguistic analysis of general texts. To aid in this task, it was decided to divide the 20 categories into three main categories: Logical Relations, Activities, and Entities. The logical relations category contains all of the classical relations such as synonyms and antonyms, hypernymic (kind of) relations, and meronymic (part of) relations. The activities category contains relationships where one of the concepts is an action, activity, or process. The other word in the pair in an activity relationship, for example, can tell who is doing the action, where the activity is taking place, what instrument is being used to complete the activity, or what the output of the activity is. The third category of entities focuses on concepts that are nouns, whether they are people, objects, or abstract concepts. The paired words in the entities category can, for example, describe the properties of the concepts (often adjectives), locations, origins, or the time or event that the concept often appears.

Since this taxonomy was intended to be used by other researchers, a set of natural language phrases have also been developed to aid the researcher in knowing which category to use. The full list of 20 semantic relationships with example word-pairs taken from the three taxonomies above and natural language phrases can be seen in Table 2.

**8. Testing of the new taxonomy**

As stated in the beginning of this paper, it seems likely that logical relations and conceptual associations can also help to identify the differences between different types of text, whether they are written by students at different stages of proficiency, or come from different text genres, or are written in different writing styles. While counting all conceptual associations as one big category can show differences in texts (Towns & Watson Todd, 2019), a more fine-grained approach could potentially provide a more detailed analysis. The purpose of this paper, then, was to create a semantic relationship taxonomy that could be used to analyze general texts, with the outcome shown in Table 2 above. A pilot study was conducted by Siripol and Towns (2021) to test the taxonomy that was developed for the current study. This pilot study analyzed excerpts from two graded readers at different grade levels (beginner and advanced) by first identifying the words that were semantically related in each text, and then determining the relationship of each word-pair link using the taxonomy provided in the previous section. From the results of the study, it appears that the pilot test was successful. Nineteen of the twenty relationship types were found in at least one of the texts, with a range of 1-14 occurrences. Most of the categories had similar frequencies in the two texts, but there were some categories that showed differences. Perhaps the most interesting finding of the pilot study was that the quantitative frequencies of various semantic relations afforded a deeper qualitative analysis. For example, the beginner's text had more Entities:Participants while the advanced text had more Entities:Locations. This is because the beginner text was focused on people talking about their daily activities at work, while the advanced text focused on the setting of the story. After noticing this difference, the researchers realized that the locations in the advanced text could be interpreted as representing the protagonist's real and metaphorical journey. In this way, it appears that not only can this taxonomy be used in a quantitative study, but it can also be used as the foundation of a qualitative study that critically investigates more literary aspects of a text.

**9. Conclusion**

This study was conducted in order to create an easy-to-use taxonomy for the linguistic classification of semantic relationships in a general-purpose text. Three taxonomies from previous studies were analyzed for their use towards this goal. The first two taxonomies were based on word associations research. One was developed by Bolognesi et al. (2017), which was in turn based on earlier work by Wu and Barsalou (2009). The third taxonomy was Semantic Relations from Sager (1990) which was originally developed to add lexicographers in creating definitions for technical terms. These taxonomies were adapted and combined to create a 20-category taxonomy for this paper. The 20 categories were divided into 3 main categories of logical relationships (i.e., classical relationship such as synonyms and hypernyms), activity relationships (i.e., relationships which include an action or process), and entity relationships (i.e., relationships between people, objects, and abstract concepts). The taxonomy was tested in a pilot study which analyzed excerpts from two different levels of graded readers and was found to be an appropriate tool for that study's goals (Siripol & Towns, 2021). Therefore, it is believed that this taxonomy can be used in any manual analysis of semantic relations in a general-purpose text for the purpose of finding evidence of differences in text quality, writing proficiency, or genre and text type.

In addition, since semantic relations such as logical relationships and conceptual associations often show what the text is about (Siripol & Towns, 2021; Towns & Watson Todd, 2019), the taxonomy described in this paper could be used for research purposes similar to those in corpus-based keyword analyses. As seen in Pojanapunya & Watson Todd (2018), keyword analyses can also be used to make judgements about the words used in a text, or to reveal more abstract issues. It seems likely that the semantic relations taxonomy may also afford similar critical analyses of texts.

**References**

Bolognesi, M., Pilgram, R., & van den Heerik, R. (2017). Reliability in content analysis: The case of semantic feature norms classification. *Behavior Research Methods, 49*(6), 1984–2001. https://doi.org/10.3758/s13428-016-0838-6

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42-65. http://doi.org/0.1016/j.jslw.2014.09.005.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods, 48*(4), 1227-1237. http://doi.org/10.3758/s13428-015-0651-7

Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning, 21*(2-3), 170-191.

Crossley, S. A., Weston, J. L., McLain Sullivan, S. T. & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28*, 282-311. http://doi.org/10.1177/0741088311410188.

Cruse, Alan. 2011. *Meaning in language: An introduction to semantics and pragmatics*. Oxford University.

Flor, M., & Klebanov, B. B. (2014). Associative lexical cohesion as a factor in text complexity. *International Journal of Applied Linguistics, 165*(2), 223-258.

Garrard, P., Lambon, R., Hodges, J., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology, 18*, 125–174.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*, 193–202.

Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia - Social and Behavioral Sciences, 198*, 546-556. http://doi.org/10.1016/j.sbspro.2015.07.477.

Higginbotham, G., & Reid, J. (2019). The lexical sophistication of second language learners' academic essays. *Journal of English for Academic Purposes, 37*, 127-140. https://doi.org/10.1016/j.jeap.2018.12.002.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge. https://doi.org/10.4324/9780203327630.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Unpublished doctoral dissertation]. Georgia State University.

Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. (2013). BLIND: A set of semantic feature norms from the congenitally blind. *Behavior Research Methods, 45*, 1218–1233.

Lewoniewski, W., Węcel, K., & Abramowicz, W. (2018). Determining quality of articles in polish Wikipedia based on linguistic features. In Damaševičius, R. & Vasiljevienė, G. (Eds.), Information and Software Technologies. ICIST 2018. *Communications in Computer and Information Science* (pp. 546-558). Springer, Cham. https://doi.org/10.1007/978-3-319-99972-2_45

Lieungnapar, A., Watson Todd, R., & Trakulkasemsuk, W. (2017). Genre induction from a linguistic approach. *Indonesian Journal of Applied Linguistics, 6*(2), 319-329. https://doi.org/10.17509/ijal.v6i2.4917.

Louis, A., & Nenkova, A. (2013). What makes writing great? First experiments on article quality prediction in the science journalism domain. *Transactions of the Association for Computational Linguistics, 1*, 341-352.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*, 474-496.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Journal, 96*(2), 190-208.

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for academic purposes programme. *Journal of Second Language Writing, 29*, 3-15. http://dx.doi.org/10.1016/j.jslw.2015.06.004.

Morris, J., & Hirst, G. (2006). The subjectivity of lexical cohesion in text. In Shanahan, J.G., Qu, Y., Wiebe, J. (Eds.), *Computing attitude and affect in text: Theory and applications*. Springer.

Moschitti, A., & Basili, R. (2004). Complex linguistic features for text classification: A comprehensive study. In *European Conference on Information Retrieval*. 181-196. Springer.

Pojanapunya, P., & Watson Todd, R. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory, 14*(1), 133-167.

Recchia, G., & Jones, M. N. (2012). The semantic richness of abstract concepts. *Frontiers in Human Neuroscience, 6*, 315.

Sager, J. C. (1990). *A Practical Course in Terminology Processing*. John Benjamins. https://doi.org/10.1075/z.44

Siripol, P. & Towns, S.G. (2021). A network-based method for identifying and categorizing semantic relationships. In *Proceedings of DRAL 4: Doing Research in Applied Linguistics*. Bangkok, Thailand.

Towns, S. G., & Watson Todd, R. (2017). Methods for Identifying Relations and Associations in Text. In *Proceedings of DRAL 3: Doing Research in Applied Linguistics*. Bangkok, Thailand, 2017, 146-155.

Towns, S. G., & Watson Todd, R. (2019). Beyond proficiency: Linguistic features of exceptional writing. *English Text Construction, 12*(2), 265-289. https://doi.org/10.1075/etc. 00029.tow.

Wolters, M., & Kirsten, M. (1999). Exploring the use of linguistic features in domain and genre classification. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*. 142-149.

Wu, L.-L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica, 132*, 173–189. doi:10.1016/ j.actpsy. 2009.02.002

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing, 28*, 53-67.

Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education, 23*(1), 31-48.

Yannakoudakis, H., & Briscoe, T. (2012). Modeling coherence in ESOL learner texts. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*. 33-43.