

# The Second-Language Essay as Cognitive Task: Complexity, Subjectivity and Emotion

CA DeCoursey<sup>1</sup>

<sup>1</sup> Nazarbayev University, Astana, Kazakhstan

Correspondence: CA DeCoursey, Associate Professor, Nazarbayev University, Astana, Kazakhstan.

Received: March 14, 2022      Accepted: November 20, 2022      Online Published: November 30, 2022

doi:10.5539/ijel.v13n1p43      URL: <https://doi.org/10.5539/ijel.v13n1p43>

## Abstract

Models of task complexity indicate the multiple processes and challenges second-language university students face when learning to write an essay. Studies of complex cognitive tasks frame emotion as an aspect of individual differences. This study used Appraisal analysis to assess subjective attitudes realised across four weeks of writing an essay, content analysis to identify how students took up instructor input, and co-frequency to identify strong connections. Results indicate that students focus on researching essay content at the expense of structure and language. They find topic sentences more difficult than thesis statements, and have difficulty collating sentence-level proficiency with the sophisticated discourse-level demands of the essay task. Content and attitude frequencies suggested that relatively little work was done in the first week, substantial work was done in weeks two and three, and realisations dropped sharply in the final week, suggesting resource-dispersing impacts in the final stretch. Results highlight the need for somatic measures of task complexity and effort, due to frequent realisations of stress and stress relief, and their co-frequency with misery.

**Keywords:** emotion, second-language writing, essay-writing, Appraisal analysis, work, identity

## 1. Introduction

Learning to write a university essay is a complex task. Students must master the metalanguage of composition—thesis statement, controlling idea, topic sentences, coherence and cohesion, rhetorical appeals, structuring inductive, deductive, Toulmin and other structures for argumentations—and then apply this terminology in their writing and research, along with constructing data representations and following bibliographic conventions (Shin, Cimasko, & Yi, 2020). Various properties of the texts they generate are assessed using multiple criteria for language, structure and content, as are the associated learning activities—concept mapping, critical reading notes, revising, peer reviewing, conferencing, pairwork and groupwork (Sawyer, 2006). Intended learning objectives in composition courses also often include transferable learning skills such as critical thinking, academic literacy, problem-solving, information literacy and intercultural sensitivity (Mestre, 2005). Writing in an L2 is also a complex task. Students must receive input in an L2, negotiate its meaning and application with peers, take in new collocations of sense, syntax and figuration alongside constant error correction, and navigate the associated inner challenges of maintaining positive motivation (Casanave, 2013). Many L2 textbooks and readings assume and model native-speaker cultural perspectives, affirming ways of being, thinking and doing which often reflect the western monolingual middle-class, meaning L2 students must integrate their local identities with international and native-speaker cultural norms (Pennycook, 2014). Clearly, taking freshman composition in an L2 is a demanding experience, and we should expect it to arouse complex subjective reactions in learners.

## 2. Literature Review

Concepts used to understand cognitive work, and thus to analyse L2 writing, arose from workplace analyses of task components, during the 1960s through 80s. This tradition defined subtasks, steps and difficulty levels, within different contexts and performance conditions (Arnold, Umanath, Thio, Reilly, McDaniel, & Marsh, 2017). A three-component model of the learning task emerged from this tradition, the Triadic Componential Framework of task complexity (Robinson, 2007). Its first component comprised cognitive factors involved in the nature of the input, and the procedures required for processing this information. Its second component comprised task completion conditions, which includes elements such as planning time, group interactions, and contributions from other sources (Lambert & Engler 2007). Its third component comprised factors within learners that shape

task difficulty levels, including operations required to process information in order to generate the required output, language variables, and working memory (Baddeley, 2003). Learners' internal subjective states received attention within the third component, particularly self-efficacy (Bandura, 1997), motivation (Dornyei, 2009), intrinsic enjoyment (Ajzen, 2005), interest (Hidi & Renninger, 2006) and anxiety (Mayer et al., 2000). In these studies, subjective attitudes are understood as a substrate for and precursor to learning behaviours (Gardner, 2006).

The Triadic Componential Framework has generated three streams of research into how people learn complex tasks. Studies with a sociocultural focus on learner identity explore task complexity through the social interaction undertaken during the process of text generation, and on the time and other resources used in acquiring and implementing the input knowledge (Swain & Lapkin, 2001). Other studies focus on learner acquisition of structural and generic features of written texts. Many of these have explored lexical acquisition and grammatical accuracy, with most studies of writing focused on the acquisition of discrete forms at the level of the clause or sentence (Barcroft, 2006; Keating, 2008; Kim, 2008). They have also modelled relationships between noticing and producing the expected output type, complexity and fluency, attentional processes and working memory, form-concept mapping, implicit and explicit feedback (Robinson, 2007; Reigeluth & Carr-Chellman, 2009). Still further studies of writing at the level of the paragraph and longer texts have focused on narrative (Gilbert, 2007; Robinson, 2007; Tavakoli & Foster, 2011; Kormos & Trebits, 2012). Finally, some studies have taken an information theory approach (Robinson, 2001). These three approaches have mapped the challenges facing L2 students, in terms of input, language, and sentence to paragraph to text type outputs. As this work has gone forward, the subjective attitudes and experiences of the learner have been foregrounded. Complex tasks such as writing a university essay demand accurate and complex output, placing a significant burden on what L2 learners must do to comprehend and negotiate the meaning of task input (Bruna & Gomez, 2009). The emotions evoked by learning tasks, including anxiety, enjoyment, boredom, pride and relief, have been studied, using a states-and-traits model of emotion (Pekrun, Goetz, Titz, & Perry, 2010; Caprara, Vecchione, Alessandri, & Barbaranelli, 2011).

Within English for Academic Purposes (EAP), this research has been used to make teaching and learning experiences more fluid and feasible. Many studies apply the understandings gleaned from these three research approaches to task-based instruction, task sequencing and syllabus design, the staging of class materials and assessments, and helping language instructors' control how they step up cognitive demand in order to optimise students' language learning and performance (Duran & Ramaut, 2006). Still, for what goes on inside the learner, "the relationship between task type or task complexity and writing performance is by no means clear" (Kukken & Vedder, 2008, p. 49).

All three research approaches within the Triadic Componential Framework have noticed the significance of individual differences, from biological varieties to cultural influences, and from psycholinguistic to neurolinguistic parameters (Filmore, Kempler, & Wang, 2014). Much of this research follows Skehan (1989) in addressing specific areas in which individual differences impact learning, especially language-learning aptitude, cognitive/learning styles, learning strategies, personality and motivation. These differences are felt to "explain variation in language learning successes under particular instructional conditions" (Robinson, 2001, p. 369). More recently, affect has been emerging as a significant factor in L2 learning. Ellis identified individual "propensities" which blended "both cognitive and affective dimension[s]" and reflected "an individual's preferred way of processing information and dealing with other people" (2004, p. 534). Self-efficacy studies have considered the relationship of self-confidence and learner anxiety to self-concept (Bandura, Caprara, Barbaranelli, Gerbino, & Pastorelli, 2003; Ehrman, Leaver, & Oxford, 2003; Revesz, 2011; Kormos, 2012). Dornyei has problematised "the seemingly comprehensive and straightforward picture of identities being stable and monolithic learner traits" (2009, p. 232), and called for qualitative and mixed-method studies to explore the complex of subjective emotional responses in defining identities (Dornyei, 2014). Thus, subjective attitudes are of interest to researchers at this time (Grabe, 2009).

The models of emotion used in studies of L2 learning tasks differ from those employed in sentiment analysis, which arose from forensic linguistic studies of stance and evidentiality, and from automated opinion-mining techniques used in marketing (Argamon, Bloom, Esuil, & Sebastiani, 2007; Yang, Lin, & Chen, 2007). These techniques are grounded in psycholinguistic and neurolinguistic studies of the biophysical base of emotions in the brain (Pavlenko, 2002), which give rise to the lexicogrammar of emotion and, through indirect morphosyntactic resources, extensive lexis for realising a great variety of subjective attitudes (Wurm & Vakoch, 1996). Psycholinguistics models the linguistic realisation of subjective states as networked sets of hierarchically taxonmised semantic classes (Fontaine, Scherer, & Soriano, 2013). Appraisal analysis, based in systemic

functional linguistics, includes the largest range of semantic sets, categories and subcategories (Halliday & Matthiessen, 2004; Bednarek, 2008). When expressing subjective opinions, people choose words and phrases from among the great variety available (Whitelaw, Garg, & Argamon, 2005). Realisations of attitude “diversify across a range of grammatical structures” (Martin & White, 2005, p. 45). They may be placed on a polarity cline from negative to positive (Wiebe, Wilson, Bruce, Bell, & Martin, 2004).

Individual differences have been determined through questionnaires administered at single moments in time, before and/or after the task (Dörnyei, 2014). Writing an essay is a process unfolding over time. Emotions would vary across this experience, which questionnaires could miss. The emotions called up as learners tackle the challenges outlined above may not reflect individual differences so much as the burden in the moment. They may be better understood as stress appraisals, a cognitive or emotional response called up when events tax or exceed learner’s ability to adapt to them comfortably (Gross, 2007). It seems useful, then, to elicit learners’ subjective attitudes as broadly as possible, across the timeframe of the production of an essay, in order to explore appraisals called up during the process. In the same way, by defining complex cognitive tasks, the streams of scholarship outlined above have prefigured the field. It seems useful to elicit learners’ cognitive appraisals, to see from their perspective how they construe this complex learning task. This study used Appraisal analysis to explore attitudes, and content analysis to explore what learners thought they were doing, across the four weeks during which they produced an essay. Research questions included: What meta critical concepts from freshman composition do students choose to realise frequently? What emotions, judgments and appreciations were frequently-realised in student reflections? Which of these co-occur? Do frequently-realised attitudes change across the four weeks of writing an essay? How does this help us understand the essay as a complex task?

### **3. Method**

Qualitative or textual data was collected over a 4-week period, and then analysed first for content, and second for subjective attitude using Appraisal analysis. Inter-coder and inter-rater reliability scores obtained for these two stages of the data analysis. Significant co-frequency scores were determined.

#### *3.1 Participants*

Qualitative data was taken from 85 Arabic native-speakers, all of whom had English proficiency at the IELTS 6.5 or TOEFL iBT83 level. All were registered in one of five sections of a freshman composition course, at an English-medium-of-instruction university in Egypt. Data from four international students was removed from the corpus prior to analysis.

#### *3.2 Sample and Saturation*

Student participants pose no intrinsic problem for validity, or for conventional measures of confidence and error, when generalising to students as a population. However, with qualitative data analysis (QDA), saturation rather than sample size is the basis for validity, as cohort members will tend to realise normative themes and attitudes for their population (Cresswell & Plano Clark, 2011). This is based in the frequency with which specific cohort themes are realised by any given participant (Thompson & Juan, 2006). The minimum sample size for achieving 95% salience, or realisation of normative themes and attitudes, is seventeen, with samples larger than 50 participants (Tran, 2017).

#### *3.3 Instrument*

Participants wrote a 100-word personal reflection weekly across four weeks. The prompt used general terms to avoid artificially defining content and attitudes realised: “What was this class like for you this week, and why?” This prompt invited participants to realise both content and subjective attitudes weekly, according to what each personally felt to be important to them (Neumann, 2013). Data was collected over a four-week period during which participants produced a research essay, and prior to its being handed in.

#### *3.4 Teacher’s Role*

The teacher’s role during the four weeks of essay production was limited. The weeks chosen for analysis were the final weeks of the course, during which classes met irregularly, with teachers in the role of consultant. The final paper of the semester was a research paper. This was chosen as students had previously produced a variety of essay types with teachers providing weekly feedback, but for this final paper, students were meant to produce these on their own, with teachers available as a resource, replying to inquiries on email or able to meet students in office hours.

#### *3.5 Data Analysis*

Data were analysed in three ways. First, content analysis was used to identify frequently-realised themes and

subunits across the four weeks. Second, Appraisal analysis was used to identify frequently-realised emotions, judgments and appreciations across the four weeks. Inter-coder and inter-rater reliability were calculated using Cohen's  $\kappa$ , both percent overall and free margin (Lombard, Snyder-Duch, & Bracken, 2002).

### 3.5.1 Content Analysis

Content analysis is used to distil referential content from extensive textual data (Bryman, 2004). The derivation of larger themes articulated into subunits allows referential patterns to be represented numerically (Roberts, 1997). In this case, the corpus was coded in its entirety, and sampling units were not used (Krippendorf, 2004). The fundamental unit of analysis was the clause, which was used as a bounded coding unit (Babie, 2001). An emergent coding frame was derived from frequently-occurring nominal groups within the corpus (Weber, 1985). Element definitions were designed as mutually exclusive, to ensure that no clause was classified within more than one category. However, themes and subunits were not exhaustive, but limited to the scope of the study. Element definitions were kept broad, intuitive and clear, to exclude cases requiring disambiguation (Denzin & Lincoln, 2000). Clauses containing manifest subunit content were counted, but clauses containing latent content were not (Stemler, 2001). Element frequencies were obtained, with in-sentence counts requiring grammatical connection to a minimum of a single nominal group (Carley, 1990).

### 3.5.2 Appraisal Analysis

Second, data was analysed using the Attitude system from Appraisal analysis. The Attitude system network comprises three sets; Affect, Judgment and Appreciation, as in Figure 1. Affect includes realisations of positive/negative emotion (Halliday, 1994). For example, "I'm really happy with my results" realises positive Affect in the category and subcategory Un/happiness: pleasure. Many subjective attitudes are realised indirectly. Judgement and Appreciation rework feelings indirectly as assessments of people, events and objects outside the self (Scherer et al., 2001). For example, in "the new assignment looks hard and I do not know how it can be done", "hard" reworks a negative emotion (Affect-Un/happiness: pleasure) indirectly as an appreciation of the quality of the assignment (Appreciation-Reaction: quality) and "can" reworks negative emotion (Affect-In/Security: disquiet) as a negative Judgment of Social esteem: capacity. The Appraisal semantic taxonomy has gained in validity, as models of emotion and evaluation arising in linguistics and psychology increasingly resemble each other (Elfenbein & Ambady, 2009).

When aggregated into a corpus, textual data may be tagged to identify attitudinal patterns (Thompson, 2004). Text-tagging is a reliable and widely-used method in computational linguistics (Read & Carroll, 2010). Software programs use extensive concordances to define hierarchical classes, which are then applied in automatic and semi-automatic text-tagging (Bednarek, 2006). Concordances are built from supervised classification tasks, taking in massive internet-based textual data through which statistical machine-learning approaches are trialed and adjusted to handle voluminous specific cases and exceptions (Pang, Lee, & Vaithyanathan, 2002). Software is an efficient means of sorting lexicogrammar into semantic classes (Polanyi & Zaenen, 2006). Emotional lexis is the easiest lexicogrammar for software to identify (Read & Carroll, 2010). Attribution rates are highly reliable, though a few semantic subcategories such as sarcasm remain difficult for software to identify (Wiebe, Wilson, & Cardie, 2005). The software Corpus Tool (CT) which incorporates the Appraisal taxonomy, was used (O'Donnell, 2008). The combination of machine attribution with trained human-tagger checking is highly reliable (Polanyi & Zaenen, 2006). Because Appraisal analysis sorts word data into 24 semantic subcategories, it offers highly refined, or delicate results (Marín-Arrese, 2004). "Appraisal theories of emotions have gained widespread acceptance in the field of emotion research" (Kuppens, Van Mechelen, Smits, De Boeck, & Ceulemans, 2007, p. 689).

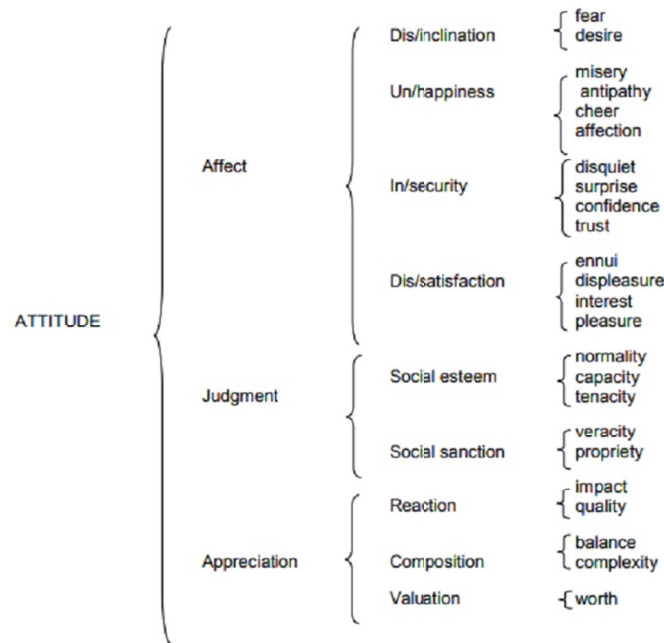


Figure 1. The attitude system

### 3.5.3 Significant Co-Frequencies

Co-frequencies were then determined, using SPSS and explored using Voyant Tools (Miller, 2018). The measure selected for use in defining significant co-frequencies, between content subunit and attitude subcategory, was Yule’s *Y* (Gries, 2008).

### 3.5.4 Reliability

The corpus was independently coded for content, and tagged for Appraisal tagged by the researcher and a research assistant. Inter-coder and inter-rater reliability were calculated, using both the free-margin, and percent-overall values as expressions of Cohen’s kappa (Lombard, Snyder-Duch, & Bracken, 2004).

## 4. Results

Corpus data was independently coded for content and Appraisal analysis, with the inter-coder reliability and significant co-frequencies then determined.

### 4.1 Content Analysis

Three themes comprising eight subunits were defined. Two themes and six subunits reflected frequently-realised input elements of instructor talk about composition in the areas of content and structure. To distinguish instructor input from content analysis, the term material has been employed as a theme header. A third theme with two subunits reflected frequently-realised student focused on the somatic elements stress, and stress relief. The number of clauses in the corpus overall containing work-related content was 1439, of which 1166 could be classified using the emergent schema, as in Table 1.

Table 1. Content analysis of essay work-related themes and subunits

THEME	SUBUNIT	WEEK 1		WEEK 2		WEEK 3		WEEK 4		T
		n	%	n	%	n	%	n	%	
structure	thesis statement	19	1.63	26	2.23	31	2.66	11	0.94	87
	topic sentence	30	2.57	58	4.97	71	6.09	33	2.83	192
	outlining & drafting	8	0.69	15	1.29	24	2.06	5	0.43	52
										331
material	research	70	6.00	111	9.52	122	10.46	41	3.52	344
	argument	24	2.06	41	3.52	53	4.55	19	1.63	137
	data	23	1.97	39	3.34	58	4.97	35	3.00	155
										636

somatic	stress	26	2.23	21	1.80	37	3.17	12	1.03	96
	stress relief	21	1.85	18	1.54	27	2.23	37	3.17	103
										199
		221	18.95	329	28.22	423	36.28	193	16.55	1166

Content analysis yielded frequency scores by subunit and week. Thematic frequency scores reveal that participants realised content as about twice as much work as structure. Content was realised about three times as frequently as the somatic impact of doing an essay.

The distinct character of each week was identified, as in Figure 2.

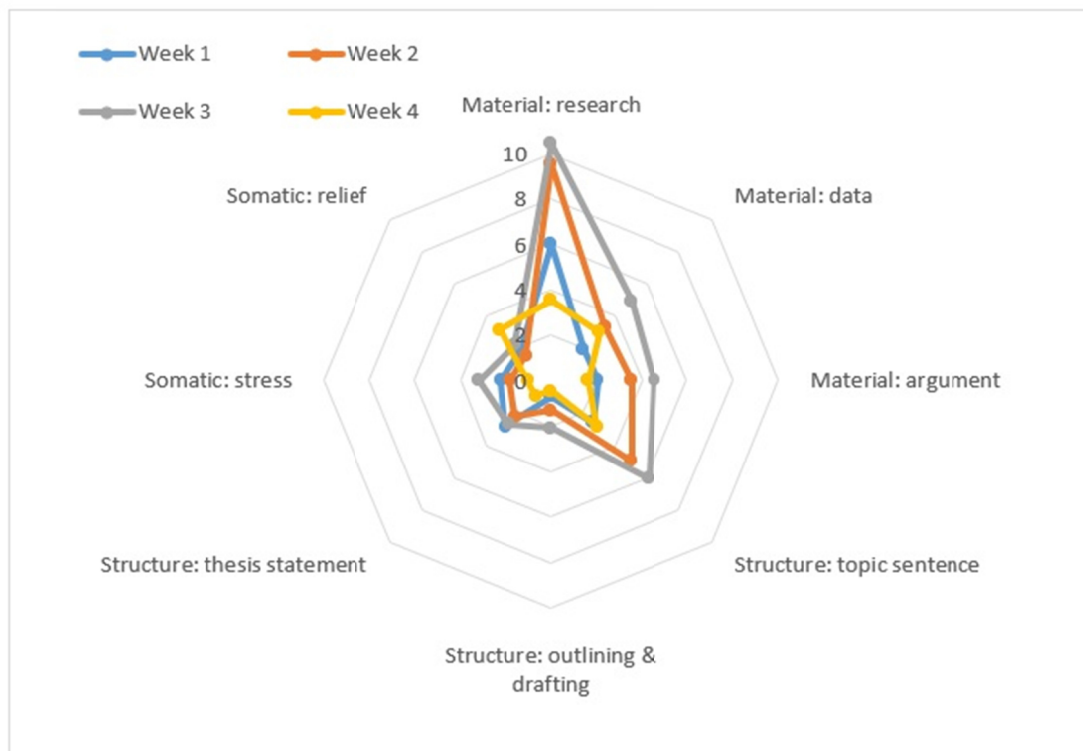


Figure 2. Frequency of theme-subunits across 4 weeks

Weeks 1 and 4 contained fewer content realisations than weeks 2 and 3. In week 1, *research* was the most frequently realised. In week 2, *research* and *topic sentence* were frequently realised. In week 3, they were again frequently realised, but slightly less so. The Week 4 subcorpus contained the least content. For Content and Structure subunits, realisations were at their highest in week 3, and reduced greatly in week 4. In the somatic theme, the subunit *stress* increased over the 4-week period. However, the subunit *stress relief* began high, increased over weeks 1 through 3, and reduced in week 4. In terms of the relative importance participants gave to content areas of the essay as a complex task, research was pre-eminent, with data and argument less than half as important. They gave greater importance to topic sentences than thesis statements or outlining and drafting, and slightly greater importance to stress relief than to stress, particularly in week 4.

#### 4.2 Appraisal Analysis

Data were then analysed using Appraisal analysis. Reflections were anonymised. Textual data from the five sections was aggregated into a corpus of 43,075 words, comprising 4 weekly subcorpora. Attitudes in the 1166 clauses containing work-related content were taxonomised by week and polarity, as in Table 2.

Table 2. Reflection sub-corpora by week

	WEEK 1	WEEK 2	WEEK 3	WEEK 4	T
No. Participants	85	78	82	60	
Subcorpus size	12 496	12 251	11 590	6 738	
Avg length	148.76	157.06	141.30	112.30	
No. attitudes	488	401	357	215	1461
no. positive	295	286	176	132	889
% positive	60.45	71.32	49.30	61.40	
no. negative	193	115	181	83	572
% negative	39.55	28.68	50.70	38.60	

Participant numbers varied across the four weeks, as did subcorpus size, average wordcount, number of subjective attitudes realised in total and as negative or positive attitudes, and attitudinal density. The corpus was tagged in CT by the lead researcher and a research assistant, both of whom had 300+ hours' experience with the software.

Realisations in eight semantic subcategories accounted for 72.48% of attitudes realised, as in Table 3.

Table 3. Eight most frequently-realised attitudes by week and rank

	ATTITUDE	WEEK 1		WEEK 2		WEEK 3		WEEK 4		T
		n	%	n	%	n	%	n	%	
1	+ capacity	55	5.19	61	5.76	63	5.95	38	3.59	217
2	+ impact	76	7.18	70	6.61	65	6.14	16	1.51	227
3	+ quality	30	2.83	34	3.21	36	3.40	39	3.68	139
4	+ complexity	42	3.97	31	2.93	54	5.10	7	0.66	134
5	- capacity	36	3.12	28	2.64	39	3.68	13	1.23	116
6	+ worth	27	2.55	21	1.98	25	2.36	31	2.93	104
7	+ interest	11	1.04	19	1.79	12	1.13	22	2.08	64
8	- misery	2	0.19	14	1.32	38	3.59	4	0.38	58
		279	26.35	278	26.25	332	31.35	152	14.35	1059

Of these, realisations in six subcategories accounting for 885 = 83.60% of all work-related attitudes were positive. Realisations in two subcategories accounting for 174 = 16.43%, indicating that negative attitudes were realised at a low rate among many of the 24 subcategories, where realisations of positive attitudes concentrated in fewer subcategories.

Positive and negative realisations of capacity accounted for 333 = 31.44% of all realisations. Judgments of capacity realise beliefs about routine future states of affairs, based on our individual past experiences (Panther & Thornburg, 1999). The modality of possibility (“can”, able to”, “have the ability to”, “know how to”, “am X enough to”, and so on) realises a writer’s beliefs about actions they have in the past mastered, routinely enact, and construe as expected future potentials (Stack, 2012). For example, realisations such as “I am able to deal with the workload” and “I can spot the minutest of details when it comes to mechanics” assumes past development of this skill, its present enactment, and the likelihood of future enactment. Similarly, negative realisations such as ““I always have so many ideas, yet am unable to summarize the quantity” and “We have practiced in the two previous classes but that is not enough to acquire the skill”, assumes past unsuccessful efforts, present inability, the likelihood of the inability continuing in the future. The purpose of a realisation such as is to express frustration with this ongoing

The lexis comprised by the reaction, composition and valuation semantic subcategories are related dimensions reflecting fundamental psychocerebral processes within perception and cognition (Eggins, 1994). Reactions of impact and quality rework personal emotion as the characteristics and impacts of objects outside the self. Thus, in “the format in which the essay should be written is especially exciting”, while “exciting” grammatically modifies “format”, clearly the emotion is felt by the writer, not by the textual property. Appreciations of complexity rework emotion through the extensive lexis found in all languages reflecting visual perception, which is used to assess object composition, both actual and abstract (Martin & White, 2004), as for example, “With this new information I will add more detail and content to my essay which will make it more effective in its meaning and message.” Appreciations of valuation realise “our considered opinions” (Martin & White, 2004, p. 57), for example “things I want to improve and make good decisions about are to make headers more precise, and

involve at least one creative diagram".

There were only two frequently-realised emotions, including positive Dis/satisfaction: interest. In "the theme is so interesting to write about or to keep searching", the emotion is framed as a quality of the topic being written about and researched. Few of these used a congruent, "I + feeling" construction placing the experience of emotion close to the self in a manner that "bear[s] a natural relation to the meanings they have evolved to express" (Halliday, 1985, p. 18). Most realisations were framed as qualities, or as comments, as in "it was interesting to read the data". Most negative realisations of Un/Happiness: misery were realised *irreal is*, and identified a trigger (Halliday, 1994), for example "I was sad that Monday's lecture ended quickly as we had no time for my group to debate".

#### 4.3 Inter-Coder and Inter-Rater Reliability

Inter-coder reliability scores for content analysis, and inter-rater reliability scores for Appraisal analysis were calculated using ReCal, as in Table 4 (Freelon, 2010).

Table 4. Inter-rater and inter-coder reliability scores

	INTER-RATER		INTER-CODER	
	p-o	f-m	p-o	f-m
corpus	0.863	0.815	0.783	0.721

Percent-overall values are weighted to assess agreement by chance, free-margin values to assess the degree of disagreement. Values greater than 0.75 are considered robust. Using content analysis for referential and Appraisal for latent content increases the overall validity of the study (Krippendorff, 2004).

#### 4.4 Significant Co-Frequencies

Co-frequencies were determined between content subunits and Appraisal subcategories. Significance was enhanced by selecting colligations of word classes over collocations, in order to exclude idiomatic phrases which, if counted, could over-representing relationships between content and attitude. For example, this would exclude the collocation "I can only stress that" being counted into the *stress* subunit and the capacity subcategory, and exclude the collocation "it is worth stressing that..." being counted into *stress* and worth. Co-frequencies were expressed using Yule's Y, where values of  $\pm 1$  describe co-frequency strength. Co-frequencies of greater than  $\pm 0.5$  are robust (Chung & Lee, 2001). Four strong co-frequencies were found, as in Table 5.

Table 5. Co-frequencies between content subunits and attitude subcategories

RANK	THEME: SUBUNIT	ATTITUDE			Y
		SET	CATEGORY	SUBCATEGORY	
1	Structure: <i>topic sentences</i>	Appreciation	composition	complexity	+0.64
2	Material: <i>data</i>	Judgment	social esteem	capacity	+0.61
3	Material: <i>research</i>	Affect	dis/satisfaction	pleasure	+0.58
4	Somatic: <i>stress</i>	Affect	un/happiness	misery	+0.52

Strongest co-frequencies were found between *topic sentences* and complexity ("this applies to the topic sentences. In addition, how to write them in much complicated vocabulary and structure"), *data* and capacity ("quantitative data can strengthen arguments"), *research* and pleasure ("Today's class was actually very amusing! It gave us an introduction to the process we have to go through to create our research"), and *stress* and misery ("this week has been so stressful and I'm so worried because I know that next week is going to be hell).

## 5. Discussion

This study generated four areas of insight into L2 composition as a cognitive task, both from content and from Appraisal analysis, as well as further understandings from the co-frequency data, and relating to differential impacts on students across the four weeks of the production process.

### 5.1 Mismatch Between Student and Instructor Understanding

First, there is an essential mismatch between how students and how their composition instructors understand what composition is primarily about. Despite extensive instructor input on the process of writing, the structuring of paragraphs, the linking of sentences and paragraphs within an overall argument, the use of devices for



coherence and cohesion, and so on, students overwhelmingly understood their essays as being about its ideational or referential content. That is, for tertiary learners, the essay is mainly the topic they chose. They understand higher level writing as defined by what they are writing about, far more than how they are writing about it. This has clear implications for their preparedness for the demands of writing in the work world, as professional genres are highly differentiated by structure.

### *5.2 Minimal Take-up of Tertiary-Level Structure*

Second, over the duration of a one-semester course, these participants had only begun to take in the highly structured nature of tertiary-level writing. These participants realised the work of research, argument and data twice as often as the work of crafting thesis statements and topic sentences, and outlining and drafting their work. Of these, they chose to realise the work of writing topic sentences about twice as often as that of writing thesis statements. This suggests they feel comfortable generating a stance (“controlling idea”) towards their chosen topic, but less comfortable chunking it, or breaking it down into manageable components to deploy sequence within an argument. This finding is consonant with the view that freshers’ confidence in and ability to be critical about thesis statements is unwarranted, and needs explicit teaching, practice and development (Guardado & Shi, 2007; Miller & Pessoa, 2016).

### *5.3 Lack of Critical Attention to Language*

Third, a notable feature of the content analysis data is the lack of realisations revealing students’ critical attention to language. Despite extensive instructor input in this area, for example to topics including grammar, vocabulary choice, style, diction, register, these items were not frequently realised. Realisations about language tended to be generalised (“the class is very enjoyable and is still increasing my knowledge of the English language as a whole”), conceived simplistically (“my language should be sophisticated”), or treated as a list item (“the essay as well aim to ensure a balance between content, mechanics, structure, and language”). L2 students are generally highly aware of error feedback (Ellis, 2008). This result contradicts Robinsons’ (2003) cognition hypothesis of concurrent increases in task complexity, attentional resources and performance gains. Instead, it resembles studies suggesting limited attentional resources, and the resource-dispersing impact of highly planned cognitive tasks (García Mayo, 2007). That is, because students’ attention was focused at the level of semantic and discourse complexity, they paid less attention to matters of sentence-level language proficiency.

### *5.4 Somatic Impacts*

Finally, the content analysis data revealed the degree to which L2 composition is experienced powerfully and somatically by fresher tertiary students. These realisations make clear both the complexity and the experienced level of demand, of learning to write at a professional level. To date, somatic measures have not been integrated into any model of task complexity, though technologies are available to measure work through real-time biomarkers such as heart rate, respiration and sweating, along with brain-imaging technologies. Pre- and post-task impacts could also be measured using sensory bracelets. These metrics could provide new insights into L2 writing as a cognitive task.

### *5.5 Time Frame Effects*

Seeing the essay as a task spread out across four weeks has generated three insights linked to the above results. First, most realisations of material and structure occurred in weeks 2 and 3, suggesting that the week 1 reflection was written prior to doing the greater part of the work of the essay. Second, the sudden decrease of content realisations, and the reduction in reflection word length in week four supports the hypothesis of limited attentional resources. In all sections, the due date for handing in the essay was after the fourth reflection. This suggests that these reflections do not reflect very fully what students were thinking, feeling and doing, as they were engaged in a gallop for the finish. This highlights the middle weeks as the best time to research what L2 learners are doing, when writing in their L2. More detailed time series studies might reveal distinct attitudinal stages associated with complex L2 writing tasks. Finally, mentions of stress peaked in week 3, and of stress relief in week four, highlighting the need for composition instructors to be aware of the somatic impacts of their course tasks.

## **6. Conclusion**

The attitude data revealed students’ overwhelming focus on ability, positive and negative, which totaled 333 = 31.45% of all attitudes realised. This is an appropriate result, as students are learning a skill. The frequent realisation of complexity and worth suggest they have an appropriate understanding of the demands of the task and its relevance to their lives and futures. Second, the attitude data revealed that students felt the impacts and qualities of the task very personally: these subcategories together accounted for 366 = 34.56% of attitudes

realised. These two areas of attitude approximate two conceptually distinct constructs in L2 motivation, intrinsic and instrumental, in a manner consonant with the view that they are interconnected and interactive in L2 learners' experiences (Ushioda & Dörnyei, 2012). The frequent realisation of misery again highlights the need to study somatic dimensions of the work associated with cognitive tasks. The frequent realisation of interest, and the clear positive polarity overall, indicate that the stress of learning to write at a professional level in an L2 may be offset. These results are in agreement with studies of individual differences which find group-level regularities. It was not possible in this study to break these down by age, gender or other factor.

Finally, co-frequencies suggest some further elements of how students experience learning to write L2 university essays. The co-frequency of pleasure with research is a very positive result, and suggests opportunities for positively motivating students to remain engaged in the demanding work of learning to write at the tertiary level. Among these demands, the co-frequency between topic sentences and complexity highlights the sentence level as the place where discourse dimensions of the task meet the language proficiency requirements. The co-frequency of data with capacity suggests tertiary students' concepts of the character of argumentation and of strong conclusions, again offering instructors ways to enhance learner understanding of the highly structured nature of professional level writing. The co-frequency of stress with misery again highlights the need to adopt a fully somatic vision and scope for cognitive tasks, as learning cannot be artificially divorced from its physical framework and reality.

## References

- Ajzen, I. (2005). *Attitudes, personality and behaviour*. New York: Open University Press.
- Argamon, S., Bloom, K., Esuil, A., & Sebastiani, F. (2007). *Automatically determining attitude type and force for sentiment analysis* (pp. 369–373). Proceedings of the 3rd language and technology conference (LTC '07), Poznan, PL.
- Arnold, K. M., Umanath, S., Thio, K., Reilly, W. B., McDaniel, M. A., & Marsh, E. J. (2017). Understanding the cognitive processes involved in writing to learn. *Journal of Experimental Psychology: Applied*, 23(2), 115. <https://doi.org/10.1037/xap0000119>
- Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829–839. <https://doi.org/10.1038/nrn1201>
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: W. H. Freeman.
- Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effect of forced output during vocabulary learning. *Second Language Research*, 22(4), 487–497. <https://doi.org/10.1191/0267658306sr276oa>
- Bednarek, M. (2008). *Emotion talk across corpora*. New York: Palgrave Macmillan. <https://doi.org/10.1057/9780230285712>
- Bruna, K. R., & Gomez, K. (Eds.). (2009). *The work of language in multicultural classrooms: Talking science, writing science*. Routledge, UK. <https://doi.org/10.4324/9780203866979>
- Casanave, C. (2013). *Controversies in second-language writing: Dilemmas and decisions in research and instruction*. Ann Arbor, University of Michigan Press.
- Chung, Y., & Lee, J. (2001). A corpus-based approach to comparative evaluation of statistical term association measures. *Journal of the American Society for Information Science and Technology*, 52(4), 283–296. [https://doi.org/10.1002/1532-2890\(2000\)9999:9999::AID-ASII073>3.0.CO;2-5](https://doi.org/10.1002/1532-2890(2000)9999:9999::AID-ASII073>3.0.CO;2-5)
- Cresswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed method research* (2nd ed.). Thousand Oaks, CA: Sage.
- Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, 19(1), 42–68. <https://doi.org/10.1075/aila.19.05dor>
- Dörnyei, Z. (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning*, 59(s1), 230–248. <https://doi.org/10.1111/j.1467-9922.2009.00542.x>
- Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge, London. <https://doi.org/10.4324/9781410613349>
- Doughty, C. (2001). Cognitive underpinnings of focus on form. In R. Peter (Ed.), *Cognition and Second Language Instruction* (pp. 204–255). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.010>

- Eggs, S. (1994). *An introduction to systemic functional linguistics*. London: Pinter.
- Ehrman, M. E., Leaver, B. L., & Oxford, R. L. (2003). A brief overview of individual differences in second language learning. *System*, 31(3), 313–330. [https://doi.org/10.1016/S0346-251X\(03\)00045-9](https://doi.org/10.1016/S0346-251X(03)00045-9)
- Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24, 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, R. (2008). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107. <https://doi.org/10.1093/elt/ccn023>
- Fillmore, C. J., Kempler, D., & Wang, W. S. (Eds.). (2014). *Individual differences in language ability and language behaviour*. Academic Press, USA.
- Fontaine, J. R., Scherer, K. R., & Soriano, C. (Eds.). (2013). *Components of emotional meaning: A sourcebook*. OUP Oxford. <https://doi.org/10.1093/acprof:oso/9780199592746.001.0001>
- Freelon, D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5(1), 20–33.
- García Mayo, M. (Ed.). (2007). *Investigating tasks in formal language learning*. Multilingual Matters, Clevedon. <https://doi.org/10.21832/9781853599286>
- Gardner, R. C. (2006). The socio-educational model of second language acquisition: A research paradigm. *EUROSLA Yearbook*, 6, 237–260. <https://doi.org/10.1075/eurosla.6.14gar>
- Gilbert, R. (2007). The simultaneous manipulation of task complexity along planning time and Here-and-Now: Effects on oral production. In M. d. P. G. Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 44–68). Clevedon: Multilingual Matters. <https://doi.org/10.21832/9781853599286-006>
- Gries, D. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403–437. <https://doi.org/10.1075/ijcl.13.4.02gri>
- Gross, J. (2007). *Handbook of emotion regulation*. Guilford, New York.
- Guardado, M., & Shi, L. (2007). ESL students' experiences of online peer feedback. *Computers and Composition*, 24(4), 443–461. <https://doi.org/10.1016/j.compcom.2007.03.002>
- Halliday, M. (1985). *Spoken and Written Language*. Geelong, Victoria: Deakin University Press.
- Halliday, M. A. K. (1994). *An introduction to functional grammar*. London and New York: Arnold.
- Halliday, M. A. K., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd ed.). Great Britain: Hodder Education.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51, 539–558. <https://doi.org/10.1111/0023-8333.00164>
- Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, 12(3), 365–386. <https://doi.org/10.1177/1362168808089922>
- Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58(2), 285–325. <https://doi.org/10.1111/j.1467-9922.2008.00442.x>
- Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, 21(4), 390–403. <https://doi.org/10.1016/j.jslw.2012.09.003>
- Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, 62(2), 439–472. <https://doi.org/10.1111/j.1467-9922.2012.00695.x>
- Lombard, M., Snyder-Duch, J., & Bracken, C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4), 587–604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In R. Peter (Ed.), *Cognition and Second Language Instruction* (pp. 69–90). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.005>
- Martin, J. R., & White, P. R. R. (2005). *The Language of Evaluation; Appraisal in English*. UK, Palgrave Macmillan.

- Mestre, J. (2008). *Transfer of learning: From a modern multidisciplinary perspective*. Greenwich, CT, Information Age Publishing.
- Miller, A. (2018). Text mining digital humanities projects: Assessing content analysis capabilities of Voyant Tools. *Journal of Web Librarianship*, 12(3), 169–197. <https://doi.org/10.1080/19322909.2018.1479673>
- Miller, R., & Pessoa, S. (2016). Where's your thesis statement and what happened to your topic sentences? Identifying organizational challenges in undergraduate student argumentative writing. *TESOL Journal*, 7(4), 847–873. <https://doi.org/10.1002/tesj.248>
- Panther, K., & Thornburg, L. (1999). The potentiality for actuality metonymy in English and Hungarian. In K. Panther & G. Radden (Eds.), *Metonymy in language and thought*. Amsterdam, John Benjamins. <https://doi.org/10.1075/hcp.4.19pan>
- Pavlenko, A. (2002). Bilingualism and emotions. *Multilingua*, 21(1), 45–78. <https://doi.org/10.1515/mult.2002.004>
- Pennycook, A. (2014). *The cultural politics of English as an international language*. London, Routledge. <https://doi.org/10.4324/9781315843605>
- Revesz, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom - based study. *The Modern Language Journal*, 95(s1), 162–181. <https://doi.org/10.1111/j.1540-4781.2011.01241.x>
- Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In R. Peter (Ed.), *Cognition and Second Language Instruction* (pp. 287–318). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.012>
- Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27–57. <https://doi.org/10.1093/applin/22.1.27>
- Robinson, P. (2001c). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, 17(4), 368–392. <https://doi.org/10.1177/026765830101700405>
- Robinson, P. (Ed.). (2002). *Individual differences and instructed language learning* (Vol. 2). John Benjamins Publishing. <https://doi.org/10.1075/llt.2>
- Robinson, P. (2003). Attention and memory during SLA. In D. Catherine & L. Michael (Eds.), *The Handbook of Second Language Acquisition* (pp. 631–678). Oxford: Blackwell. <https://doi.org/10.1002/9780470756492.ch19>
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics*, 43, 1–32. <https://doi.org/10.1515/iral.2005.43.1.1>
- Robinson, P. (2007a). Criteria for classifying and sequencing pedagogic tasks. In M. d. P. G. Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 7–26). Clevedon: Multilingual Matters. <https://doi.org/10.21832/9781853599286-004>
- Robinson, P. (2007b). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, 45(3), 193–213. <https://doi.org/10.1515/iral.2007.009>
- Robinson, P. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (Vol. 2). John Benjamins Publishing. <https://doi.org/10.1075/tblt.2>
- Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching*, 45(3), 161–176. <https://doi.org/10.1515/iral.2007.007>
- Sawyer, K. (Ed.). (2006). *Cambridge handbook of the learning sciences*. New York, Cambridge University Press. <https://doi.org/10.1017/CBO9780511816833>
- Shin, D. S., Cimasko, T., & Yi, Y. (2020). Development of metalanguage for multimodal composing: A case study of an L2 writer's design of multimedia texts. *Journal of Second Language Writing*, 47, 100714. <https://doi.org/10.1016/j.jslw.2020.100714>
- Skehan, P. (1989). *Individual differences in second-language learning*. London: Edward Arnold.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

<https://doi.org/10.1177/003368829802900209>

- Skehan, P. (2001). Tasks and language performance assessment. In B. Martin, S. Peter & S. Merrill (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 167–185). Harlow: Pearson Education.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, 1–14  
<https://doi.org/10.1017/S026144480200188X>
- Stack, G. (2012). The language of possibility and existential possibility. *The Modern Schoolman*, 50(2), 159–182.  
<https://doi.org/10.5840/schoolman19735026>
- Swain, M. (1998). Focus on form through conscious reflection. In D. Catherine & W. Jessica (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 64–81). New York: Cambridge University Press.
- Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In B. Martin, S. Peter & S. Merrill (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Assessment* (pp. 98–118). Harlow: Pearson Education
- Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 61(s1), 37–72. <https://doi.org/10.1111/j.1467-9922.2011.00642.x>
- Thompson, E. C., & Juan, Z. (2006). Comparative cultural salience: measures using free-list data. *Field Methods*, 18(4), 398–412. <https://doi.org/10.1177/1525822X06293128>
- Tran, V., Porcher, R., Tran, V., & Ravaud, P. (2017). Predicting data saturation in qualitative surveys with mathematical models from ecological research. *Journal of Clinical Epidemiology*, 82, 71–78.  
<https://doi.org/10.1016/j.jclinepi.2016.10.001>
- Ushioda, E., & Dörnyei, Z. (2012). Motivation. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 396–409). London: Routledge.  
<https://doi.org/10.1057/9781137032829.0009>
- Whitelaw, C., Garg, N., & Argamon, S. (2005). *Using appraisal groups for sentiment analysis. Proceedings of the conference on information and knowledge management.* ACM Press.  
<https://doi.org/10.1145/1099554.1099714>
- Wickens, C. D. (1989). Attention and skilled performance. In H. Dennis (Ed.), *Human Skills* (pp. 71–105). New York: John Wiley.
- Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational Linguistics*, 30(3), 277–308. <https://doi.org/10.1162/0891201041850885>
- Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2–3), 165–210. <https://doi.org/10.1007/s10579-005-7880-9>
- Wurm, L., & Vakoch, D. (1996). Dimensions of speech perception: Semantic associations in the affective lexicon. *Cognition and Emotion*, 10(4), 409–423. <https://doi.org/10.1080/026999396380204>
- Yang, C., Lin, K. H., & Chen, H. H. (2007). *Building emotion lexicon from weblog corpora* (pp. 133–136). Proceedings of the ACL 2007 demo and poster sessions. Prague, Czech Republic.  
<https://doi.org/10.3115/1557769.1557809>

### Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).