

## Using automated and fine-grained analysis of pronoun use as indicators of progress in an online collaborative project

Kate Thompson, CoCo Research Centre, The University of Sydney, kate.thompson@sydney.edu.au  
Shannon Kennedy-Clark, Learning and Teaching Centre, Australian Catholic University, shannon.kennedy-clark@acu.edu.au

Nick Kelly, CoCo Research Centre, The University of Sydney, nick.kelly@sydney.edu.au  
Penny Wheeler, Learning and Teaching Centre, Australian Catholic University, penny.wheeler@acu.edu.au

**Abstract:** Multimodal discourse analysis has been shown to be useful in adding depth to our understanding of the processes of computer supported collaborative learning. We take one of the codes used in a multimodal coding scheme (CPACS), and apply automated data extraction techniques to a large corpus of data aimed at one code in the micro-level (pronouns). The results of this are plotted over time, and patterns of pronoun use identified for further investigation using an in-depth systemic functional linguistics (SFL) approach. Complications concerned with singular and plural second person are discussed, and patterns of pronoun use that indicate movement of the group from one phase of design work to the next are identified. Further refinement of the techniques of automated extraction are required to capture additional patterns noted in the SFL analysis.

### Introduction

Multimodal discourse analysis adds depth to our understanding of the processes of computer supported collaborative learning. The use of a coding scheme such as CPACS (Kennedy-Clark & Thompson, accepted) combines macro- and micro-level approaches to discourse analysis in a systematic way. In this paper our focus is at the micro-level, particularly the use of pronouns in online group interactions. Other work has focused on the macro components of CPACS, such as decision-making or content (see Kennedy-Clark & Thompson, accepted). We apply automated extraction of pronoun usage to a well-researched corpus of data from a month-long online group collaboration on a design task. Patterns of pronoun use were identified, and an in-depth analysis from a systemic functional linguistics perspective was applied. We related these micro-level patterns to the macro-level patterns previously identified in the data (Reimann, 2009; Thompson & Kelly, 2012).

### Background information

Without entering into the debate of the actual definition of a pronoun, personal pronouns can be viewed as a person marker (Kummerow, 2012): that is, as a part of speech, pronouns identify speaker(s), hearer(s)/addressee(s) and ‘spoken of’. One of the functions of pronouns is to establish a continuity of reference; that is, they show how locally constructed reference forms are used to construct relationships between episodes and sequences of events over time: as Kitzinger, Shaw & Toerien (2012) state, “People don’t just find themselves having a focus – rather, they create it” (p. 133). An analysis of the use of person markers is of interest in CSCL as it enables researchers to construct a visual representation of how participants arrange their actions and relationships within an online group through the choices they make in presenting themselves and others in the discourse. Traditional English grammars distinguish pronouns in terms of person, number, gender (in the third person) and case (Table 1):

Table 1: Traditional grammatical categorization of English pronouns

SINGULAR				PLURAL		
Person	Subject	Object	Possessive	Subject	Object	Possessive
First	I	me	mine	we	us	ours
Second	you	you	yours	you	you	yours
Third	he / she / it	him / her / it	his / hers / its	they	them	theirs

Systemic functional linguistics (SFL) focuses on description and emphasizes choices within a specific situation or cultural context (the ‘systemic’ in its title summarizing the concept that each instance of writing or talk or other text is a selection from a set of alternative features which are systematically related to each other by the particular language or language subset). The foundation of SFL is four connected observations on language, that language is (Eggins, 2005): (i) *functional*, that it serves a use; (ii) *semantic*, that its function is to make meaning; (iii) *contextual*, that meaning is influenced by social and cultural context; and (iv) *semiotic*, that using language involves a choice of which signifiers to use. These principles encourage an “activist bent” (Jacobs & Ward, 2000) in its practitioners and educators, who can, for example, help students examine unfamiliar

academic registers as the first step to mastery, or reveal how language enables justice to be restored or to fail to be done, although SFL analysis can also be used, as here in an examination of CSCL, for humbler purposes.

In SFL, personal pronouns can be analyzed as key features in the ‘interpersonal’ metafunction of the text, “enacting our personal and social relationships with the other people around us” (Halliday & Matthiessen, 2004, p. 29), and as linked to other interpersonal components such as the Mood element (which in English is made up of the Subject and Finite operator of a clause, and realizes the declarative, interrogative, or imperative mood of a clause); modality (the uncertainty or possibility in a clause, for example ‘can’ or ‘must’), and the role of clauses in offering and exchanging services and information. The other two linguistic metafunctions in SFL theory are the ideational, which looks at the categories and processes being identified and experienced by the speakers, and the textual, those parts of a text that provide cohesion and organization for the discourse as a whole. Personal pronouns can be examined for the agency that they realize in the clause, and also for their cohesive role in providing a continuous thread of reference to the participants in the conversation, written text or, as here, computer-mediated chat.

Word counting has been used in text analysis in the social sciences for many decades. Two notable examples are DICTION (Hart, 2001) and LIWC (Chung & Pennebaker, 2012; Pennebaker & Francis, 1996), both of which use pre- and user-defined dictionaries to count and subsequently analyze corpora of text for features such as attentional focus, emotionality, social relationships, thinking styles, and individual differences (Tausczik & Pennebaker, 2010). Natural language processing has significantly advanced over the past few decades (Jurafsky & Martin, 2008; Manning & Schütze, 1999), and advanced techniques have recently begun to be applied to the learning sciences, such as the mapping of words in vector spaces to discover use of concepts through clustering (Sherin, 2012) and the use of parts-of-speech (PoS) tagging for segmentation and classification. In this work we make use of the Python programming language (Bird, Klein, & Loper, 2009) to apply some of these techniques to our analysis of pronouns used during CSCL. We use these tools as aids in reconstructing the contextual use of person markers in relating to other members of the group, before more closely examining their functions in this context via a grammatical analysis based on systemic functional linguistics.

## Methods

The corpus of data used in this analysis is the online chat record of a group of four postgraduate students (one male, three females) collaborating over one month on a design task. The task was to identify an existing system dynamics model and turn it into an educational resource both by adding to the model and providing materials to accompany the model. Students had access to a chat and a wiki to support their work. Previous research has described the data collection (Reimann, Weinel & Thompson, 2007) and the patterns of decision-making in the chats (Reimann, Frerejean & Thompson, 2009) as well as in the wiki (Thompson & Kelly, 2012). This paper returns to focus just on the chat data. The collaborative exercise took place over one month in 2006; students met as a whole group on eight occasions, and 2182 utterances were used in the automated process.

The Collaborative Process Analysis Coding Scheme (CPACS) was used to code the synchronous chat data. (For a full description of the coding scheme, see Kennedy-Clark & Thompson, accepted). This coding system considered several SFL phenomena, such as those described by Halliday (1994) and elaborated upon by Martin and Rose (2007). CPACS functions on two levels. At the macro level, the action code concerns goal identification and solution, and the content code covers the content of the utterance. At the micro level, the attitudinal code describes the type of attitude each utterance was taking towards the problem-solving design; the tense is the marker of temporality; the modality is the degree of certainty; and pronouns relate to personal references. In this study we use the pronouns code: 1 (first person singular); 2 (second person singular); 3 (third person singular); 4 (first person plural); 5 (second person plural); and 6 (third person plural).

We hypothesized that by combining the macro elements of discourse (e.g. Poole & Holmes, 1995) with the micro or grammatical levels that are often the target of discourse analysis (e.g. Nivre et al. 1999) we could obtain a better understanding of the conversations than if we analyzed the macro levels only, using a coding system such as the Decision Function Coding System (Poole & Holmes, 1995) as we had done in our earlier analysis (see, for example Kennedy-Clark & Thompson, 2011; Thompson, Kennedy-Clark, Markauskaite, & Southavilay, 2011). We found that by drilling down into the discourse and examining the conversations at a finer grain size we could get a detailed understanding of the function of language in establishing roles and responsibilities within a group.

Naturally occurring classroom language and patterns of discourse produced by the students were mined for linguistic and social insights about how the online groups managed their shared task through the language that they used (Gee, 2005; Schallert, et al., 2009). SFL analysis was carried out by considering correspondences in the data between the variables of DFCS, personal pronoun use, and author, and investigating the language in use at this point. To carry out these analyses, one researcher worked on the utterances as filtered by the quantitative analyses, and provided additional columns of annotation on specific linguistic aspects, namely: modality; mood of the clause, that is, imperative, interrogative, or declarative; annotations to the automated

pronoun identification data, where, for example, the author has omitted the Subject that should be understood and produced an elliptical utterance, particularly by omitting a first person singular pronoun; cohesion, by marking the chain of reference for each neutral third person pronoun used; and process type, from the three principal categorizations in SFL of material; mental; and relational processes.

These general categorizations are ones that are immediately evident to the reader, and these types of annotations are rapidly made: by working directly on the filtered data source, the researcher can compare the occurrence of different linguistic aspects within the target variables, counting occurrences and looking for 'marked' clauses, that is, linguistic constructions which show a non-typical structure. This level of human linguistic analysis was possible because the preceding automated analysis and visual interpretation of the quantitative aspects of the linguistic data reduced the number of segments and sequences that require individual assessment, and because the categorizations are at the most general level of SFL analysis. Aspects of appraisal and analysis of circumstantial elements (such as realizations of location, manner, or cause), both of which can provide deep insights into interactions and experiences as represented in more highly structured oral and written texts, did not form part of this SFL analysis.

## Analysis

### Automated extraction of data and identification of patterns

Two phases of automation were used in the analysis of this corpus of chat text<sup>(1)</sup>. The first involved counting pronouns, whilst the second used PoS tagging to assist in understanding the word counts. Pronouns function as both person markers and indications of focus. A count of the pronouns uttered by each author gave an overview of how the author positioned their own experience of the collaboration, whom they addressed during the group exchanges, and when they addressed entities outside the exchange. The count was conducted using a Python script to iterate through chat text and count each utterance. The pronouns counted were each assigned to a CPACS pronouns code: 1 (my, me, mine); 2 (you); 3 (her, him, it, he, she); 4 (us, we, ours) and 6 (they, them). The result of this analysis was a count of each CPACS code for each author. A limitation of this automated approach to pronoun counting is that it loses the context of the utterance. This means that it always assumes that the use of a pronoun by a participant is in their own voice, when in some cases it may be a quote of something that another participant or third party has said. An analysis based upon use of the words 'said' and 'says' shows that this use of another voice occurs 23 times in the corpus.

Further analysis was possible by breaking these numbers down in two ways: (i) seeing how they changed over time; and (ii) seeing how they changed within context. Students met in eight separate chat sessions and separating the count of pronouns in each session allowed for the analysis of changes that occurred over time. In addition, each utterance within this corpus of text had previously been grouped into five tasks: choosing, adding, implementing, overall and coordinating (Reimann, Frerejean & Thompson, 2009). The counting of pronouns was indexed to these categories, allowing for an analysis of the changing use of pronouns within these different contexts.



Figure 1: visualization of pronoun use for each participant, for each session

Figure 1, in combination with results from coding and counting not presented here, was used to identify instances of variation. Figure 1 shows the percentage of pronoun use (the CPACS code is along the y-axis,

session number along the x-axis) for each member of the group. Each section (for example, between 1 (me, my, our) and 2(you)) represents 100% of the pronouns extracted. An issue with second person is that the same word is used for both singular and plural, and so all second person was recorded as '2': this is the first complication to be discussed below. Five patterns were observed during examination of the extracted pronoun data. (1) Pronouns and tasks: the combined percentages were different when examined across tasks (not presented in Figure 1). Higher percentages of use were observed in coordination and implementation than choosing or additions. Within coordination, Charles and Jane used this more than other members, while Beth had the highest proportion in implementation. (2) Use of first person changes over time: if we examine the patterns across time, within each pronoun code, it can be seen that the use of first person, while remaining quite consistent, differs amongst the group members. (3) The task of choosing had a different distribution of pronouns than other tasks: coding and counting results indicated that with respect to task, choosing had a different distribution amongst team members than other tasks. (4) Session 5 was different to other sessions with respect to pronoun use: Figure 1 shows a different distribution amongst team members in Session 5. (5) First person plural is used differently over time: Sessions 7 and 8 show a slightly lower percentage across all group members than other sessions. The use of third person showed a higher proportion apparent when participants were discussing implementation.

### **'You' – plural or singular?**

PoS tagging (Mu, et al., 2012; Schmid, 1994) was utilized in the deeper analysis of the word 'you' within the corpus. Because English no longer has a singular and plural form of the second personal pronoun, a simple count of the occurrences of 'you' does not distinguish between the author's focus upon one person or upon some or all of the people in the exchange. A Python script was used to apply PoS tagging to the corpus<sup>(2)</sup>, to produce an output consisting of, each time the word 'you' is uttered; the two words before and after this utterance of the word 'you', and the PoS indicated. The assumption in doing this was that the PoS tagging would reveal patterns that could allow for automation of the context of the word 'you', and it was partially successful in this, highlighting two linguistic features particularly relevant to the successful conduct of computer-mediated discussion: a marked use of the vocative, and the conversational workarounds that English speakers use to indicate plural reference for the 'you' pronoun. In the latter case, the PoS tagging clearly revealed the uses of 'you both' (1 instance), 'you all' (4 instances), 'you guys' (25 instances: note this reference is non-gender-specific) as well as 'you girls' (6 instances) and 'you ladies' (4 instances), both of these authored by the sole male speaker within the group of 4 collaborators, and all of these examples of the author disambiguating the 'you' pronoun to indicate all addressees in the context of the discussion in order to check for their consent, for example, or direct their action. (In subsequent human analysis, reviewing the PoS tagging output, an additional refinement of this distinction between 'you' singular and plural was identified in the use of 'u', a text-chat specific variant of the pronoun, which was used by one author as a singular 'you' reference only.) The PoS tagging also highlighted where authors were employing the vocative as a means to target their utterance to only one of their collaborators: a few examples of this linguistic choice, one frequently found in this corpus, are 'jane, do you have that link handy?' (segment 14); 'what d'ya say beth and jane?' (98); 'beth do you what [want] to start with your questions?' (2958); 'lydia, jane, what about you guys?' (5065); 'beth you mentioned you had some resources' (7743); and 'charles what do you think about beths question?' (6212).

This marked usage of the vocative in CSSL is paralleled by a similarly marked rejection of the use of personal pronouns in favour of names as the subject or object of a clause. For example, in standard conversation or writing, the context often makes unnecessary the specific naming of your partner in an action: it is the de-contextualised nature of CMD that makes it preferable to choose 'lydia and me' instead of standard 'we' in an utterance such as 'In that way Lydia and me should find additional resources' (2989) or to use a name rather than a third person pronoun to emphasize to which person's utterance your response refers: 'my that's right is with charles answer' (6510) or to reject 'you', when Jane and Lydia are both present, in favour of 'we need to see if jane or lydia want to add anything / amalgamate anything' (7279).

### **Guiding behavior**

An important way in which personal pronouns provide a shortcut to an analysis of what is going on in the discussion is by indicating the kind of exchanges occurring in the text, and the salient personal pronoun is 'you'. Figure 1 shows that both Charles and Jane are strong users of 'you', particularly in segments that are coded as overall and coordination phases in the DFCS. SFL sees an analysis of what is being 'exchanged' by these 'you' clauses as a fundamental indication of the roles and statuses adopted by the authors of these clauses.

Clauses can be structured as an exchange of information or as an exchange of goods-and-services: most simply, the choices can be represented as giving information (making a statement) or giving goods or services (making an offer); or as demanding/requesting information (questioning) or demanding goods or services (issuing a command, also discussed in the previous section). An example of how clauses function in these exchanges can be seen in Table 4.

Table 4: Utterances from Session 5, Jane's use of 'you' as part of overall decisions (planning, trouble-shooting)

Author	Clause	Role in exchange	'You' refers to:
Jane	did you tell lydia about what Kate told you guys ?	question – asking for information	'you' (here, Sayer and Subject), 'you guys' (here, Audience) = Beth and Charles
Jane	have you guys seen the storytelling feature how it works?	question	you guys (Subject) = Beth, Charles and Lydia
Jane	beth you were interest in seeing how storytelling works see the immigration and Easter Island models	statement, command ('see the ... models')	you (Subject) = Beth
Jane	but I don't see how lydia and I help you guys	==	you guys = Beth and Charles
Jane	I could work with you charles	offer	you = Charles

It is clear, in tracing the referent(s) of the 'you' pronoun, that additional mapping beyond pronoun counting and PoS tagging would be required to precisely outline who was working with whom, but what is evident from looking at the clauses as exchange is the role that the author can variously assign herself/himself.

Another example where the distribution indicates a strong use of 'you' is Charles (see Table 5), in coordination mode, requesting goods-and-services from others in the group. It appears that his earliest demands on his group members (here, Beth) are couched as a request rather than a requirement, but he does not use this strategy in later sessions.

Table 5: examples of Charles' use of 'you' in coordination

Session	Clause	Role in exchange
1	sure beth. are you going to link it in the wiki?	Structured as a request for information, this is actually functioning as a command (asking for Beth's action).
2	jane, do you want to put in your response to beth's questions first	Request for information in structure, by implication a command for action, an instruction
2	if you like set up a forum.	Suggestion: a command ('set up a forum') modulated by a concessive clause
2	jane, can you also archive the notes	Command structured as a yes/no question
3	for sure beth! you get some sleep girl!	Command
4	ok. first let me tell you that you have made a huge contribution to the team	Both a command (for 'you') and an offer (from 'me'): 'let's'
6	ok. jane. were you ok about tomorrow night at 10:30?	Request to attend, structured as a yes/no question
7	can you give me a link?	Request for action

The examples shown in Table 5 are in fact not the typical clause structures for realizing a command, as they are marked for person. Usual or unmarked imperatives cannot be identified in an automatic count of pronouns, as they do not contain the Subject component of the clause: take, for example, the instances of unmarked imperatives from Charles in the first session (see Table 6).

Table 6: Imperative clauses (unmarked): author, Charles, coordination

Session	Clause	Role in exchange
1	check page now	Command
1	bullets in now = refresh	Command (= 'Now, refresh!')
1	check now!	Command

### Making suggestions

We can extend the discussion of the clause as exchange by examining a marked difference in first person pronoun use within the 'choosing' component (DFCS coding) for team members than in the other DF tasks. Using SFL tools to probe this more closely, we can identify the linguistic choices accompanying this first person pronoun use (chiefly, the use of 'I' or 'we' in focus or thematic position) as contributing to the successful use of 'suggestion'.

There are two choices that the group must make to progress with the collaborative task: the selection of model, and the selection of audience or class level. The selection of model comes first, and, it being early in their joint history (completely contained in Session 1), group members are initially unwilling to baldly assert a choice: instead, they make suggestions. The linguistic choices required for successful suggestions are seen by SFL as part of the interpersonal resources of language, and, matching pronoun use to DFCS choosing, and following this with human analysis, we found that strategies used in suggestion and negotiation include a modulation in the Finite verb used when 'I' was the Subject (for example, using modals with a lower representation of certainty ("Charles: i agree. having a look at the some the models we might just need to pick the one that best suits us"): mitigation of the imperative mood (that is, softening the command, as in "Beth: not yet, but come prepared with one each and then vote, i suppose" (segment 9); or, "Beth: how about we give each other a day or two" (segment 3); and omission of first person singular pronouns to conceal who is the Actor, in SFL terms, in these processes (as Charles does repeatedly as he reports his additions to the wiki, perhaps concealing his dominant role here: "Charles: just updated that element"; "Charles: just put in the link to the October 5<sup>th</sup> archive", and so on).

### Creating active artifacts

A higher pattern of third person usages within segments coded as implementation warrants a closer look at these occurrences. The English third person singular pronoun distinguishes its referents by genders, and by simply using the filters in the pronoun count view, it is apparent that the human referents for these pronouns are straightforward: 'he' = the supervising male teacher for the activity, 'she' = the female teacher or, in one sequence, one group member who was temporarily absent. More frequent is the neuter pronoun 'it', and, in those utterances related to the task of implementation, the referent is the group's collaboratively created digital artifact and/or the task of creating this artifact. For example, when Beth directs the group's attention to the visuals and sound effects that she has added to the model in the wiki, they respond:

Jane: **it's** very good, girls (10238)

Charles: will just open **it** now (10239)

Beth: **it's** a little rough...i need to adjust **it** now i've spoken to Kate (10240)

Beth: and then **it** goes into storytelling mode... and by pressing the space bar, the 'story' appears step by step according to the sequel in trac (10252)

Lydia: the female learners will **luv it** yah :) (10253)

By this stage in the operation of the group, the design task is getting close to completion, and this is indicated by the increased reference to the model they have worked on as one of the 'participants', that is, in SFL transitivity terms, one of the entities involved in the events or processes being referred to, and its emergence as a possible subject and Actor (as in 10252).

### Contributing independently and within the group

The pronoun count indicates a decrease in use in first person plural in the later sessions, and a follow-up by a human reader indicates that in part this is because the participants have organized themselves into pairs, and report on the pair's activity as, for example, 'lydia and me'. The addition of a broad classification of process types (the major categories of SFL transitivity analysis, that is material, mental, verbal and relational processes) associated with the first person plural pronoun reveals that the actions the group is depicted as taking are not often material, and certainly not physically effective processes. For example:

Charles: we did some refinements [Material]

Charles: we came up with a "test some values and come up with something better as an activity basically [Material]

Charles: can we get away with two? [Material]

Charles: I think we can probably whip another scenario up pretty quickly. [Material]

Lydia: shd we add a short story to explain the stock and flow concept? [Material]

These clauses with Material processes are most importantly realizing the actions of the group members in *creating* or extending the enhanced model: 'came up with'; 'whip up'; and 'add'.

In clauses with relational processes, the pronoun 'we' is more likely to be referring to the group of four participants as a whole (Table 7). The relational process subsystem of 'possession' is realized strongly here, as well as the idea of representation, an identification process again related to the idea of creation. Metaphors of process here are relational processes realizing mental states: 'got any ideas' = 'thinking'; 'are we on the sane [same] page' = 'agree'

Table 7: Examples of relational and verbal processes from the chat

Author	Clause	Process type
Jane	I think we all have that impresion	relational
Charles	we need the stocks to be different for each scenario.	relational
Lydia	if we represent the scenarios with graphs,	relational
Jane	in trac we have limited space to upload files	relational
Jane	charles are we going to have 1 or two scenarios	relational
Charles	got any ideas for it?	relational
Lydia	are we on the sane page?	relational
Charles	we are nearly there!	relational
Jane	should we talk about our progress ?	verbal
Beth	lydia, we need to talk about the text	verbal
Charles	so we still agree that [we get them done on Wednesday]	verbal
Lydia	is there anything else we need to discuss?	verbal
Charles	we chat more tomorrow	verbal

By contrast, the first person singular pronoun is extensively employed in these last two sessions, particularly in segments coded as ‘implementing’ in the DFCS, and this may indicate that the group members are operating independently, having negotiated particular sets of tasks during the earlier sessions of choosing, planning and coordinating. Looking at the processes which the group members self-describe themselves as performing in broad SFL categories, it is evident that ‘I’ is associated with more effective material actions. Note the over-representation of Beth in using ‘I’: she has a significant share of the utterances coded as implementation and particularly utterances using ‘you’ as the object of a request for help. The other types of processes done by ‘I’ in these sessions are, in the segments coded for implementation, evenly split between mental, verbal, and relational, with just a few instances from each category.

## Conclusions

This initial combination of methods of analyzing this data has indicated areas for follow-up research in other corpuses of data. There is some evidence provided here that changes in patterns of pronoun use could indicate changes in phases of design work (from choosing to additions, and also the beginning of implementation). In several instances, automation did not identify the patterns identified by human raters. There is significant scope for further application of automation to systemic functional linguistics in education and refinement of this work, as was carried out with ‘you’, and reported in this paper. One of the key areas for development is in the calibration of the diverse methods available with human input. Once we know what patterns to look for, we will be better able to identify automated techniques that could be used to help students manage their own progress, as well as instructors to monitor the progress of groups through their collaborative work. Future work too needs to tie these findings more explicitly to the macro-level of discourse analysis.

## Endnotes

<sup>1</sup> All scripts used in the analysis were written in the Python programming language and are available by request

<sup>1</sup> The Python package Natural Language Toolkit was used for the PoS tagging, specifically the algorithm trained on the Wall Street Journal corpus of over 270 000 words (Bird, et al., 2009)

## References

- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*: O'Reilly Media, Incorporated.
- Chung, C. K., & Pennebaker, J. W. (2012). *Linguistic Inquiry and Word Count (LIWC): Pronounced “Luke,” and Other Useful Facts Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 206-229): IGI Global.
- Eggins, S. (2005). *Introduction to systemic functional linguistics*: Continuum.
- Gee, J. P. (2005). *An Introduction to Discourse Analysis: Theory and method* (2nd ed.). London: Routledge.
- Halliday, M. A. K. (1994). *Introduction to Functional Grammar*. London: Arnold.
- Halliday, M. A. K. & Matthiessen, C. (2004). *An Introduction to Functional Grammar* (3rd ed.). London: Arnold.

- Hart, R. P. (2001). Redeveloping DICTION: theoretical considerations. *Progress in communication sciences*, 43-60.
- Jacobs, G. M., & Ward, C. (2000). Analysing Student-Student Interaction from Cooperative Learning and Systemic Functional Perspectives. *Electronic Journal of Science Education*, 4(4). Retrieved <http://ejse.southwestern.edu/article/view/7639/5406>
- Jurafsky, D., & Martin, J. (2008). *Speech and Language Processing* (2nd Edition) (Prentice Hall Series in Artificial Intelligence): Prentice Hall.
- Kennedy-Clark, S., & Thompson, K. (2011). What Do Students Learn When Collaboratively Using A Computer Game in the Study of Historical Disease Epidemics, and Why? . *Games and Culture*, 6(6), 1-25.
- Kennedy-Clark, S. & Thompson, K. (accepted) Using multimodal discourse analysis to identify patterns of problem solving processes in a computer-supported collaborative environment. In Sudweeks, F. & Lim, H. (Eds) *Innovative Methods and Technologies for Electronic Discourse Analysis*.
- Kitzinger, C., Shaw, R., & Toerien, M. (2012). Referring to Persons Without Using a Full-Form Reference: Locally Initial Indexicals in Action. [Article]. *Research on Language & Social Interaction*, 45(2), 116-136. doi: 10.1080/08351813.2012.673376
- Kummerow, D. (2012). The Person That Isn't: On Defining the Third-Person, Negatively and Positively So. *Australian Journal of Linguistics*, 32(2), 259-290. doi: 10.1080/07268602.2012.669099
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*: MIT press.
- Martin, J. R., & Rose, D. (2007). *Working with Discourse: Meaning Beyond the Clause* (2nd ed.). London: Continuum International.
- Mu, J., Stegmann, K., Mayfield, E., Rosé, C., & Fischer, F. (2012). The ACODEA framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International Journal of Computer-Supported Collaborative Learning*, 7(2), 285-305. doi: 10.1007/s11412-012-9147-y
- Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion*, 10(6), 601-626.
- Poole, M. S., & Holmes, M. E. (1995). Decision Development in Computer/Assisted Group Decision Making. *Human Communication Research*, 22(1), 90-127.
- Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 4, 239-257.
- Reimann, P., Frerejan, J. & Thompson, K. (2009). Using process mining to identify models of group decision making in chat data. Paper presented at the CSCL Conference, Rhodes, Greece.
- Schallert, D. L., Chiang, Y.-h. V., Park, Y., Jordan, M. E., Lee, H., Janne Cheng, A.-C., Song, K. (2009). Being polite while fulfilling different discourse functions in online classroom discussions. *Computers & Education*, 53(3), 713-725. doi: 10.1016/j.compedu.2009.04.009
- Sherin, B. (2012). Using computational methods to discover student science conceptions in interview data. Paper presented at the Proceedings of the 2nd International Conference on Learning Analytics and Knowledge.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24-54.
- Thompson, K., & Kelly, N. (2012). *Processes of decision-making with adaptive combinations of wiki and chat tools*. Paper presented at the The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012).
- Thompson, K., Kennedy-Clark, S., Markauskaite, L., & Southavilay, V. C. (2011). Capturing and analysing the processes and patterns of learning in collaborative learning environments. Paper presented at the 9th International Conference on Computer Supported Collaborative Learning. CSCL2011: Connecting computer supported collaborative learning to policy and practice, Hong Kong.

## Acknowledgments

Part of this work was funded by the Australian Research Council grant FL100100203.