The Hierarchy of Detective Fiction: A Gramulator Analysis

Travis A. Lamkin and Philip M. McCarthy

Department of English, University of Memphis
Memphis, TN, USA
{talamkin, pmmccrth}@memphis.edu

Abstract

Closely related genres have complex interrelations. An antecedent genre can constrain a subsequent genre, but changing rhetorical situations can lead to distinctions between an antecedent and its descendent. In this study, we assess two genres of detective fiction to determine their hierarchical relation to one another. We use the Gramulator, a computational tool that identifies indicative lexical features, to explain the relationship between *whodunit fiction* and *hardboiled fiction*. We conclude, based on the indicative lexical features of the expositions in texts, that the two are sibling genres.

Introduction

This study assesses the heterogeneous and hierarchical relationship between two genres of fiction. More specifically, we are interested in the types and distributions of the lexical features that are indicative of *whodunit fiction* and *hardboiled fiction*. Whodunit and hardboiled are generally considered sister terminal nodes of the parent node *detective fiction*, itself a child of *crime fiction* (Rzepka 2005). Additionally, because whodunit is an antecedent to hardboiled, it is possible that whodunit constrains and influences hardboiled (Jamieson 1975). However, Jamieson notes that when the communicative context changes, the antecedent's rhetorical responses are not longer relevant. As such, in this study, we address the question "is hardboiled fiction a sister-genre or sub-genre of whodunit fiction?"

Genres are heterogeneous. Even while a text (as a whole) may be recognized as *of a genre*, many of the features that compose that text, and presumably make the recognition possible, are often indicative of other, quite different genres (McCarthy et al. 2009; McCarthy 2010). Such heterogeneity is only to be expected; after-all, different genres are not different languages; and so, genres are not simply denoted by an instance of indicative language, but instead reflect the "presence, prevalence, and prominence" of numerous exemplars: many of which are no doubt ubiquitous (Davies and Elder 2004; McCarthy 2009).

But genres aren't only heterogeneous (i.e., replete with numerous examples of linguistic features that may be more common in other genres). We also have to consider that genres are commonly organized into hierarchies. For example, Karlsgren and Cutting (1994) analyzed *the mystery novel* as a subgenre of *fiction*; which in turn was considered a subgenre of *imaginative prose*. Thus, the heterogeneity of a genre needs to be considered both horizontally (i.e., across sibling genres) and vertically (i.e., across parent and child genres).

Complicating genre classification further is relative prevalence. We have to consider that a greater prevalence of Genre X than Genre Y in Text T does not entail that Text T is a member of Genre X. For example, science texts often contain a greater raw frequency of indicative narrative features than they do of indicative science features (McCarthy 2010). Thus, genre can also be identified by the prominence of a less frequent element. This recognition, presumably, allows us to consider some genres as subgenres, rather than genres in their own right. Thus, the subgenre of *spy fiction* (e.g., *Smiley's People*) is considered a subgenre of *crime fiction*, into which also fall the subgenres of *police fiction* and *detective fiction* (Priestman 2003; Rzepka 2005).

Understanding the diverse compositionality of genres may lead to improved computational tools, search facilities, and textual evaluations. It may also play a role in reading development because, apparently, the balance of genre features in a text is an important factor of genre recognition. And if a text does not contain an appropriate mix of genres, it is possible that readers' comprehension may be negatively affected when processing the text, presumably because certain expectations are not met. (Bhatia 2002; Graesser, Olde, and Klettke 2002).

Detective Fiction

Detective fiction began in the 19th century in the United States and England. At the center of these texts is always a crime and usually a murder, and the solution to the crime is either beyond the ability of the police or beneath their interest. The detective's role is to explain what is

unexplainable by any other character, through brilliance or moral superiority. Merely including a detective is not enough to identify a text as detective fiction. The focus must be on the detective solving a mystery (Rzepka 2005).

Detective fiction includes two prominent subgenres, which are the subject of the current study: whodunit and hardboiled. Chronologically, the whodunit genre preceded hardboiled, beginning with Poe's "Murders in the Rue Morgue" in 1841 (Palmer 2001). By the 1920s, a perceived lack of realism in whodunit helped to give raise to the hardboiled genre. Despite the emergence of hardboiled fiction as a reaction to whodunit, most taxonomies of detection fiction view whodunit and hardboiled as sister genres (Rzepka 2005, Thomas 2005). However, as suggested by Jamieson (1975), antecedent genres (here, whodunit) can constrain and influence subsequent genres. Thus, it is reasonable to hypothesize that hardboiled fiction may be a subgenre of whodunit rather than a sibling.

Assessing the relationship between whodunit and hardboiled is challenging because here are any number of theoretical differences between the two genres. For example, the whodunit detective has "genius," whereas the hardboiled detective has "unshakeable morality." As such, in this study, we begin and at the beginning, and focus on the scene setting featured in the exposition (or opening) of the stories.

Setting is clearly one of the major differences between whodunit and hardboiled. In whodunit fiction, the setting is the "locked room." Sometimes the setting is literally a locked room, but it can also be a quiet manor in the countryside, or the Orient Express. What is important for the setting is confinement; the murder, the clues, the suspects, and the investigation itself cover a defined and limited space into which there is no intrusion from the outside world. In contrast, the major innovation of hardboiled was to bring murder and violence into a real world setting, "out of the drawing room and into the alley" (Chandler 1985). As McManis (1978) argued, authors wanted to present a familiar, believable setting, something quite different from a train traveling through 19th century Europe or an isolated island in the Baltic Sea.

Contrastive Corpus Analysis

Contrastive Corpus Analysis (CCA) is the name given to the method through which the meaningfulness of lexical features is generated as a process of relativity. The principle of CCA is that any discourse unit (e.g., text-type, register, genre, variety, or section of text) is best understood, and perhaps only understandable, within the context of its contrast to some other discourse unit (McCarthy, Watanabe, and Lamkin in press). CCA typically compares two corpora that are closely related while remaining theoretically distinguishable, even if the points of distinction have yet to be identified. For this study, the discourse unit under analysis is the corpus, and

specifically the corpora of whodunit and hardboiled detective fiction.

Discourse units (here, corpora), can be analyzed at any discourse level (e.g., word, sentence, construct); however, in this study, we followed Henry and Roseberry (2001), Hopkins and Dudley-Evens (1988) and Upton and Conner (2001) who argue that *moves* (Swales 1990) are a useful form for classifying genres. Moves are functional components of a text such as *introducing a character*, *setting a scene*, or *providing background*. Given our interest in the expositions of the fiction, an analysis at the move level is clearly appropriate.

Moves are situated in linguistic features (formulaic words, phrases, constructions, etc.), and by identifying those linguistic features, we aim to identify the features that distinguish one genre from the other. As such, this study also builds on research by McCarthy (2010), and Herring and colleagues (2006). These studies demonstrate that computational tools can be used to assess genre relationships between two corpora. If the computationally-identified linguistic features from whodunit identify different moves from the features in hardboiled, we will evidence that the two genres are different.

Corpora

To address our research question, we formed two contrastive corpora: whodunit and hardboiled. The whodunit corpus is composed of 210 texts, and the hardboiled corpus is composed of 250 texts. For this study, we focused entirely on the expositions (i.e., openings) of the texts. Naturally, texts are of varying lengths, so the extent of the exposition may vary greatly. As such, we approximated the openings by following Biber (1989) and selecting the first 2,000 words of each text, ending at the nearest end of a sentence, and excluding any prologue. In the whodunit corpus, there are 16 authors. In the hardboiled corpus there are 32 authors. This ratio of authors is representative of the genres, as whodunit authors tend to be prolific, and hardboiled authors less so (Holquist 1971). All texts were gathered in January 2010 from available digital texts.

The Gramulator

Recent developments in computational linguistics and discourse processing have made it possible for researchers to develop a wide range of sophisticated techniques for textual analysis. Many of these techniques have been gathered together in a single tool called *the Gramulator*. The Gramulator is a freely available qualitative and quantitative computational textual analysis tool. It is designed to allow researchers and materials designers to identify indicative lexical features of texts and text types. More formally, the Gramulator is designed to identify

differential linguistic features of correlative text types (McCarthy, Watanabe, and Lamkin in press).

The Gramulator processes text using a variety of frequency weighted n-grams. Bigrams are generally the primary level of analysis, although other forms such as trigrams and quadgrams can be requested. Typically, the most highly rated n-grams (in terms of weighted frequency) do not differ greatly from corpus to corpus (e.g., *of the* is either the highest or near highest ranked bigram in most analyses). As such, highly ranked n-grams are seldom diagnostic in a contrastive corpus analysis because they are highly common to both corpora. Despite their own lack of diagnostic ability, the highest ranked n-grams help to identify the *differentials* or *indicative features* of a corpus. Differentials is the name used to describe those word clusters that are indicative of the text type under analysis, relative to the text type being compared. For example, *the body* is indicative of whodunit fiction, relative to hardboiled, whereas *the gun* is indicative of hardboiled relative to whodunit. Although parameters can be adjusted, in this study, we set differential as those n-grams that are among the most highly ranked in one corpus (i.e., among the top 50% of n-grams) but are uncommon to the contrasting corpus (i.e., *not* among the top 50%). The differentials are derived following the principals of machine differential diagnostics (Rahati and Kabanza 2010). Specifically, the Gramulator disregards all lexical features that are highly common to both corpora (called *reciprocals*). Reciprocals are characteristic of both corpora, therefore they are diagnostic of neither. The remaining n-grams are the most highly ranked n-grams that are present in just one of the two corpora (McCarthy et al. in press). These n-grams are used collectively as an index, in t-tests and other statistical procedures; and individually (or in groups) to establish traits of the text type, such as setting, character, or temporal features.

Other textual analysis tools such as Coh-Metrix (Graesser et al. 2004) and LIWC (Pennebaker and King 1999) are undoubtedly useful and have a rich history in genre categorization (Crossley, Louwerse, McNamara 2009). However, those tools tend to focus on predefined measures of psychological constructs such as coherence and anxiety; whereas we are primarily interested in the *function* of the linguistic features that define and distinguish genres relative to each other. Thus, for identifying indicative linguistic features, the Gramulator in uniquely suited to the current study.

The Gramulator comprises eight modules, two pre-processing and six post-processing modules. The two pre-processing modules are *the Cleanser* and *the Sorter*. The cleanser was used to help clean the texts of extraneous information, whereas the Sorter was used to randomly divide the corpora into training and test sets for validation purposes. For this study, we also used *the Evaluator* and *the Concordancer* from the post-processing modules. The Evaluator calculates incidence values of the derived indicative features. As such, this module of the Gramulator is similar in function to the Coh-Metrix and LIWC tools, although unlike those tools, the evaluator is also designed to perform basic statistical analysis such as t-tests. The concordancer allows indicative features to be viewed in context and compared to both corpora under analysis. The concordancer is designed to incorporate Fisher's Exact test for relative comparisons of frequencies.

Results

Using the Sorter module, we randomly divided each group of texts into a *training* set and a *test corpus*. Each whodunit set contains 105 texts, and each hardboiled set contains 125 texts (see Table 1). We tested each test corpus (e.g. WD2 and HB2) against each of the training indices (i.e. WD1 and HB1). We assessed the results of each test with a within-texts t-test. The results were as predicted and validated the indices: The whodunit test texts were better explained by the whodunit training index than by the hardboiled training index. Similarly, the hardboiled test texts were better explained by the hardboiled training index than by the whodunit training index. In Gramulator nomenclature we express these assessments as follows: WD2→WD1(HB1): $M = 2.637$, $SD = 0.712$, WD2→HB1(WD1): $M = 2.386$, $SD = 0.882$; $t= 2.237$, $p= 0.027$, Cohen's $d =1.762$; HB2→HB1(WD1): $M = 4.281$, $SD = 1.280$, HB2→WD1(HD1): $M = 1.681$, $SD = 0.482$; $t= 21.203$, $p<0.001$, Cohen's $d=2.682$ where 2 denotes a test corpus; and 1 denotes a training set.

Table 1: Description of the corpora

Corpus	Training	Test	Avg. Words	Authors
Whodunit	105	105	2060.4	16
Hardboiled	125	125	2015.6	32

Having validated the indices, we turned to determining whether hardboiled fiction is a sub-genre of whodunit fiction. Using the Evaluator module of the Gramulator and a series of t-tests, we tested the hardboiled test corpus (HB2) using the whodunit index (WD1) and the whodunit test corpus (WD2) using the hardboiled index (HB1). In Gramulator nomenclature we express these assessments as follows: HB2→WD1(HB1) and WD2→HB1(WD1). The purpose of this test was to determine which corpus best explained the values generated from the contrasting index. Our results suggest that the whodunit test corpus is better explained by the hardboiled index: WD2→HB1(WD1): $M = 2.386$, $SD = 0.882$; HB2→WD1(HB1): $M = 1.681$, $SD = 0..482$; $t= -7.798$, $p= <0.001$, Cohen's $d= -0.723$).

To supplement our analysis, we conducted a logistic regression on the data sets. The results suggest that each of the indices are significantly predictive of the two test corpora ($p < 0.001$), but neither index is significantly more

predictive than the other. These results provide evidence that whodunit and hardboiled are more equally related than the *t*-tests (see Table 2).

Table 2: Binary logistic regression results

		B	S.E.	Wald χ^2	Exp(B)
Step 1	Hardboiled	-152.029	19.551	60.465*	.000
	Constant	4.670	.614	57.778*	106.667
Step 2	Whodunit	241.311	40.864	34.871*	6.309
	Hardboiled	-154.080	25.261	37.205*	.000
	Constant	-.437	.962	.206	.646

Note: * denotes *p*<0.001

We also conducted discriminant analysis to supplement our results. Standardized canonical discriminant coefficient for the predictors were highly similar (Whodunit: -0.657, Hardboiled: 0.697). The results of the regression and discriminant analysis suggest that neither whodunit nor hardboiled is a parent genre to the other.

Differentials

After accounting for both artifacts of the texts (i.e., differentials that only appear in certain texts) and flexi-grams (i.e. differentials where one part is flexible: *my mother/my father*, *a guy/the guy*), the five most frequent differentials for the hardboiled corpus were *my office*, *the/a guy*, *you got*, *was wearing* and *the bar*. The highest ranked five differentials for the whodunit corpus were *upon the*, *the young*, *he has*, *at once* and *no doubt*. As shown in Table 3, some bi-grams obviously belong to longer differentials. For example, *the young* and *young man* are likely to be two bi-grams from *the young man*, similarly, *was wearing* and *wearing a* are two bi-grams from *was wearing a*.

The differentials follow several different patterns. Some (e.g. *my office*, *upon the*) refer to places or more generally to spatiality. Others (e.g. *the/a guy*, *the young man*) refer to people. According to Zwaan, Langston and Graesser (1995), readers build their situation model based on five dimensions, *causality*, *temporality*, *spatiality*, *protagonist goals*, and *intentionality* which are integrated across the text. Not all of these five dimensions can be tracked across our corpora because the corpora comprise only the opening moves of the text. Causality, intentionality, and temporality, in particular, are not available in corpora comprising only beginnings because we are not able to track relationships between developing ideas and events. Any situation model based solely on the exposition is necessarily incomplete. Instead, we focused our analysis on spatiality and identifying characters.

The dimensions of spatiality and characters are already identified as the most indicative (Palmer 2001, Chandler 1985). Although they do not exhaust the avenues for analysis, they represent a reasonable point of departure. Palmer argues that whodunit texts privilege setting, while Chandler argues that hardboiled texts privilege characters.

Spatiality

We predicted there would be greater frequency of *spatial* differentials in the whodunit corpus, given that the literature identifies whodunits as privileging setting (Palmer 2001). The highest ranked whodunit differential, *upon the*, seems to be bear this out.

Table 3: Highest ranked differentials

Whodunit		Hardboiled	
Rank	Differential	Rank	Differentials
1	upon the	1	my office
2	the young	2	the/a guy
3	he has	3	you got
4	at once	4	was wearing
5	no doubt	5	the bar
6	young man	6	wearing a
7	i shall	7	i nodded
8	and with	8	you get
9	my dear	9	he told
10	may be	10	don't know
11	you may	11	look like
12	which had	12	i put
13	had made	13	while he
14	not been	14	started to
15	and of	15	he didn't
16	not be	16	my face
17	very well	17	of coffee
18	would not	18	i didn't
19	man of	19	i found
20	gave a	20	me as

When we examined the differentials using the concordancer, we found that many instances of *upon the* actually refer to *temporality* (see Table 4). However, differentials in the hardboiled index, such as *my office* and *the bar* are purely spatial, referring to both *office* and *bar* as places (see Table 5). Although spaces are treated differently between the two genres, the spatial differentials reveal that setting is as important in hardboiled as in whodunit. This is in contrast to authors such as Palmer (2001) suggest that setting is of greater importance in whodunit texts. Further, there was no

Table 4: Non-spatial uses of *upon the* in the Whodunit corpus

one of the many adventures i have shared with poirot was that of our investigation into the strange series of deaths which followed _upon the_ discovery and opening of the tomb of king men-her-ra.
i had called upon my friend sherlock holmes _upon the_ second morning after christmas, with the intention of wishing him the compliments of the season.
that the man was highly intellectual is of course obvious _upon the_ face of it, and also that he was fairly well-to-do within the last three years, although he has now fallen upon evil days.
this is wanting in the police report, where more stress is laid, perhaps, _upon the_ platitudes of the magistrate than _upon the_ details, which to an observer contain the vital essence of the whole matter.

Table 5: *my office* and *the bar* in the Hardboiled corpus

it was about three blocks from _my office_ building that i saw a cop car double-parked and the two buttons in it
i was just i was just leaving _my office_ downtown when she called, and i picked her up.
mason said cautiously, "i was out of _my office_ when mrs. hastings called to see me."
the kid behind _the bar_ was in his early twenties and looked as if he had never had a drink.
i sidestepped the bellman by _the bar_ and opened the door myself.

significant difference between the amount of spatial differentials across the two corpora ($p = 0.262$).

For example, there is the treatment of specific places in both genres. In hardboiled texts, authors refer to *the neighborhood* in which a character lives in ways related to the character, such as *my neighborhood* or *the old neighborhood*. Very few whodunit texts (3 out of 105) refer to any kind of neighborhood, and when they do it is only *the neighborhood* ($p<0.001$). The equivalent differential in the whodunit texts appears to be *the village*, which is treated as a place of origin (much like *my neighborhood*) or as a destination, though those villages are never named. Hardboiled texts refer to *the village* even more rarely than whodunits refer to *my neighborhood* (2 out of 125). On those rare times they do, *the village* refers to a foreign place, such as Vietnam.

Character identity

Given that characters and their motives are important factors in a hardboiled story (Chandler 1985), we predicted more references to characters in those texts. Overall, the numbers of references to characters were hard to verify, especially once we excluded artifacts of the texts (i.e., names of characters). Whodunits, especially, referred to characters by their full names, or last names and an honorific. On the other hand, hardboiled texts includedoften only the last names of most characters, and so those references are not included in bi-grams or tri-grams.

There is one clear reference to characters at or near the top of each differential list: *the/a guy* and *the young man*. *The/a* guy appears in 33.6% of the hardboiled texts, but in only 4% of the whodunit texts ($p < 0.001$). *The young man* appears in 19.2% of the whodunit texts and 7.2% of the

hardboiled texts ($p = 0.008$). *The/a guy* appears significantly more often in hardboiled texts than *the young man* appears in whodunit texts ($p = 0.014$).

These results support the descriptions given by Chandler (1985) and Rzepka (2005), and Palmer (2001). Because of their focus on presenting useful information for a puzzle, whodunit texts are more precise. Hardboiled texts are meant to reflect reality, or at least a believable world, where people often describe a man as "a guy" or "the guy" rather than with any specific descriptor like "the young guy."

Hardboiled authors appear to present more believable worlds through their treatment of characters as well. More than a quarter (32 out of 125) of the texts feature the differential *was wearing*. This is significantly more often ($p<0.001$) than in the whodunit corpus, where *was wearing* is featured in less than one percent of the texts. *Was wearing* is evenly split between descriptions of the narrator's clothing and the clothing of people the narrator meet.

Discussion

Our results suggest that whodunit and hardboiled may be sibling genres (i.e. neither is hierarchically above the other). Although more research needs to be done, we have found evidence that whodunit and hardboiled are sibling genres, rather than a parent and child. However, the results of the *t*-tests suggest that hardboiled may be overwhelming whodunit in terms of prevalence. More research needs to be done to confirm this conclusion.

Given that we limited ourselves to the expositions, our results may only apply to the opening moves. Further research on different sections of the texts (e.g. the endings) may present a quite different story.

Future studies will also investigate other genres of detective fiction. It is possible that *police procedurals* are related to both whodunit and hardboiled texts. Through understanding further relationships, we could present a three-dimensional picture of multiple sister genres' relationships to one another. The current study presents an important step towards a better understanding of the complex relationships in literature hierarchies.

References

Bhatia, V.K. 1997. Applied genre analysis and ESP. In T. Miller (Ed.), Functional approaches to written text: Classroom applications. Washington, DC: USIA.

Bhatia, V.K. 2002. Applied genre analysis: A multi-perspective model. *Ibérica*, 4: 3-19.

Biber, D. 1988. *Variation across speech and writing.* Cambridge University Press: Cambridge, UK.

Biber, D. 1989. A Typology of English Texts, *Linguistics,* 27: 3-43.

Chandler, R. 1985. The Simple Art of Murder. In H. Haycraft Ed., *The Art of the Mystery Story*. New York, NY: Carrol.

Crossley, S.A., Louwerse, M., and McNamara, D.S. 2009. Identifying linguistic cues that distinguish text types: A comparison of first and second language speakers. *Language Research, 42,* 361-381.

Davies, A. and C. Elder eds. 2004. *The handbook of applied linguistics*. Oxford: Blackwell.

Graesser, A., Olde, B., and Klettke, B. 2002. How does the mind construct and represent stories. In M. C. Green, J. J. Strange, and T.C. Brock Eds., *Narrative impact: Social and cognitive foundations*. Mahwah, NJ: Lawrence Erlbaum.

Henry, A. and Roseberry, R. L. 2001. Using a small corpus to obtain data for teaching a genre. *Small corpus studies and ELT: theory and practice*. Amsterdam: John Benjamins.

Herring, S. and Paolillo, J. C. 2006. Gender and genre variations in weblogs. *Journal of Sociolinguistics*. 10: 439-459.

Holquist, M. 1971. Whodunit and Other Questions: Metaphysical Detective Stories in Post-War Fiction. *New Literary History*. 3: 135-156.

Hopkins, A. and Dudley-Evans, A. 1988. A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes*. 7: 113-122.

Jamieson, K. M. 1975. Antecedent Genre as Rhetorical Constraint. *Quarterly Journal of Speech, 61*, 406-415.

Karlsgren J. and Cutting, D. 1994. Recognizing text genres with simple metrics using discriminant analysis. *International Conference on Computational Linguistics Proceedings of the 15th conference on Computational linguistics.* Kyoto: 1071-1075.

Malmgren, C. D. 1999. The Crime of the Sign: Dashiell Hammett's Detective Fiction. *Twentieth Century Literature*. 45: 371-384.

McCarthy, P. M. 2010. Special Presentation of the Gramulator. Presented at the South-eastern Conference on Linguistics.

McCarthy, P. M. 2010. GPAT paper: A Genre Purity Assessment Tool. In Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference. 247-252. Menlo Park, CA: International Joint Conferences on Artificial Intelligence, Inc.

McCarthy, P.M., Myers, J.C., Briner, S.W., Graesser, A.C., and McNamara, D.S. 2009. Are three words all we need? A psychological and computational study of genre recognition. *Journal for Computational Linguistics and Language Technology.*

McCarthy, P.M., Watanabe S., and Lamkin, T.A. in press. The Gramulator: A tool for the identification of indicative linguistic features. In P.M. McCarthy and C. Boonthum Eds. *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.

McManis, Douglas R. 1978. Places for Mysteries. *Geographical Review*. 68: 319-334.

Meyer zu Eissen, S. and Stein, B. 2004. Genre classification of web pages.KI 2004: Advances in Artificial Intelligence. 3238: 256-269.

Min, H.C. and McCarthy, P. M. 2010. Identifying Varietals in the Discourse of American and Korean Scientists: A Contrastive Corpus Analysis Using the Gramulator. In Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference. 247-252. Menlo Park, CA: International Joint Conferences on Artificial Intelligence, Inc.

Palmer, J. 2001. Tracing Bodies: Gender, Genre, and Forensic Detective Fiction. *South Central Review*. 18,:54-71

Priestman, M. 2003. *Cambridge Companion to Crime Fiction*. Cambridge: Cambridge University Press.

Rahati, A., & Kabanza, F. (2010). Persuasive dialogues in an intelligent tutoring system for medical diagnosis. In Proceedings of the 10th Annual Intelligent Tutoring Systems International Conference. 51-61. Berlin: Springer.

Rehm, G. 2002. Towards Automatic Web Genre Identification. Proc. of the 35th Hawaii International Conference on System Sciences.

Rzepka, C.J. 2005. *Detective Fiction*. Malden, MA: Polity.

Swales, J.M. 1990. *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Thomas R.R. 2005. *Detective Fiction and the Rise of Forensic Science*. Cambridge, UK: Cambridge Univ. Press

Upton, T., and Connor, U. 2001. Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre. *English for Specific Purposes: An International Journal*. 20: 313-329.

Zwaan, R.A., Langston, M.C., and Graesser, A.C. 1995. The construction of situation models in narrative comprehension. *Psychological Science*. 6: 292-297.