School of Languages, Cultures and Societies

Faculty of Arts, Humanities and Cultures

UNIVERSITY OF LEEDS

Standard Cover Sheet for Assessed Coursework

From January 2017 all assessed coursework in LCS must be accompanied by this coversheet (*unless you have a specific learning disability, in which case you should use instead the Marking Awareness coversheet*).

This coversheet should be included as the first page of your submission. Please copy and paste it into the beginning of your assessment (Ctrl+A, then Ctrl+C to copy its entirety). The contents of the coversheet do not count towards your submission word count.

Please complete each of the following sections:-

Student ID:	200521304
Module Code:	5350M
Module Tutor's Name:	Leendert Plug
Assessment description (eg essay 1):	Dissertation
Word count	9631

Please note that by submitting this piece of work you are agreeing to the University's Declaration of Academic Integrity. You can read the Declaration here.

NB Please ensure that you name your file with your SID, the module code, seminar/language tutor's name (if required) and a brief assessment description e.g. 200987654 MODL1234 seminar/language tutor name (if required) Essay 1.pdf

Designing a Framework for Measuring Inference Generation and Communicative Efficacy During Board-game Play

SID: 200521304

Linguistics and Phonetics

School of Languages, Cultures and Societies

University of Leeds

Submitted in part fulfilment of the

requirements for the degree of

MA in Linguistics and English Language Teaching

Module LING5350M

Dissertation (Linguistics and Phonetics)

August 30, 2022

Supervisor: Dr Cecile de Cat

Acknowledgements

With enormous thanks to Dr Cecile de Cat. Without whose serendipitous conversation with Dr Paula Clarke, I would not have been working even remotely close to this area. I could not have produced this work without her boundless enthusiasm, patient guidance, and advice to constantly recontextualise nerves as excitement.

Very special thanks also to Dr Peter Hart and Dr Paula Clarke of the School of Education for their assistance in organising and recruiting for data collection, as well as allowing me to muscle in on their own research project. I hope what follows is of interest and of use!

Loving thanks also go to my partner Charlie for their support, encouragement, and patience during a very stressful period for us both, and for keeping me grounded through numerous bouts of impostor syndrome. And also for their generous loan of a copy of Mysterium to run the study with!

Last but never least, eternal love and thanks to my daughter Maia for her incalculable role in inspiring me to return to education and pursue a Masters degree in Linguistics, many years after my undergraduate degree. I hope we never stop playing games together.

"It's not the notes you play, it's the ones you don't play."
- Miles Davis

Contents

ABSTRACT: Forthcoming research intends to investigate the effect of regular board game use on the reading comprehension abilities of children entering secondary school. This dissertation describes the design of a prototype framework for linguistic analysis which can be developed and applied to such research in the future. An early pilot session involving a tape-recorded board game session was carried out, and the transcript of that session was analysed for inference generation and communicative efficacy. The results of that analysis are presented and discussed from a perspective of future utility and with a view to further development.

1 Introduction

Forthcoming research from the University of Leeds' School of Education intends to investigate the possibility that regular intervention with certain board games can be used to scaffold the reading comprehension skills of children with weaknesses in this area. This dissertation is concerned with the initial stages of designing a framework for quantifying inference generation and communicative efficacy during gameplay, for possible development and use in that same research. An early pilot session of the proposed intervention was run by this author, involving a recorded session of a board game, transcription of that session, and analysis of that transcript. In this paper, the rationale for the development of that framework is laid out, as well as a presentation and discussion of the results of the recorded gameplay session, and the perspective of the present author on the possible facility and utility of this approach for future research efforts.

Beyond the simple act of decoding words and phrases into meaningful information, a significant aspect of the comprehension process is being able to build a mental representation of what the discourse is about and the modification and manipulation of such mental models (Glenberg et al., 1987). Being able to construct such representations relies in part upon the ability to infer meaning and information. We may define inference as "the process by which information is retrieved or generated to fill in information that is left implicit" (Kendeou, 2015, p161). In other words, it is the mechanism by which the human brain reads between the lines. Inference is typically described as a two-stage process in which current information first ACTIVATES previously acquired information into working memory (either from earlier in the discourse or from long-term memory), and then that activated information is INTEGRATED into the current content (Cook & O'Brien, 2014). Activation and integration are usually conceived as parallel, overlapping, and continuous, and not necessarily as two completely distinct cognitive processes (Kintsch, 1988). Both of these operations are assumed to be passive and automatic, but of the two, integration has the potential to be influenced by outside, strategic or attentional processes (Kendeou, 2015). The successful application of these two processes is essential to achieving discourse comprehension. Further, as a facet of reading comprehension, inferential ability has been identified as one of the primary predictors of proficiency for both adults and children (Kendeou, Bohn-Gettler, et al., 2008; Kendeou, 2015).

Because of the consistency of the initial activation process across all definitions and descriptions of inference, and the main distinction being in the source of information which drives that activation, it has been argued that a categorisation system for inferences is of limited value (Gerrig & O'Brien, 2005). Nevertheless, a broad array of inferential taxonomies and distinctions have been drawn over the years, based on numerous differentiating factors. For example, the distinction between BACKWARD INFERENCES and FORWARD ELABORATIONS. The former type links information in the focal statement (ie. the one currently being read or heard) to information which is kept in memory. The latter anticipates the future relevance of information, highlighting and keeping it prominent in the mind of the perceiver (Fincher-Kiefer, 1993; Kendeou, 2015; van den Broek, Fletcher, et al., 1993). A more controversial distinction is made between AUTOMATIC and CONTROLLED inferences, or those which are generated without conscious effort by the comprehender, and those which require consideration and thought to draw (Graesser, Singer, et al., 1994; Kendeou, 2015). A final example of a taxonomic distinction is made between LOCAL and GLOBAL inferences: those which are required for ensuring that discourse remains comprehensible at the linguistic, sentence-to-sentence level, and those which are required for comprehension at a higher, contextual level (Graesser, Millis, et al., 1997; Kintsch, 1988). In spite of Gerrig & O'Brien's (2005) protestations then, a number of different taxonomies have arisen, and will no doubt continue to arise. Kendeou (2015) suggests that the inference taxonomies are important only in the context of the discourse characteristics which may encourage or discourage the generation of one variety over another (p165).

A brief note on terminology. Significant portions of the current research into inference generation and comprehension are performed on written text, mainly for practical and methodological reasons (Pickering & Garrod, 2004). The words TEXT and DISCOURSE are used throughout this paper, and are used respectively to mean *written language* and *spoken language*, unless otherwise stated.

2 Literature Review

This section will introduce the concepts used to formulate the present study. We will first look briefly at discourse comprehension as a whole, discussing the concepts of mental representations and alignment of the same between interlocutors. Then we turn to Relevance Theory and what it tells us about how individuals select language to communicate and refer to entities in their environment. Finally we will discuss inference, how it relates to discourse comprehension, and provide some examples of specific types of inference which will form the beginnings of our taxonomy for categorisation of the data gathered for this study.

2.1 Discourse Comprehension

In order to comprehend anything, written or spoken, an individual must construct a mental representation of the discourse, decoding and converting linguistic information into a connected model of what the discourse was about. This representation can then be manipulated by its owner, who may demonstrate comprehension by—for example—paraphrasing the discourse, or answering questions about its content (Glenberg et al., 1987; Kintsch, 1988). A variety of levels of discourse representation have been identified across the literature, though the most commonly accepted distinctions are made between the SURFACE CODE, a preservation of the exact syntax of a message, typically only retained for a short period of time unless a specific wording has significant import as to the meaning of the message (for example during a joke); the TEXTBASE, containing the decoded propositions transmitted by a message, but absent of the specific syntactic structure which delivered them; and the SITUATION MODEL, the overall content that a full set of messages (for instance a story or a conversation) is about, which is constructed through interactions between the text itself and the comprehender's background knowledge of the world (Graesser, Millis, et al., 1997; Kintsch, 1988; van Dijk & Kintsch, 1983). All three of these levels of representation work in tandem to render any discourse as coherent in the mind of the experiencer: a discourse must be rendered coherent at both the local, surface code level, and at the global, situational level.

Much of the research into comprehension has been performed using written monologues (eg. Glenberg et al., 1987; Kendeou & van den Broek, 2007a; Trabasso & Suh, 1993), due in part to the experimental complexity in controlling the language use involved in naturalistic discourse, but also because of a theoretical derivation from the generative school of linguistics. Much as generative linguistics makes use of sentences which have been isolated and decontextualised, so too does much of the psycholinguistic study into comprehension (Pickering & Garrod, 2004). However, the overwhelming majority of human communication is performed in context, and through dialogue. Clark (1996) conceptualises discourse as a joint activity, where two or more participants work together (rather than in isolation) to accomplish a particular task, and in doing so relies on all participants assuming that every other participant has certain information available to them. Clark refers to this information as COMMON GROUND, and defines two broad categories: COMMUNAL COMMON GROUND, which is linked to membership in a cultural community such as nationality, ethnicity, hobby, politics, or subculture; and PERSONAL COMMON GROUND, which is derived from one's immediate perceptual environment.

It is critical to note that common ground is neither fixed nor static, but constantly evolves over the duration of a discourse. Suppose that I meet somebody at a party, and after an hour of conversation, discover that they too are an enthusiast for the children's TV show The Magic Roundabout. Our communal common ground has now grown to incorporate a shared assumption about each other's background knowledge relating to The Magic Roundabout, and our ability to refer to any of this information has been facilitated by this assumption. In doing so, we have both updated our situation models about the nature of the discourse, and brought them into alignment (Zwaan & Radvansky, 1998).

This same process of alignment has been shown to happen for personal common ground as well. Garrod & Anderson (1987) designed a co-operative game in which pairs of players had to guide each other through a maze containing blockages and gates. Each dyad had to co-ordinate to activate switches which were only visible on their own screen to unblock passages for the other participant. Garrod & Anderson found that pairs of participants very quickly co-developed a framework for communicating their spatial locations to each other, for instance relying upon a traditional *x,y* coordinate system (eg. "C4"); or else describing broader locations such as "the L on its side" (p190). In their discussion the researchers note that each participant pair arrived at its own particular spatial model, and that for every pair, both participants arrived at the same model. This result is reminiscent of the results of Krauss & Weinheimer's (1964; 1966) classic finding that referent terms for abstract drawings were shortened over the course of successive task repetitions, dropping from an average length of ten words to two. In addition, Krauss and Weinheimer found that this shortening was only possible in a scenario when feedback from the listener (even something as simple as "uh-huh") was available to the speaker, and that when speaking into a tape-recorder, no such shortening occurred. Thus we can see that the creation and development of meaning is a joint process.

Alignment of mental models is not an immediate operation, and nor is it permanent. After a trio of experiments designed to assess the roles of recency, frequency, accessibility, and informativeness in selecting referring expressions, Brennan & Clark (1996) conclude that speakers agree upon CONCEPTUAL PACTS, which are "temporary agreements about how they and their addressees are to conceptualize that object" (p1491). Brennan & Clark found that speakers would alter these agreements either to adapt to changing task demands, new addressees, or to improve the communicative efficiency of the agreed upon referent. How speakers collaborate to decide upon what referent to use to talk about a given concept is a matter for Relevance Theory, which we will discuss in the next section.

2.2 Relevance Theory

Relevance Theory states that the act of communication is aimed toward the maximisation of a cognitive effect from a stimulus (ie. linguistic input), wherein a larger cognitive effect entails greater relevance; while minimising the amount of processing effort required for an interlocutor to understand (Clark, 2013; Sperber & Noveck, 2005; Sperber & Wilson, 1986). Sperber and Wilson (1986) define a cognitive effect as a change to an individual's mental representation of the world (p265). For example, hearing the utterance *Florence's plane will arrive at four fifteen* means that the listener's representation of Florence's time of arrival has now been impacted and changed. The larger the alteration to the mental representation, the greater the cognitive effect, and thus the greater the relevance of the utterance. Further, processing effort and relevance are inversely correlated:

(1) Florence's plane will arrive at four fifteen.

(2) Florence's plane will arrive after four.

(3) Florence's plane will arrive forty-five minutes before five.

At the surface, all of these statements might be considered equally relevant to the listener. (1) is in fact the most relevant of these three. (2) is entailed by (1), meaning that (1) produces all (and more) of the possible conclusions derivable from (2), and so the cognitive effect of (1) is the greater. Meanwhile (3) is identical to (1) in terms of the information which can be derived, but the additional effort required to calculate the meaning of (3) lowers its relevance significantly (Van der Henst & Sperber, 2005). This then is the COGNITIVE PRINCIPLE of Relevance Theory: "human cognition tends to be geared towards the maximisation of relevance" (ibid., p143).

The second principle of Relevance Theory is COMMUNICATIVE: "every utterance conveys a presumption of its own optimal relevance" (ibid., p.144). Optimal relevance is a function of the processing effort by the listener as well as the communicative preferences and abilities of the speaker. The listener is entitled to the assumption that any utterance they hear is sufficiently relevant to be worth processing. Meanwhile, the speaker wishes to be understood and has a vested interest in making her utterance simple to comprehend, and to provide evidence for the cognitive effect she wishes to achieve (Sperber & Wilson, 1986; Van der Henst & Sperber, 2005). The presumption of optimal relevance then, leads to a general inferential heuristic for comprehending any utterance:

1. Follow the path of least effort in constructing and testing hypotheses about intended meaning.

2. Stop when your expectations of relevance are satisfied (Sperber & Noveck, 2005, p7).

This is the Relevance-Guided Comprehension Heuristic. Let us consider an example:

(4) Dougal has a big cat.

Upon hearing (4), one might derive two possible meanings: that Dougal has a pet domestic cat which is large, or that Dougal has a pet lion, tiger, panther, or similar. In order to resolve this ambiguity, the Relevance-Guided Comprehension Heuristic states that a listener would not need to derive both possible meanings and select the more likely of the two. Instead, following the path of least effort will lead most speakers to infer the former, least effortful meaning first, and if it is deemed most relevant to the context (for instance, in a situation where none of the interlocutors are involved in zookeeping), that meaning will be selected. However this does not mean that the burden for comprehension is placed solely on the listener. Sperber & Wilson (1986) analogise communication to ballroom dancing: one partner leads the dance in order to avoid problems in coordination. An individual communicating a message such as (4) must make correct assumptions about the information and context that the audience already has available to them, and that they are likely to use in the comprehension process. Arnold (2008) corroborates this with her finding that when choosing referring expressions, "speakers consider who they are talking to, making global assumptions about who the addressee is individually, what social groups they belong to, and what they are likely to know as a result" (p520).

We can see that the acts of communication and comprehension involve not just the alignment of mental representations, but certain unspoken presuppositions about the mental states of our conversation partners. We suppose that what a speaker says is sufficiently relevant enough to be worth processing, but also that listeners will be able to comprehend the meaning of our words, which we select in accordance to our assumptions about the cognitive states of others. This is naturally not a one-time operation: feedback is a critical aspect of communication and jointly creating meaning (Brennan and Clark, 1996). Indeed, feedback need not even be strictly lexical: hedging, hesitancy, and disfluency in speech are all markers that what follows may require increased processing capability, or will be something unfamiliar to the listener (Arnold et al., 2007). In order to provide feedback as to what we interpret from a heard utterance, including to interpret non-lexical signals such is disfluency, we need to be able to infer meaning.

2.3 Inference

The cognitive view of reading comprehension outlined by Kendeou et al. (2014) divides the processes involved into two categories: lower- and higher-level. Lower-level processes are those which are involved in decoding the written word into meaningful language units: word identification, fluency, and vocabulary knowledge. Meanwhile higher-level processes combine those units with each other and—where necessary—additional knowledge to create a mental representation of the text, and include inference generation, comprehension monitoring, and working memory (Kendeou, 2015). To put it another way, the lower-level processes are concerned with comprehension at the surface code and textbase levels, while the higher-level processes are concerned comprehension at the textbase and situation model levels (see section 2.1 above). Viewed from this perspective, we can already see that the two levels of cognitive processing are interlinked, and must have an influence upon each other. Indeed, these processes typically occur in parallel, but are usually researched in isolation (Kendeou, van den Broek, Helder, et al., 2014; van den Broek, Fletcher, et al., 1993). A number of factors can influence these processes and—it follows—an individual's comprehension ability.

Prior knowledge and awareness of genre can both have an impact upon the levels of comprehension shown by a reader (Kendeou & van den Broek, 2007b). Meanwhile, reading a text which does not align with one's pre-existing conceptions (or misconceptions) can lead to lower levels of comprehension (Kendeou & van den Broek, 2005). Individuals have their own criteria for comprehension, and inferential abilities are impacted by these standards. If an individual perceives that—for example—a pronoun does not have a clear referent, or a proposition does not have a readily apparent causal antecedent, then controlled, strategic processes will be activated to attempt to remedy the lack of coherence (P. van den Broek et al., 2011). In addition to individual standards of coherence, the demands of a given task will also affect which kinds of inference are generated, according to an individual's own personal goals in comprehending a discourse (van den Broek, Lorch Jr., et al., 2001). Carlson et al. (2014) state that coherence can be divided into two camps: local coherence, where current information is connected with very recent information from the discourse; and global coherence, where current information is linked with earlier discourse information beyond that which immediately preceded it, as well as associating that same information with broader world knowledge (see also Graesser, Singer, et al., 1994).

Inference is the skill which is central to building and maintaining these standards of coherence. At this stage in our journey, then, we may define two broad categories of inference, borrowing the local/global distinction to name them. Local inferences operate at the surface code and textbase levels, connecting current information with information active in the working memory, typically from up to the three immediately previous sentences (ibid.). Global inferences incorporate relevant information both from within the discourse and from long-term memory, such as as general world knowledge or thematic understanding (Kendeou, 2015; van den Broek, Fletcher, et al., 1993). What follows is a discussion of these two categories of inference and a selection of illustrative examples. However it must be acknowledged again that this is but one taxonomy of inferences among many (Kendeou, 2015), and furthermore, even within such a binary distinction as local/global, inferences occur on a spectrum between the two poles. Additionally, activation of such inferences does not constitute an on/off switch, but are better thought of as existing on a continuum of degrees of activation (Gundel et al., 1993; Kintsch, 1988). With that equivocation in place, let us now discuss a binary system of inference taxonomy for use later on.

2.3.1 Local Inferences

Local inferences are required for maintaining coherence and comprehension at the surface code and textbase levels of representation. Rickheit et al (1985) refer to this as "lexical decomposition" (p18), where text is broken down into semantic units and the relationships between them inferred. These inferences are regarded as fully automatic, necessary for establishing coreferents during comprehension, and happen without the comprehender being aware (Kintsch, 1988; Paul van den Broek, Kendeou, et al., 2005). According to Swinney & Osterhout, such inferences are "mandatory, immediate, and uninfluenced by world knowledge" (1990, p21). Inferences which are included under this category are the resolution of ellipsis and anaphora, as well as the comprehension of semantic roles, and of referent chains.

2.3.1.1 Anaphor Resolution

We can define an anaphor as a noun phrase (NP) which takes its meaning from another NP in the same utterance (Carnie, 2013).

(5) Zebedee trimmed his moustache.

A local inference is required to resolve the meaning of the underspecific pronoun *his*. Due to Zebedee's status in the discourse (ie. as the subject of the utterance), Zebedee is activated in the listener's working memory and upon encountering the pronoun later in the utterance, the listener can integrate the given information (Zebedee) with the new (he) to resolve the ambiguity of the referent (Garnham & Oakhill, 1992).

2.3.1.2 Ellipsis

Ellipsis is the deletion of an already-uttered phrase in subsequent structures where it would otherwise be repeated (Carnie, 2013).

(6) Ermintrude's favourite food is cheese. Dylan's is mushrooms.

The intuitive understanding of (6) is easy enough to see: we are dealing in favourite foods. However, the second clause omits the phrase *favourite food*, and yet the meaning is still preserved. Another local inference must be made, this time across two clauses. The information about Ermintrude's favourite food is still in working memory, and the omission in the second clause causes the comprehender to automatically seek a suitable antecedent to fill in the blank (Fox, 2000).

2.3.1.3 Semantic Roles

The semantic (or theta) role is the name given for the particular role that the syntactic arguments of an utterance play in that sentence, for example *agent*, *theme* or *patient* (Aarts, 2001).

(7) While Mr McHenry was cooking cabbages fell from the sky.

Garden path sentences such as (7) first activate one meaning before a reevaluation takes place to correctly interpret the sentence. Christianson et al (2001) assert that the initial interpretation of garden path sentences remains activated in short term memory even after the whole sentence is understood. This implies that upon hearing each word, the listener's inferential mechanisms are drawing upon semantic knowledge of that word and predicting how and when the necessary arguments will arise. Hence, the initial difficulty in processing a garden path sentence is the result of an incomplete integration of activated predictions.

2.3.1.4 Referent Chains

The term REFERENT CHAINS was coined by Chastain (1975) as a way to describe the phenomenon of referential expressions referring to the same entity shifting as discourse progresses, often shaped by multiple speakers.

(8) A: Buxton the blue cat shall be crowned king.
B: He'll be the greatest one that ever lived!

(9) A: The upside-down martini glass in a wire stand.
B: The inverted martini glass.
A: The martini glass.
B: The martini.

Example (8) shows how anaphoric terms can be used across turns to refer to entities which are already activated and in focus in the discourse. B substitutes *Buxton the blue cat* to *he*, and *king* to *one*. Gundel et al.'s (1993) Givenness Hierarchy proposes that the form that the referent term takes indicates its prominence in the mental representations of the speakers. Thus in (8) *he* indicates that the referent is in focus, but in a similar way to anaphoric inference, the pronoun takes its meaning from an utterance earlier in the discourse, and relies on participants being able to infer what is being referred to.

Although not an example of naturalistic discourse, (9) shows how referring expressions became shorter over time in Krauss & Weinheimer's (1964) study into changing referent phrases as discussed previously in section 2.1. The entities remained activated and relevant in the participants' mental representations and thus shorter and shorter referring terms could be used, relying on the fact that interlocutors could infer which image was being referred to by the attenuated expression.

2.3.2 Global Inferences

(10) Mr Rusty wrote a letter to his MP. He wasn't very happy with her response.

In addition to the need for local inferences to comprehend the use of the anpahoric pronouns present in (10), additional information not available in the surface code must be applied to truly comprehend the meaning of the whole utterance. The phrase *his MP* does not have the literal meaning of *the or an MP owned by Mr Rusty*. Instead, a global inference using background world knowledge must be generated to create a meaning something like *the elected member of parliament for the constituency in which Mr Rusty is resident* in the mind of the listener. Additionally, the second clause in (10) requires a further inference about the nature of her response. The listener must activate information about the meaning of MP (ie. that an MP refers to a human individual to which one may send a letter and expect a reply from), and integrate it with the pronoun *her* to achieve comprehension of the utterance (Arnold, 2010). Thus, we can see that global inferences are necessary for comprehension at the textbase and situation model levels of representation. While local inferences are viewed as automatic and mandatory, global inferences can also be strategically applied by the reader or listener to integrate wider knowledge with in-focus information (Kendeou, 2015; van den Broek, Fletcher, et al., 1993). According to Graesser & Kreuz's (1993) constructionist view, such inferences are not generated *during* text comprehension, but at some point after, dependant upon the goals of the reader. It is possible for global inferences to be generated automatically as well, however: when comprehending discourse, the experiencer constructs a connected mental model of each proposition in the text and infers—for example—causal links between each node. Nodes with a larger number of connections contribute more to the situation model of a discourse than do those with fewer connections (Kendeou, van den Broek, White, et al., 2007; Paul van den Broek, Kendeou, et al., 2005). In addition to causality, other types of global inference include thematic inferences; character goals, beliefs, and emotions; spatial relations; and authorial intent.

2.3.2.1 Thematic

Thematic inferences are ones which integrate together large portions of a discourse which convey a broader meaning or point of the text (Graesser, Singer, et al., 1994). Such inferences can be generated automatically to the extent that the discourse is adequately coherent (Gundel et al., 1993), but also require more cognitive processing than some other varieties of inference (McGinnis et al., 2007).

2.3.2.2 Character

In fictional narratives, character inferences integrate explicit textual information with more general interpersonal and sociocultural knowledge to generate assumptions about a character's goals, emotional state, thoughts, motivations, personality traits, and so on. According to McGinnis et al (ibid.), such inferences are essential for situation model development, and in more experienced or confident readers such inferences are made automatically (Paul van den Broek, Kendeou, et al., 2005).

2.3.2.3 Authorial Intent

In a taxonomy of inferences generated by a passsage of text, Graesser et al (1994) identify the author's intended overall meaning and attitude towards the subject matter as one of the least studied categories of global inference (p376). There is continued debate as to whether or not these inferences are generated on-line (ie. during and as a part of the main comprehension process) or not (Graesser & Britton, 1995). According to Gibbs (1995) both literary and everyday language is governed by the use of metaphor, and the specific choice of metaphor used by an author or speaker will shape the kinds of inference drawn by the comprehender, as well as the wider information which is activated.

2.3.2.4 Spatial inference

Typically arising from the activation of information which is coexistent or implied by information contained within the focal statement, spatial inferences are generated to build the situation model during discourse (van den Broek, Fletcher, et al., 1993). However as shown by Morrow et al (1990), spatial inferences are not required to construct the textbase level of representation. For example, the choice of prepositions and motion verbs in the statement *John walked past the kitchen to the bedroom* has an impact upon the reader's conceptualisation of where John begins his motion, but no effect on their propositional representation of the statement.

2.3.3 Summary

The above is by no means an exhaustive list of possible inferences. As has been stated, the types of inference which can be and are generated automatically are largely dependent on the contents of the discourse, and it is evident from example (10) that global inferences are not exclusively strategic and effortful, but can (and indeed must) also occur automatically. The role of inference generation ability in reading comprehension is also clear, and as stated a number of factors can impact an individual's ability to generate inferences. Next we turn to the design of a board game-based intervention which will attempt to quantify some of the inferences generated during gameplay.

3 The Present Study

As stated in the introduction, the purpose of the present study is to propose a possible framework which can be developed for use in subsequent research projects into board games and their use in scaffolding reading comprehension skills in school-age children. As such this study may be conceived of as a pilot programme, operated to identify weaknesses and strengths in the approach described, as a starting point for further development, rather than to necessarily generate comparative data among the participants involved.

Figure 1: An example of one of the ghost player's vision cards

Participants were recruited to play one round of the co-operative board game Mysterium (Asmodee, 2015). During the game of Mysterium, up to seven players work together to try to solve a murder. One player acts as the ghost of the murder victim, and must communicate with the remaining players solely through the use of abstract vision cards to help them to identify the murderer, location of the crime, and the weapon used. The non-ghost players must work collaboratively in their deductions about the visions they each receive and how they might link to their own, predetermined correct answer from a number of different possible alternatives. Once all players have collected their own set of one culprit, location and weapon, a final round takes place wherein the ghost points all players to one final correct answer from the available possibilities. During the game the ghost is not permitted to speak to the players, lest they give away clues as to the correct answers. Apart from the rulebook, the game contains no written elements, and is based entirely on visual stimuli, requiring players to describe to each other what they see in order to forge associations between vision cards, and suspect/location/weapon cards. It is for this reason that the game was identified as a possible intervention tool for research into scaffolding reading comprehension.

3.1 Participants

Six adults (n=6; aged 28-44; M=36.8) participated in the study on a voluntary basis, with the present researcher taking on the role of ghost and game master. All participants were recruited with the assistance of colleagues from the University of Leeds' School of Education, and were either staff members of the School, or were postgraduate students at the School and known to at least one of the staff members present. Of the six, two participants were male and four were female. Half of the participants had English as a first language, while the remaining half did not. However, given the University's requirements for English language proficiency at a postgraduate level, this was not judged to be a concern. This sample naturally constitutes a convenience sample. However given the eventual aim to use this framework with children rather than adults, the data gathered should therefore represent idealistic results.

Informed consent was elicited from all participants and no subsequent requests for removal from the study were received. An example of the information sheet and consent form can be found in Appendix B.

3.2 Procedure

One copy of the board game Mysterium (Asmodee, 2015) was used. The game was mostly unmodified, however the included egg-timer mechanic was not used in order to maximise the amount of player interaction and possible data. In addition, participants were instructed to think aloud to the best of their ability, to attempt to capture any inferences generated about the visual stimuli they were presented with. The gameplay session was recorded using a voice-recorder placed in the middle of the table used to play the game. This recording was then transcribed by the present author using the CLAN software package, before porting to a CSV file to be coded and manipulated. Each complete utterance was recorded on its own row, and evaluated according to the below coding schema. Any utterances by the researcher were removed from the dataset.

During this session, a full game of Mysterium was not completed. At the end of the first portion of the game (in which players generate their own possible solution to the murder mystery), it was agreed by all participants that this would be an appropriate stopping point due to time constraints and prior commitments. Regardless, according to the rules of play, the final part of the game requires that the players no longer talk with each other, and thus would not have generated any further significant data.

3.3 Coding Schema

To preserve anonymity, during coding each participant was assigned a three-letter code according to the colour of their player token from the recording session: black (BLA), blue (BLU), purple (PRP), white (WHT), and yellow (YLW).

3.3.1 Inference

Given the well-attested import of the ability to generate inferences as a facet of reading comprehension (eg. Kendeou, 2015), each utterance was evaluated for the presence of a number of different inference types. The inferences which were coded are described below. Each sub-type of inference was further categorised with a mid-level category, creating a 3-tier system for each inference coded, for example *local* > *grammatical* > *anaphor*. The reason for this is to attempt to capture the types of inference which participants make at a number of levels of specificity, and to allow for the expansion of this schema to incorporate more inference categories in the future. This schema is summarised in a table which can be found in Appendix A.

3.3.2 Local Inferences

Local inferences are necessary for maintaining coherence at the level of the surface code and textbase representations and occur commonly in everyday discourse. These are typically automatic and require less cognitive effort to generate (eg. Swinney & Osterhout, 1990). It was decided that three types of local inference should be coded: anaphora, ellipsis, and referent chains.

Anaphora were coded when a noun phrase (NP) in the utterance took its meaning from another NP in the same utterance (Carnie, 2013). Typically these were identified by the presence of a pronoun in the utterance. However, if a NP took its meaning from a NP in a *different* utterance, then it was not coded as an anaphor. In these instances a referent chain was coded. Furthermore, across utterances a referent chain was coded for sequential referential expressions referring to the same entity (Chastain, 1975). This includes the use of deictic and anaphoric expressions such as *it*, *that*, or *this*. However, where our taxonomy requires an anaphor to occur within the same utterance as the NP which provides its meanings, anaphoric expressions appearing across utterances referring to the same entity are categorised as referent chains. This decision was taken to attempt to distinguish between grammatical inferences required for coherence at the surface-code level, and more cohesive inferences required for making communication more efficient. In addition, where explicit NP referents had been shortened (eg. *hot air balloons* becoming *balloons*), these too were categorised as referent chains. Finally, Ellipsis was coded for utterances where an already uttered NP or verb phrase (VP) were omitted in subsequent structures where that phrase would have been repeated (Carnie, 2013). Ellipsis is another inference required for grammatical coherence, and so together with anaphora, these two were given a category of grammatical inference, while referent chain inferences were given a category of cohesive inference.

3.3.3 Global Inferences

Global inferences are required for making more effortful deductions about a discourse, whether that is relating to the themes of an overall text, or connecting causal relations between two propositions (eg. Paul van den Broek, Kendeou, et al., 2005). For this study, an adapted version of the taxonomy used by McGinnis et al. (2007) and Graesser et al. (1994) was used. The taxonomy used by Graesser et al. (ibid.) contains 11 possible global inferences, and McGinnis et al. (2007) expand this further to incorporate 17 different categories of global inference. However, as both of these studies were performed using narrative text rather than more naturalistic discourse, the decision was taken to omit or combine some of these categories. For example, while Mysterium has an overall theme of murder mystery, there is very little causal relation between a vision card and its suspect/location/weapon partner, so causal inferences were not included in this taxonomy. The following categories were selected for coding global inferences: character, place, object state, and thematic.

CHARACTER inferences are derived and combined from the schema of McGinnis et al. (2007), and incorporate any reference to a portrayed person's physical characteristics, social/work roles, emotional state, or current activities. Given that the first stage of the game requires players to identify the suspect which has been assigned to them, character inferences are expected to be a common occurrence. Similarly, the second stage requires players to identify a location which has been assigned to them, so PLACE inferences were coded when participants produced the name of a setting or a place in reference to their visual stimulus. It is worth repeating that the game contains little to no written elements, and so players must make their own deductions about what they see. One player's basement might be another's attic, thus place inferences are a relevant metric to measure. Mentions of the state or properties of an object were coded as OBJECT STATE inferences. Finally, due to the turn structure of the game some participants will likely have multiple vision cards pointing towards a single correct answer, and it is anticipated that there will be deductions attempted as to what binds all of the visions together in pointing towards a single result. Expressions of this kind were coded as THEMATIC inferences.

Figure 2: An example of one of the player suspect cards used in the data collection session

Figure 3: An example of one of the player location cards

During coding, it was found that participants made frequent associations between the objects depicted on their cards and the purpose or apparent use for those objects, so an additional category of INSTRUMENT inferences was added to account for this. This name is derived from the coding schema used by Graesser et al. (1994), and is not strictly related to the idea of the semantic role of *instrument* (eg. Haegeman, 2006). Instrument inferences were coded when a player referred to an object that would be used to accomplish an explicitly stated action. Furthermore, it was found that the object states category as initially conceived was too broad, and incorporated simple descriptions of objects. It was amended to discount simple descriptions of colour, shape, or size, and only those utterances which referenced a *reason* for the state of an object were coded.

Character, place, object state, and instrument inferences are described by Graesser et al. (1994) to be ELABORATIVE inferences, and as with the local inferences, a category by the same name has been added to incorporate these four. Thematic inferences meanwhile come under the category of EVALUATIVE inferences.

3.3.4 Communicative Efficacy

In addition to inferences generated by participants, each utterance was evaluated for communicative efficacy. Because the game requires communication about the nature of the visions and their counterparts on the table, the majority of which are highly abstract, conceptual pact (CP) development is an appropriate aspect to this schema, as players try to figure out how to refer to elements of cards in play. Each utterance was considered for the presence of a noun phrase (NP) CP which was used to describe a feature of one of the cards currently in play. Each possible CP was then evaluated for whether it was being proposed or being requested. Hedging, disfluencies, explicit questions, and rising intonation as in questions were all regarded as markers for requesting or confirming a CP from others. A CP proposal was marked by the explicit naming of an object or feature, and a lack of hedging or disfluency. Any recorded CP was then evaluated for alteration, rejection, acceptance, or use.

ALTERATION was coded when a referent term was changed by any participant after its first appearance in the discourse, indicating relevance of the entity being referred to, but unacceptability of the referent term. Where a CP was altered, the participant who supplied the alteration was recorded, as well as a judgement as to the reason the initial CP was not sufficiently acceptable, such as under- or over-specificity, or a lack of lexical availability for the speaker. For example, one participant produced the utterance *maybe it's to do with fire; like that could breath fire, and the last picture ...*, for which the fire-breathing referent was subsequently altered to *dragon* by a different player later in the discourse. ACCEPTANCE was coded for explicit stated agreement such as *oh yeah*, or *you're right* to a proposed or requested CP by a participant other than the proposer, indicating the acceptability of a CP. USE was coded when a participant other than the CP proposer used the same term to refer to the same entity. When a CP was coded as used, which participants also produced that CP was also recorded.

REJECTION was initially intended to be coded when a proposed or requested CP went unused by another participant subsequently in the discourse, indicating unacceptability of the term. However, it was frequently ambiguous if a CP had been rejected due to its unacceptability or simply not used because the discourse focus had moved on before there was time for players to evaluate it. Therefore, rejected CPs were not coded and instead were given the label UNUSED.

4 Results and Discussion

4.1 Results

After investigator turns were removed from the transcript, a total of 557 (M=96.17, s=39.19) utterances remained. Of that 557, 410 (M=68.33, s=27.65) utterances contained one or both of inference and conceptual pact. The average utterance

Player	Turns	Total Inferences (per turn)	Local Inferences (per turn)	Global inferences (per turn)
BLA	116	142 (1.22)	55 (0.47)	87 (0.75)
BLU	98	129 (1.32)	43 (0.44)	86 (0.88)
PRP	56	80 (1.43)	19 (0.34)	61 (1.09)
RED	161	184 (1.14)	62 (0.39)	122 (0.76)
WHT	87	116 (1.33)	41 (0.47)	75 (0.86)
YLW	59	50 (0.85)	6 (0.10)	44 (0.75)
Mean	96.17	116.83 (1.22)	37.67 (0.37)	79.17 (0.85)
SD	39.19	47.15 (0.21)	21.37 (0.14)	26.56 (0.13)

Table 1: Global, local, and total inferences generated

Player	Turns	Local Inferences (per turn)	Anaphor (per turn)	Ellipsis (per turn)	Referent Chain (per turn)	Grammatical inferences (per turn)	Cohesive Inferences (per turn)
BLA	116	55 (0.47)	21 (0.18)	3 (0.03)	31 (0.27)	24 (0.21)	31 (0.27)
BLU	98	43 (0.44)	20 (0.20)	9 (0.09)	14 (0.14)	29 (0.30)	14 (0.14)
PRP	56	19 (0.34)	6 (0.11)	3 (0.05)	10 (0.18)	9 (0.16)	10 (0.18)
RED	161	62 (0.39)	28 (0.17)	6 (0.04)	28 (0.17)	34 (0.21)	28 (0.17)
WHT	87	41 (0.47)	17 (0.20)	4 (0.05)	20 (0.23)	21 (0.24)	20 (0.23)
YLW	59	6 (0.10)	3 (0.05)	1 (0.02)	2 (0.03)	4 (0.07)	2 (0.03)
Mean	96.17	37.67 (0.37)	15.83 (0.15)	4.33 (0.05)	17.50 (0.17)	20.17 (0.20)	17.50 (0.17)
SD	39.19	21.37 (0.14)	9.54 (0.06)	2.80 (0.03)	11.02 (0.08)	11.58 (0.08)	11.02 (0.08)

Table 2: Local inferences generated

length was 11.74 (s=6.8) words. In order to make the data more easily comparable, a per-turn figure was generated where appropriate for each measurement by taking the total number of (for example) thematic inferences produced by one participant, and dividing that by the total number of turns taken by that same participant. A total of 675 (M=112.5, s=47.01) inferences were recorded, at an average rate of 1.17 inferences per participant per turn (s=0.21). Of the 675 inferences, 226 (M=37.67, s=21.37; per turn M=0.37, s=0.14) were local and 475 (M=79.17, s=26.56; per turn M=0.85, s=0.13) were global. These results are summarised in Table 1.

Of the local inferences, 121 were grammatical (M=20.17, s=11.58) while 105 were cohesive (M=17.5, s=11.02). Anaphora were the more common of the grammatical inferences generated with a total of 95 (M=15.83, s=0.15), while 26 ellipses were recorded (M=4.33, s=2.8). Meanwhile for the global inference category, 318 were elaborative (M=53, s=22.18), and 157 were evaluative (M=26.17, s=6.37). Character inferences were recorded 48 times (M=11.33, s=0.12), place 68 (M=11.33, s=0.12), object state 104 (M=17.33, s=0.17), and instrument 98 (M=26.17, s=6.37). Tables 2 and 3 show a breakdown of these results per participant for local inferences and global inferences respectively.

A total of 156 conceptual pacts were recorded across the duration of the tran-

Player	Turns	Global Inferences (Per Turn)	Character (Per Turn)	Place (Per Turn)	Object State (Per Turn)	Instrument (Per Turn)	Thematic (Per Turn)	Elaborative Inferences (Per Turn)	Evaluative Inferences (Per Turn)
BLA	116	87 (0.75)	6 (0.05)	14 (0.12)	20 (0.17)	25 (0.22)	22 (0.19)	65 (0.56)	22 (0.19)
BLU	98	86 (0.88)	6 (0.06)	10 (0.10)	19 (0.19)	18 (0.18)	33 (0.34)	53 (0.54)	33 (0.34)
PRP	56	61 (1.09)	13 (0.23)	5 (0.09)	7 (0.13)	9 (0.16)	27 (0.48)	34 (0.61)	27 (0.48)
RED	161	122 (0.76)	11 (0.07)	15 (0.09)	34 (0.21)	28 (0.17)	34 (0.21)	88 (0.55)	34 (0.21)
WHT	87	75 (0.86)	8 (0.09)	12 (0.14)	19 (0.22)	13 (0.15)	23 (0.26)	52 (0.60)	23 (0.26)
YLW	59	44 (0.75)	4 (0.07)	12 (0.20)	5 (0.08)	5 (0.08)	18 (0.31)	26 (0.44)	18 (0.31)
Mean	96.17	79.17 (0.85)	8 (0.10)	11.33 (0.12)	17.33 (0.17)	16.33 (0.16)	26.17 (0.30)	53 (0.55)	26.17 (0.30)
SD	39.19	26.56 (0.13)	3.41 (0.07)	3.56 (0.04)	10.48 (0.05)	9.03 (0.04)	6.37 (0.11)	22.18 (0.06)	6.37 (0.11)

Table 3: Global inferences generated

Player	Turns	Total CPs (per turn)	Requested (per turn)	Proposed (per turn)	Total A + U (per turn)	Accepted (per turn)	Used (per turn)
BLA	116	28 (0.24)	3 (0.03)	25 (0.22)	9 (0.08)	3 (0.03)	6 (0.05)
BLU	98	30 (0.31)	2 (0.02)	28 (0.29)	15 (0.15)	7 (0.07)	8 (0.08)
PRP	56	15 (0.27)	3 (0.05)	12 (0.21)	6 (0.11)	0 (0.00)	6 (0.11)
RED	161	38 (0.24)	4 (0.02)	34 (0.21)	18 (0.11)	7 (0.04)	11 (0.07)
WHT	87	25 (0.29)	7 (0.08)	18 (0.21)	9 (0.10)	3 (0.03)	6 (0.07)
YLW	59	20 (0.34)	3 (0.05)	17 (0.29)	11 (0.19)	5 (0.08)	6 (0.10)
Mean	96.17	26 (0.28)	3.67 (0.04)	22.33 (0.24)	11.33 (0.12)	4.17 (0.04)	7.17 (0.08)
SD	39.19	8.02 (0.04)	1.75 (0.02)	8.12 (0.04)	4.41 (0.04)	2.71 (0.03)	2.04 (0.02)

Table 4: Conceptual pacts by participant

script, of which 134 (M=22.33, s=8.12) were proposed CPs and 23 (M=3.67, s=1.75) were requested CPs. Thus on average participants were around 6 times as likely to propose a CP as they were to request one. 52 of the counted CPs were coded as unused, as it was ambiguous as to whether or not they were accepted by other participants before the discourse focus switched. 21 CPs were marked as altered. Of these 9 (42.8%) were marked as overspecific, 5 (23.8%) as underspecific, and 7 (33.4%) were judged as being altered due to lexical availability issues. No correlation was observed between quantity of turns taken or CPs generated and number of altered CPs.

Table 4 shows a breakdown by participant of the number of requested, proposed, accepted, used, and combined accepted & used CPs. As use of a CP implies acceptability, it was thought useful to use a combined measure as well as the two individual ones. As a proportion of total produced CPs, the average amount of accepted and/or used CPs was 43%.

4.2 Discussion

The overall categories of local and global inference are not directly comparable. While both make use of the same cognitive mechanisms, it is widely attested that local inferences are more automatised and consume fewer mental resources than do global inferences, which are more strategic, slow, and effortful (Kendeou, 2015; Kendeou, van den Broek, Helder, et al., 2014; Paul van den Broek, Beker, et al., 2015; van den Broek, Fletcher, et al., 1993). Of the local inferences, ellipsis was by far the least commonly occurring. It is unlikely that this is due to difficulty in producing such inferences, and instead is likely to be a simple result of the fast-paced nature of the game, with high levels of interaction between participants and generally short turns.

As has been stated—and as can be gleaned from the exemplar images in Figures 1-4—the game of Mysterium is one which contains a significant number of abstract visual elements, in particular the vision cards distributed by the ghost. This necessitates the players to work together to establish a system to refer to each of the cards on the table. As we know from Relevance Theory, communication is geared towards the maximisation of a cognitive effect for the minimum processing effort (Sperber & Wilson, 1986; Clark, 2013), so we would expect to see participants trying to refer to any entity in the most efficient ways possible, from a communicative perspective, but also a processing perspective. After the initial coding, the referent chain (RC) category was revisited to investigate what proportion of the 105 total inferences made use of deictic referents such as *it*, *this N*, or *that*. 88 out of the 105 (83.8%) were found to contain pronomial terms rather than explicit NP referents. Gundel et al.'s (1993) Givenness Hierarchy proposes six cognitive statuses which are derived from the form of referring expression used in natural discourse: in-focus, activated, familiar, uniquely identifiable, referential, and type identifiable (p275). For example, if a speaker uses *it* in discourse, the Givenness Hierarchy states that the referent is currently in focus. The implication of this is that the participants in this study were particularly good at monitoring the discourse status of any given referent, being able to use pronomial referents (often accompanied by gesture) more than three quarters of the time. However, according to Arnold (2010), speakers tend to prefer underspecific forms such as pronouns when they are experiencing an increased cognitive load in addition to situations where the referent's discourse status is already prominent. Thus the demands of the task may have limited the players' ability to generate more specific referents than these underspecific pronouns. A third factor to be considered is the alignment of referring expressions described by Brennan & Clark (1996) (see also Pickering & Garrod, 2004). Once gesture accompanied by a pronomial NP such as *this one* was established as an effective means of communication, all players will have aligned to this approach rather than attempting a different system for referring.

This was best illustrated by the third stage of the game, which involves the players identifying a weapon from a range of possible alternatives. While not captured by the quantitative data, at this later part of the data collection session, there was a noticeable shift towards referring to the weapons by saying (for example) *the stool* or *the secateurs*, instead of relying upon gesture. Indeed, the weapons in Mysterium are far more easily identifiable when compared with the locations or suspects (see Figs. 2-4 for examples of each), as they are singular, familiar looking objects, each uniquely identifiable from the set of alternatives. There was some brief debate over how to refer to one particular weapon which resembled a small statue, but this began by referring to *the statue of the lady* and quickly became simply *the statue.* All of this suggests that the players were consistently motivated with attempting to communicate in the most efficient way possible. During the earlier, more abstract stages, effort could be saved by using gesture, which was more reliable a way of referring than making verbal descriptions which would rely on assumptions about what was manifest to other players about particular cards. When participants reached the weapon-finding stage, where the alternatives are easily and uniquely identifiable, requiring considerably less assumption about the shared mental state of the other participants, all

Figure 4: An example of one of the player weapon cards

switched to using *the N* to refer. This was marked by a reduction in the rate of requested conceptual pacts, and is consistent with the findings of Brennan & Clark (1996) who found that a single dominant conceptualisation lead to lower rates of hedging, hesitancy and disfluency compared to when multiple competing options were available. However one example from early in the transcript stands out as an apparent rejection of this preference for dominant conceptualisations.

(11) BLK: So that persons looks like they've got, erm, they're sabotaging the other person's balloon by *stabbing a hole* in it
Several unrelated turns
RED: Going back to yours (addressing BLK), it's obvious but there's the red thread, and that reminds me of the red thread [gestures between BLK vision card and suspect card] and you've got the ropes
BLU: I thought that [gestures to same suspect] was BLK's as well until she said about the *pokey thing*... that made me wonder about [gestures to alternative suspect], just because it has a *pokey thing*
WHT: It's a *syringe*
BLK: What's the other *pokey thing*?
RED: This thing? I guess
Several unrelated turns
RED: The only thing I see that's similar there is the wood on the handle and the wood on the *pokey pole* PRP: There is a *vaccine* here

In the above example, referring expressions relating to the same entity have been italicised. Notice that the referring expression is never truly stable, and that several very explicit conceptual pacts (syringe, vaccine) are proposed and at least go unused, if not outright rejected. The object being referred to was indeed a syringe, and yet the closest to stability that this CP comes is the overall concept of stabbing or poking. This appears to be at odds with Brennan & Clark's (ibid.) suggestion that a dominant conceptualisation is preferable, and it is possible that the demands of the activity had lead to some participants (specifically RED, BLU, and BLK) at identifying *poking* as the most salient feature, and rejecting more specific interpretations which may limit their ability to succeed at the game. Indeed, the most commonly rated reason for an altered CP was overspecificity (42.8%), suggesting that the participants were inclined towards making strategic inferences, and thus keeping their inferential options (as it were) open by activating concepts and themes (such as *pokiness*) rather than concrete exemplars such as *syringe* or *vaccine.*

That the most frequently generated global inferences were thematic is perhaps unsurprising, given the highly associative nature of the game. Finding links between cards necessitates finding commonalities and themes which are present across multiple images. This is in stark contrast with the results of McGinnis et al (2007) whose study found that thematic inferences were the least common of all those quantified. However that and similar studies have focused on inferences generated during text comprehension, and tend to define thematic inferences as relating to the themes of the overall text, and thus an average of 0.77 thematic inferences for a 12-sentence text (as in McGinnis et al) is less surprising. Participants in this study had many more opportunities to generate thematic associations between not only their visions and targets, but those of other players as well. In a charming example of the connected, associative nature of activation when generating inferences (eg. Paul van den Broek, Kendeou, et al., 2005), the following interaction occurred towards the mid-stage of the game:

(12) BLU: Ah OK. I have a Rapunzel-like tower, with a black knight
BLK: Oh, secateurs! They can cut hair!

No mention of hair had previously been made, but hearing *Rapunzel* clearly activated the concept of cutting hair for BLK, and thus lead to the association with the secateurs. This integration of background knowledge from beyond what is explicit in the discourse is a textbook example of global inference, and as one of several examples of similar interactions present in the transcript, demonstrates the potential for Mysterium as an intervention tool for scaffolding the inferencing aspect of reading comprehension.

4.3 Limitations and Future Considerations

The approach used here has not been without difficulties. First and foremost, the use of only audio recording for such a visual game proved challenging to code.

Participants' reliance on referring to cards in play by pointing or tapping and saying *this* or *that one* rather than coming to agree upon a referent term was linguistically relevant, inasmuch as it was more efficient to refer in this way than to try to construct more complex referent terms. But keeping track of which *this* referred to which *that* made the coding process slow and difficult. For example, when referring to the selection of weapon alternatives, BLK produced the utterance *what's that thing on the end there.* It was judged that this utterance was down to lack of visibility (the weapon cards are smaller than the others, and BLK had earlier made mention of their poor eyesight), rather than uncertainty over what the depicted object was. Ambiguities such as this one could be easily mitigated in future iterations of this approach by taking a video recording. An alternative option, which could completely prevent the use of pointing as a method of identifying referents could be to run the game online using a service such as Tabletop Simulator or Board Game Arena. This however adds a technical overhead and expense to the process of organising an otherwise simple game session, to mention nothing of the known pragmatic difficulties associated with digital communication (eg. Seuren et al., 2021). Additionally, this has the possibility to stifle production, given the cognitive load of both the task and attempting to refer to abstract images (Arnold, 2010). A simpler solution may be to add a rule which bans gesture altogether, thus encouraging participants to try to come up with purely spoken ways of referring to cards. Given the potential use of such a framework for use in a reading comprehension study, a domain in which gesture is all but impossible to use, such an alteration may indeed be desirable.

The participants who did not have English as their first language produced notably fewer turns than those who did, almost half as many on average (M=67.3, s=13.9 vs. M=125, s=26.5). As was stated previously, the University of Leeds requires all of its students who do not speak English as a first language to demonstrate sufficient language proficiency in order for them to begin their studies. Thus, there are no concerns as to the *quality* of the language produced by these particular participants. Indeed, it is noteworthy that the participant who produced the highest number of conceptual pacts per turn did not have English as a first language (YLW; 0.34 CPs per turn). However, task demands can severely affect L2 processability (eg. Freeman et al., 2022). Furthermore, although Vonk et al. (2022) found that affective state has little impact upon the processing and production of referring expressions, it remains possible that students of any linguistic or cultural background may find playing games with members of staff in their department intimidating. The impact of these factors is not easily quantifiable, except to note again the disparity in the quantity of turns produced. As such, for any future iterations of this approach, a more homogeneous group of participants in terms of L1, age, and SES would be desirable. And while no effect of prior friendship has been observed in this study, some consideration ought to be given to the effect that existing relationships will have on the observed communicative competence between friends (eg. Gorman et al., 2013). Additionally, the MANY MINDS PROBLEM described by Cooney et al. (2020) argues that as communicative groups grow larger, coordination becomes correspondingly more difficult. While operating an intervention of this sort is likely to be impractical in a dyadic situation, the many minds problem should not be discounted, and the use of smaller groups may be advisable.

During the early design of this study, one possible avenue for quantifying inferential ability was vocabulary analysis. The breadth and depth of an individual's vocabulary has a strong relationship with that individual's inferential ability (Federico Sterpin et al., 2021). However, due to time and resource limitations it was not possible to conduct any significant vocabulary analysis with this dataset. With that said, a comparison of the vocabulary choices of participants against a frequency list of words generated from a corpus—such as the British National Corpus—may provide additional insight and an additional dimension with which inferential abilities can be quantified.

Finally, when measuring conceptual pacts, while some small amounts of information as to which other participants made demonstrable use of each CP were recorded, it may prove beneficial in future to record the distance between first and second use of a referential expression. This would allow for an estimation of the changing relevance of a CP over time, and could be a useful additional dimension of quantifiability. Similarly, measuring how rapidly a CP was altered and how many iterations it went through before reaching its final form may provide insight into participants' assumptions about what is sufficiently or insufficiently relevant.

5 Conclusion

This dissertation was concerned with the early development of a framework which can be used to quantify the inference generation and communicative efficacy of players in board game situations, as a means to scaffolding reading comprehension in children with specific difficulties in this area. The methodology of recording, transcribing, and analysing a single board game session proved to generate enough utterances for analysis at least in the context of a pilot study, although as has been stated, the lack of a video element to the data gathering made coding more difficult, and steps should be taken to mitigate that deficit and the analytical challenges it presents in future iterations.

Given the role of this paper as a first step in the development of a framework, it was not within scope to make comparative analyses between the individual participants. The results showed that overall, participants generated an average of 1.22 inferences of various types per turn. While this figure should probably not be used as a baseline for comparison if and when this framework is applied to a study involving children, it does demonstrate that the game Mysterium provides ample opportunity for generating inferences. Additionally, given its lack of written stimuli, the game appears to constitute a good choice for use as an intervention with those experiencing reading comprehension difficulties. The choice to remove the round timer for this session was useful for maximising the data generated, however there may be a benefit to using the timer: according to van den Broek et al. (1993), time-constrained reading activities can be used to produce an approximation of an individual's lower boundary for inferential capacity. With or without this constraint, in the present author's view, Mysterium has been shown to be a good selection for a board game to generate a broad selection of inferences and communicative interactions, although this should not preclude other games with similarly high amounts of player interaction and relatively low mechanical overhead from being trialled.

6 References

Aarts, Bas, 2001. *English Syntax and Argumentation.* Hampshire: Palgrave.

Arnold, Jennifer E., 2008. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, **23** (4), pp.495–527.

Arnold, Jennifer E., 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass*, **4** (4), pp.187–203.

Arnold, Jennifer E., Hudson Kam, Carla L., and Tanenhaus, Michael K., 2007. If you say 'thee uh' you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33** (5), pp.914–930.

Brennan, Susan E. and Clark, Herbert H., 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22** (6), pp.1482–1493.

Carlson, Sarah E., van den Broek, Paul, McMaster, Kristen, Rapp, David N., Bohn-Gettler, Catherine M., Kendeou, Panayiota, and White, Mary Jane, 2014. Effects of comprehension skill on inference generation during reading. *International Journal of Disability, Development and Education*, **61** (3), pp.258–274.

Carnie, Andrew, 2013. *Syntax: A Generative Introduction.* Wiley-Blackwell.

Chastain, Charles, 1975. Reference and context. *Language, Mind, and Knowledge*, **7**, pp.194–269.

Christianson, Kiel, Hollingworth, Andrew, Halliwell, John F., and Ferreira, Fernanda, 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology*, **42**, pp.368–407.

Clark, Billy, 2013. *Relevance Theory.* Cambridge University Press.

Clark, Herbert H., 1996. *Using Language.* Cambridge University Press.

Cook, Anne E. and O'Brien, Edward J., 2014. Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*, **51** (1-2), pp.26–49.

Cooney, Gus, Mastroianni, Adam M., Abi-Esber, Nicole, and Wood Brooks, Alison, 2020. The many minds problem: disclosure in dyadic versus group conversation. *Current Opinion in Psychology*, **31**, pp.22–27.

Federico Sterpin, Lucas, Soledad Ortiz, Sofía, Formoso, Jésica, and Barreyro, Juan Pablo, 2021. The role of vocabuluary knowledge on inference generation: a meta-analysis. *Psychology of Language and Communication*, **25** (1), pp.168–193.

Fincher-Kiefer, Rebecca, 1993. The role of predictive inferences in situation model construction. *Discourse Processes*, **16**, pp.99–124.

Fox, Danny, 2000. *Economy and Semantic Interpretation.* MIT Press.

Freeman, Max R., Blumenfeld, Henrike K., Carlson, Matthew T., and Marian, Viorica, 2022. First language influence on speech perception depends on task demands. *Language and Speech*, **65** (1), pp.28–51.

Garnham, Alan and Oakhill, Jane, 1992. Discourse processing and text representation from a "mental models" perspective. *Language and Cognitive Processes,* **7** (3-4), pp.193–204.

Garrod, Simon and Anderson, Anthony, 1987. Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition*, **27**, pp.181–218.

Gerrig, Richard J. and O'Brien, Edward J., 2005. The scope of memory-based processing. *Discourse Processes*, **39** (2), pp.225–242.

Gibbs, Raymond W. Jr, 1995. Metaphor as a constraint on text understanding. In: Britton, Bruce K. and Graesser, Arthur C., eds. *Models of Understanding Text.* London: Psychology Press, pp.215–240.

Glenberg, Arthur M., Meyer, Marion, and Lindem, Karen, 1987. Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, **26**, pp.69–83.

Gorman, Kristen S., Gegg-Harrison, Whitney, Marsh, Chelsea R., and Tanenhaus, Michael K., 2013. What's learned together stays together: speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, **39** (3), pp.843–853.

Graesser, Arthur C. and Britton, Bruce K., 1995. Five metaphors for understanding text. In: Britton, Bruce K. and Graesser, Arthur C., eds. *Models of Understanding Text*. London: Psychology Press, pp.341–351.

Graesser, Arthur C. and Kreuz, Roger J., 1993. A theory of inference generation during text comprehension. *Discourse Processes*, **16** (1-2), pp.145–160.

Graesser, Arthur C., Millis, Keith K., and Zwaan, Rolf A., 1997. Discourse comprehension. *Annual Review of Psychology*, **48**, pp.163–189.

Graesser, Arthur C., Singer, Murray, and Trabasso, Tom, 1994. Constructing inferences during narrative text comprehension. *Psychological Review*, **101** (3), pp.371–395.

Gundel, Jeanette K., Hedberg, Nancy, and Zacharski, Ron, 1993. Cognitive status and the form of referring expressions in discourse. *Language*, **69** (2), pp.274–307.

Haegeman, Liliane, 2006. *Thinking Snytactically*. Oxford: Blackwell.

Kendeou, Panayiota, 2015. A general inference skill. In: O'Brien, Edward J., Cook, Anne E., and Lorch, Robert F. Jr, eds. *Inferences During Reading*. Cambridge University Press, pp.160–181.

Kendeou, Panayiota, Bohn-Gettler, Catherine, White, Mary Jane, and van den Broek, Paul, 2008. Children's inference generation across different media. *Journal of Research in Reading*, **31** (3), pp.259–272.

Kendeou, Panayiota and van den Broek, Paul, 2005. The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, **97** (2), pp.235–245.

Kendeou, Panayiota and van den Broek, Paul, 2007a. The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, **35** (7), pp.1567–1577.

Kendeou, Panayiota and van den Broek, Paul, 2007b. The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, **35** (7), pp.1567–1577.

Kendeou, Panayiota, van den Broek, Paul, Helder, Anne, and Karlsson, Josefine, 2014. A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, **29** (1), pp.10–16.

Kendeou, Panayiota, van den Broek, Paul, White, Mary Jane, and Julie, Lynch, 2007. Comprehension in preschool and early elementary children: Skill development and strategy interventions. In: McNamara, Danielle S., ed. *Reading Comprehension Strategies: Theories, Interventions and Technologies*. London: Lawrence Erlbaum, pp.27–46.

Kintsch, Walter, 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, **95** (2), pp.163–182.

Krauss, Robert M. and Weinheimer, Sidney, 1964. Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, **1**, pp.113–114.

Krauss, Robert M. and Weinheimer, Sidney, 1966. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, **4**, pp.343–346.

McGinnis, Debra, Goss, R. Justin, Tessmer, Courtney, and Zelinksi, Elizabeth M., 2007. Inference generation in young, young–old and old–old adults: Evidence for semantic architecture stability. *Applied Cognitive Psychology*, **22**, pp.171–192.

Morrow, Daniel G., Bower, Gordon H., and Greenspan, Steven L., 1990. Situation-based inferences during narrative comprehension. In: Graesser, Arthur C. and Bower, Gordon H., eds. *Inferences and Text Comprehension.* Academic Press Inc., pp.123–135.

Pickering, Martin J. and Garrod, Simon, 2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, **27**, pp.169–226.

Rickheit, Gert, Schnotz, Wolfgang, and Strohner, Hans, 1985. The concept of inference in discourse comprehension. In: Rickheit, Gert and Strohner, Hans, eds. *Inferences in Text Processing*. North-Holland: Elsevier Science Publishers, pp.3–49.

Seuren, Lucas M., Wherton, Joseph, Greenhalgh, Trisha, and Shaw, Sara E., 2021. Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics*, **172**, pp.63–78.

Sperber, Dan and Noveck, Ira A., 2005. Introduction. In: Noveck, Ira A. and Sperber, Dan, eds. *Experimental Pragmatics.* Palgrave Macmillan, pp.1–24.

Sperber, Dan and Wilson, Deirdre, 1986. *Relevance: Communication & Cognition.* Oxford: Blackwell.

Swinney, David A. and Osterhout, Lee, 1990. Inference generation during auditory language comprehension. In: Graesser, Arthur C. and Bower, Gordon H., eds. *Inferences and Text Comprehension.* Academic Press Inc., pp.17–33.

Trabasso, Tom and Suh, Soyoung, 1993. Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes*, **16**, pp.3–34.

van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., and White, M. J., 2011. When a reader meets a text: The role of standards of coherence in reading comprehension. In: McCrudden, M.T., Magliano, J., and Schraw, G., eds. *Relevance Instructions and Goal-Focusing in Text Learning*. Greenwich: Information Age Publishing, pp.123–40.

van den Broek, Paul, Beker, Katinka, and Oudega, Marja, 2015. Inference generation in text comprehension: Automatic and strategic processes in the construction of a mental representation. In: O'Brien, Edward J., Cook, Anne E., and Lorch, Robert F. Jr, eds. *Inferences During Reading*. Cambridge University Press, pp.94–121.

van den Broek, Paul, Fletcher, Charles R., and Risden, Kirsten, 1993. Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes*, **16** (1-2), pp.169–180.

van den Broek, Paul, Kendeou, Panayiota, Kremer, Kathleen, Lynch, Julie, Butler, Jason, White, Mary Jane, and Pugzles Lorch, Elizabeth, 2005. Assessment of comprehension abilities in young children. In: Paris, Scott G. and Stahl, Steven A., eds. *Children's Reading Comprehension and Assessment.* Mahwah: Lawrence Erlbaum, pp.107–130.

van den Broek, Paul, Lorch Jr., Robert F., Linderholm, Tracy, and Gustafson, Mary, 2001. The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, **29** (8), pp.1081–1087.

Van der Henst, Jean-Baptiste and Sperber, Dan, 2005. Testing the cognitive and communicative principles of relevance. In: Noveck, Ira A. and Sperber, Dan, eds. *Experimental Pragmatics.* Palgrave Macmillan, pp.141–171.

van Dijk, Teun A. and Kintsch, Walter, 1983. *Strategies of Discourse Comprehension.* New York: Academic Press.

Vonk, Patrick, Goudbeek, Martijn, and Krahmer, Emiel, 2022. The influence of affect on the production of referring expressions. *Language, Cognition, and Neuroscience*, **37** (3), pp.348–364.

Zwaan, Rolf A. and Radvansky, Gabriel A., 1998. Situation models in language comprehension and memory. *Psychological Bulletin*, **123** (2), pp.162–185.

A Inference Taxonomy Coding Table

Inference Type			Description	Examples from transcript (bolded and italicised)
Local	Grammatical	Anaphor	A noun phrase which takes its meaning from another NP in the utterance	So this guy is on a weird cycle, with three hot air balloons behind ***him***
		Ellipsis	Deletion of an already-uttered NP or VP in subsequent structures where it would otherwise be repeated	This is also in an old shed, or (***in***) something wooden
	Cohesive	Referent Chain	Sequential referential expressions referring to the same entity	There's ***two hot air balloons*** ... > The ***balloons***
Global	Elaborative	Character	Reference to a character's physical traits, social/work roles, emotional state, or current activities	She's got a ***big long neck*** I thought that lady was a ***tailor*** Also she doesn't have ***any human feeling*** He's ***carrying ropes***
		Place	Reference to the name of a place or setting	It's like the ***Eiffel Tower*** in the background there
		Object state	Reference to the state, condition, or a property of an object	It's more ***forky*** than ***wedgey***
		Instrument	An object used to accomplish an explicitly stated action	Because for tailoring we use like ***ribbons*** a lot
	Evaluative	Thematic	Summaries of an overall theme	***It seems like about travelling and countries***

B Information and Consent Form

UNIVERSITY OF LEEDS

Information sheet and consent form

LING5350M Dissertation (Linguistics and Phonetics)

Research project: Developing a Framework for Measuring Inference Generation During Gameplay

13/7/22

You are being invited to take part in a research project. Before you decide to participate it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Feel free to ask if there is anything that is not clear or if you would like more information. Thank you for taking the time to read this.

1. What is the purpose of the project?

This project aims to develop a framework for measuring linguistic interactions between players during board games. This framework may be used in future research into assisting children who have reading comprehension difficulties, by investigating the possibility that board games can help to scaffold their inferencing abilities. At this development stage, we are seeking ideal data from adult gameplay, to use in order to model the specific kinds of interaction that we can expect to see, and to create a quantitative system which can be applied over the course of future longitudinal study.

2. Why have I been chosen?

We are recruiting adult speakers of English as a first language. It is up to you to decide whether or not to take part. Refusal to take part will not affect your rights in anyway. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form). If you decide to take part you are still free to withdraw at any time, without penalty or loss of benefits and without giving a reason.

3. What will happen if I take part?

You will be asked to play the board game Mysterium, a co-operative board game where one player acts as a ghost to communicate with the remaining players through the use of abstract vision cards, in order to solve a murder. You may already be familiar with the game, however the rules of the game may be slightly modified to improve the facility of the data for the purposes of this study.

4. What happens when the research study stops?

If the study ends before it is completed, you will, of course, be told why. Once the data collection is completed, there will be some time spent in the analysis and interpretation. The anonymised findings will appear in the lead researcher’s dissertation, and may later be shared with the academic and relevant professional communities through articles in academic journals, or presentations at conferences. They may also be mentioned in future research grant applications by staff from the University of Leeds. You will be able to contact us after the data collection is finished if you have questions or would like to hear the outcome of the study.

5. Will our taking part in this project be kept confidential?

All information which is collected about you during the course of the research will be kept strictly confidential. Any responses/data you provide which are disseminated will be fully anonymised so that you cannot be recognised from it. All information and results are kept in a secure location.

6. What happens to the data collected after the research project is finished?

The data for this project will be stored in a restricted-access location on the University of Leeds' computer network, for as long as is required by the University's assessment regulations. (This is at least until the student identified below has graduated.) During this time the data will not be accessible to anyone except University of Leeds staff. After this time all electronic data and personal details will be deleted, and hard copies destroyed.

7. What will happen if I change my mind about participating?

You are free to withdraw from participation at any time with no need for explanation and no penalty, up to one month after the point of data collection. If so, please contact the student identified above who will immediately withdraw your data and personal information from their project; all such information in both electronic form and/or hard copy will be destroyed.

8. Who is organising and funding the research?

This research is being undertaken by Billie O'Dwyer, based at the department of Linguistics and Phonetics at Leeds University, and is supervised by Dr Cecile De Cat. This project is being conducted as part of a module that has been reviewed and approved by the Faculty Research Ethics Committee at the University of Leeds (Ethics reference: PVAR 17-128).

Contact for further information:

Billie O'Dwyer

Department of Linguistics and Phonetics, School of Modern Languages and Cultures, Michael Sadler Building, University of Leeds, Leeds, LS2 9JT.

Tel: 07564 892960
Email: ml10wod@leeds.ac.uk

UNIVERSITY OF LEEDS

Consent to take part in Developing a Framework for Measuring Inference Generation During Gameplay

	Add your initials next to the statement if you agree
I confirm that I have read and understand the information sheet dated 13/7/22 explaining the above research project and I have had the opportunity to ask questions about the project.	
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences, up to one month after the data are recorded. In addition, should I not wish to answer any particular question or questions, I am free to decline. I understand that if I withdraw, all electronic data will be permanently deleted and hard copies destroyed.	
I give permission for members of the research team to have access to my anonymised responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research. I understand that my personal information will be kept strictly confidential.	
I understand that after the study is completed, all electronic data will be permanently deleted and hard copies destroyed.	
I agree to take part in the above research project and will inform the lead researcher should my contact details change.	

Name of participant	
Participant's signature	
Date	
Name of lead researcher	
Signature	
Date	

Please answer the following questions to the best of your knowledge. If you would rather not answer some (all) of the questions, then just leave these slots blank.

1. *What is your age and date of birth?*

2. *Is English your first language? If not, what is/are your first language(s)?*

3. *How often do you play board games?*
   (Daily / Several times a week / Once a week / Several times a month / Once a month / Less than once a month)

4. *Have you played Mysterium before, and if yes, approximately how many times?*