

A Computational Narrative Analysis of Children-Parent Attachment Relationships*

Iraide Zipitria¹ and Nerea Portu-Zapirain²

- 1 Psychology Faculty,
The University of the Basque Country (UPV/EHU)
Tolosa Hiribidea 70, Donostia (Spain)
iraide.zipitria@ehu.es
- 2 Faculty of Philosophy and Education,
The University of the Basque Country (UPV/EHU)
Tolosa Hiribidea 70, Donostia (Spain)
nerea.portu@ehu.es

Abstract

Children narratives implicitly represent their experiences and emotions. The relationships infants establish with their environment will shape their relationships with others and the concept of themselves. In this context, the Attachment Story Completion Task (ASCT) contains a series of unfinished stories to project the self in relation to attachment. Unfinished story procedures present a dilemma which needs to be solved and a codification of the secure, secure/insecure or insecure attachment categories. This paper analyses a story-corpus to explain 3 to 6 year old children-parent attachment relationships. It is a computational approach to exploring attachment representational models in two unfinished story-lines: *The stolen bike* and *The present*. The resulting corpora contains 184 stories in one corpus and 170 stories in the other. The Latent Semantic Analysis (LSA) and Linguistic Inquiry and Word Count (LIWC) computational frameworks observe the emotions which children project. As a result, the computational analysis of the children mental representational model, in both corpora, have shown to be comparable to expert judgements in attachment categorization.

1998 ACM Subject Classification I.2.0 General-Cognitive Simulation; J.4 Social and Behavioral Sciences-Psychology

Keywords and phrases latent semantic analysis, LIWC, representational models, attachment relationships, unfinished stories

Digital Object Identifier 10.4230/OASICS.CMN.2014.251

1 Introduction

The way children view the world is based on the relationships established with their caretakers. Children perceive themselves according to the way they perceive their relationships with caretaker. But, how do we get to know the way children see themselves and the way they are perceived by others? Children-parent attachment relationships are reflected in their speech and attitudes. Children live in a fantasy world. Because of this, we can connect with their emotions through stories and narrative. These emotions are relevant because the understanding of the world developed in infancy will persist over time. Thus, the most

* This work was partially supported by the Eusko Jaurlaritza (Basque Government) SAIOTEK program grant (S-PE13UN178).



© Iraide Zipitria and Nerea Portu-Zapirain;
licensed under Creative Commons License CC-BY

5th Workshop on Computational Models of Narrative (CMN'14).

Editors: Mark A. Finlayson, Jan Christoph Meister, and Emile G. Bruneau; pp. 251–268

OpenAccess Series in Informatics



OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

popular methodology to explore children attachment relationships with their parents are unfinished stories [25].

Hence, the aim of this study is to explore and analyse attachment relationships on the basis of the representational models of attachment. The significance of attachment theory [2, 1], considered one of the most important theoretical and empirical constructions in the field of socio-emotional development, is based on the formulation of internal working models of oneself and one's relationships, in close connection with behaviours and feelings.

The internal working models reflect the construction of a mental representation of the world, based on the generalisation of the interactions children experience with their attachment figures during their early relations with the adults that satisfy their needs, and include the internalization of specific attributes and expectations of both their own behaviour (feeling loved, accepted and protected) and the behaviour of their attachment figures. They also constitute a pattern for the relationships that individuals will then establish throughout the rest of their lives [15].

During early childhood, as a result of new cognitive-representative, communicative, social and motor skills, children enter a new phase in the development of attachment, known as goal-corrected partnership. During this complex phase, changes which enable a greater diversity of behaviour occur. Thus, it has been observed that attachment behaviours are activated less frequently and with less intensity, since dyadic modulation patterns for ensuring emotional balance are well established; in other words, physical contact, while still necessary, gradually develops into psychological contact. The relationship is internalized and becomes representational. Children become more autonomous and emotionally self-regulating. The moral self emerges, reflected in child's ability to defer behaviour, abide by rules and correct their behaviour in the absence of the attachment figure. Nevertheless, the most significant aspect of this new phase is that members of the attachment pair begin to operate in accordance with a set of shared plans and objectives, thus fostering a closer, more intimate relationship which lays the ground-work for the development of more complex partner or social relations, which later extend to peer relations and relations with other significant adults [5, 28].

The mental model of the relationship is more elaborate and better adjusted to reality. Coupled with increased communicative skills, the cognitive changes that occur enable a more appropriate expression of demands, the communication of internal states and dialogue. Little by little, children begin to be able to infer the goals and understand the intentions, feelings and emotions of their attachment figure. These contents are all incorporated into this more complex structure, thus enabling children to operate internally both with their own perception and representation, and those of their attachment figure.

Based on the proposals forwarded by [2], other authors (e.g. [3, 4]) have made interesting contributions which have helped underscore the importance of the gradual emergence of models and their changes during early childhood. This is a particularly significant period for children's development and growth because it is during this time that certain components become consolidated as "scripts" or knowledge structures.

Attachment during early childhood has been studied very little. Today, however, it is the subject of several interesting investigations, and according to [2], there is still much to be discovered in this sphere. Focusing on these developmental attachment patterns and the contribution of parents may provide some guidelines for exploring this age range in more detail.

Affective states (attachment) of 3 to 6-year-old children can be evaluated by means of projective measures [30]. These tests are most adequate in the evaluation of affective states for this age range. Children express their feelings using the Attachment Story Completion Task (ASCT) [29]. Based on the story, children reflect the type of affective relationship.

However, the evaluation of the information by experts, despite the systematic detail description by [29], is difficult to measure and extremely time consuming. Therefore, the use of computational means for the comprehension of projective measures could be a beneficial application of computational language analysis for the detection of this psychological phenomena. But, is it possible? Can we model the required semantics for attachment representation? Is a computational analysis of semantic representation of children-parent attachment relationship possible? Are those implicit emotions sufficiently present in language to be detected by language representation models?

There are theories of language (embodied cognition) that argue that meaning can be captured only by grounding linguistic symbols (words) in the human body and its interaction with the environment. Other theories argue that meaning can be captured by their relation to other symbols (words) [16]. However, language has also been considered both symbolic and embodied, both processes converge and meaning is encoded in both featural and distributional information [16]. Hence, some authors suggest that the relevant information to extract semantic categories is coded redundantly on perceptual and linguistic experience. Thus, distributional models based on semantic knowledge are based on regularities or word co-occurrences. The more similar the contexts in which two words appear, the more similar their meaning. In contrast, feature-based semantic representation is a list of descriptive features which represent meaning [23].

Hence, context semantic representation of language based on word co-occurrence is modelled by Latent Semantic Analysis (LSA) [13], Hyperspace Analogue to Language (HAL) [17] or the Topic model [9]. This study takes a LSA approach to model the distributed semantics behind children-parent attachment relationship representation. Latent Semantic Analysis (LSA) is a statistical corpus-based natural language understanding technique. LSA has been widely used to model semantic similarity in a variety of contexts. Amongst others, LSA has been successful simulating text comprehension and text coherence [8]. LSA was developed by [6] and later found to be comparable to humans in similarity judgements by [13] and [14]. The first achievement was in information retrieval, where LSI gained an improvement between 10% and 30% in the capability to retrieve documents with equivalent meaning but with different words, TREC3 [7]. In addition, LSA has shown to be capable to deal with complex psychological phenomena such as metaphor [10] and predication [11]. Apart from being capable of gathering documents containing the same key-words, LSA is able to gather documents with semantically similar words to the key-word.

However, LSA does not take into account word-order, and does not take into account certain linguistic structures such as negation [12]. There are non semantic linguistic structures which are specifically relevant to secure versus insecure attachment categorization; therefore, alternative approaches should be explored. Thus, linguistic cues have also been gathered by means of the Linguistic Inquiry and Word Count (LIWC) [20, 27]; which has proved to be successful detecting meaningful measures in categories such as attentional focus, emotion and social relationships based on linguistic features [27]. In terms of semantics, LIWC produces linguistic indicators in a feature-based approach.

Thus, is it possible to analyze children-parent attachment experiences by computational means? Is it possible to discriminate between secure and insecure attachment by computational means? In order to address those questions, this paper is organised as follows. In Section 2, there is a description of the studied corpora and the studied attachment relationship categorization. Section 3 describes the different computational means used to analyze attachment emotions. Section 4 contains the data analysis and the results obtained in the studied corpora. Finally, Section 5 refers to final discussion and future work possibilities.

2 Children story corpora

The corpora was created based on stories collected in a previous study [22]. The sample selected comprised stories produced by 111 children (55 boys and 56 girls) from Irun (a town in the Basque Country, Spain), aged between 3 years 9 months and 6 years 3 months. All were in either the 2nd or 3rd year of preschool. All the children in the sample were from intact two-parent families and had lived with both their parents from birth. Parents' consent was requested and received before the trials were administered.

The "Attachment Story Completion Task" (ASCT) [29] was used in this study. The aim of this instrument is to assess participants' mental representation of themselves in relation to attachment to parents and the pattern of communication established in children aged 3 to 6. The most important difference between this measure and classification systems known as "Doll Play" [19] is that in this one, both the father and the mother are main characters in the stories, thus enabling attachment styles to be assessed individually for each parent.

The ASCT procedure consist of a series of story stems (*The stolen bike*, *The present*, *I'm sorry*, *A fight at school* and *A monster in the bedroom*) which are presented and narrated by the researcher using a set of dolls which represent a family in different circumstances. In this paper, the collected stories are about *The stolen bike* and *The present*.

2.1 Corpus one (*The stolen bike*)

The collected stories are about *The stolen bike* (see Figure 1): A teenager he/she does not know steals the bicycle that the child's parents have given him/her (the story represents fear or a external threat). The child is asked to complete the story. Stories feature the father or the mother separately, and are presented in a counterbalanced order. The story has its theme and situation designed to activate attachment.

Some of the initially collected stories were not included in this corpus due to cross-linguistic issues. In some stories some or most of the speech was produced in Basque language and this combination made computation more complex. All the stories were kept in the same language. The resulting corpus is composed of 184 stories (each participant produced a story for each of the parents) and a total of 24550 words: 12061 words are dedicated to father stories and 12492 words to mother stories. Finally, in addition to verbal information, there were expert judgements associated to all the stories [22], where the story attachment levels were categorised and rated using the "Attachment Story Completion Task" [29] coding.

2.2 Corpus two (*The present*)

The story is about "*The present*" (see Figure 2): Upon arriving home from school, the child gives his/her parents a present that he/she made for them (a positive emotional interaction between the child/parent pair, based on a positive social signal emitted by the child).

In the same way as in corpus one, some of the initially collected stories were not included in the corpus. Some for the same cross-linguistic effect and a few due to a technical issue. The resulting corpus is composed of 170 stories (each participant produced a story for each of the parents) and a total of 18071 words: 8757 words are dedicated to father stories, while 9314 words are dedicated to mother stories. In the same way as in corpus one in there were expert judgements associated for all the stories.



■ **Figure 1** Set of dolls representing a family in *“The stolen bike”*.



■ **Figure 2** Set of dolls representing a family in *“the present”*.

2.3 Attachment relationship categories:

Each story ending given by participants is categorised as secure, secure/insecure or insecure and has its own scoring criteria, which are outlined below:

2.3.1 Secure response:

A secure response (4 or 5 points) is scored on the basis of the helpfulness, swiftness and responsiveness of parents' spontaneous response, happy ending to the story, any mention of positive sensations and a positive interaction between the parent/ child pair (care, consolation, assurance that the child is still loved, etc.).

2.3.2 Secure-insecure response:

A secure-insecure response (3 points) is scored on the basis of: not asking for help, absence of active engagement by parents, mention of only negative sensations (anger, physical punishment, etc.), feeling of not being loved, not feeling responsible for their actions, etc.

2.3.3 Insecure response:

A insecure response (1 or 2 points) is when there is no interaction between the parent/child pair or when said interaction is negative.

In ASCT [29] authors suggest that the secure-insecure category needs to be defined to either the secure or the insecure option. Because of this, for data analysis only secure and insecure discrimination will be studied.

3 Computational Analysis of the Narratives involved in the ASCT stories

The analysis of the security emotions exhibited by children in narratives are analysed computationally by means of Latent Semantic Analysis, Linguistic Inquiry Word Count and programs to detect pauses and response eliciting questions.

3.1 Latent Semantic Analysis

LSA bases its knowledge on a corpus where LSA learns word similarities. Next, the vector representation of each word is measured statistically, based on the occurrence of words in the corpus. Finally, the cosine between vectors, measures text to text similarity.

First, we need a corpus which represents the desired semantic knowledge. Next, we will need to be able to measure similarity under the most adequate dimensions for our study. Similarity measures can be computed based on word to word, word to document or document to document comparisons.

3.1.1 Terms

LSA does not learn every word contained in a corpus. Only those terms whose meaning can be learned from context will be understood through LSA. The words understood by LSA are terms [6]. However, what condition should a word have to be considered a term? It should have a minimum amount of characters (default 2), it needs to appear in a document a minimum amount of times (default 1) and it should appear at least in a minimum amount of documents (default 2).

3.1.2 Documents

Every portion of text contained between two blank lines or a file will be considered a document. It will normally be a paragraph, although, if any portion of text (sentences, words, etc.) is contained between two blank lines in every case will be considered a document.

LSA considers the contexts of each term documents in which terms are contained. Therefore, documents selected in the text should be semantically sound.

3.1.3 Dimensions

The dimension ratios tend to be in the range between 50 and 1500 dimensions. Although the closest ratios to human measures tend to be between 100 and 400 dimensions [31]. Nevertheless, the most adequate dimensionality is typically chosen observing how close the similarity measures are under the different dimensions to human decisions.

3.1.4 Weight

In terms of weight, although we have the possibility to choose between three local and three global weights, LSA tends to use $\mathbf{log}(i, j)$ as a local value and $\mathbf{entropy}(i)$ as a global value [14].

LSA considers term weight: local and global. The **local weight** $L(i, j)$ measures the relevance of the i term in the j document and the $G(i)$, **global weight** looks at the relevance of the i term in the whole corpus. Every document has the same level of relevance.

LSA allows three main modes to compute local and global weights:

Local weights can be measured as

1. **tf**: *term frequency*: $L(i, j) = \mathbf{tf}(i, j) = m_{ij}$. It reflects how many times the i term appears in the j document. It is precisely what the source \mathbf{M} matrix measures. Therefore using this method does not make any local change over the source matrix. The greater the frequency of a term in a document, the higher the weight it will have locally.
2. **log**: $L(i, j) = \mathbf{log}(i, j) = \log(m_{ij} + 1)$. The *log* function makes it possible to reduce the difference between frequencies. Then, when writing its logarithm, instead of the frequency values, matrix values become a little more uniform in the distribution.
3. **bin**: $L(i, j) = \mathbf{bin}(i, j) = \min\{m_{ij}, 1\}$. This local measure eliminates frequencies from the m_{ij} matrix, replacing them with binary values. Therefore, if the i value appears at least once in the j document, a 1 value will be added. If the term does not appear, a 0 value will be added.

Global weights can be measured as

1. **none**: $G(i) = \mathbf{none}(i) = 1$. This global weight is constant, therefore the same for every term in the corpus. It aims to give the same relevance to every term in the corpus. When selecting $\mathbf{tf}(i, j)$ as a local value and $\mathbf{none}(i)$ as the global value, the \mathbf{M} source matrix is left in its original configuration.
2. **normal**: $G(i) = \mathbf{normal}(i) = \frac{1}{\|\mathbf{t}_i\|} = \frac{1}{\sqrt{\sum_{j=1}^n m_{ij}^2}}$. When applying t_1, \dots, t_m weight to the terms, all the term vectors are normalised, thus all the vectors have the same length.
3. **idf** or “Inverse Document Frequency”: $G(i) = \mathbf{idf}(i) = \frac{gf(i)}{df(i)}$, where $gf(i)$: “global frequency” measures the term appearance in the whole corpus. And $df(i)$: “document

frequency” looks at how many documents contain each term.

$$df(i) = \sum_{j=1}^n \min\{m_{ij}, 1\}$$

4. entropy:

$$G(i) = \text{entropy}(i) = - \sum_{j=1}^n p_{ij} \log_2(p_{ij}), \quad \text{where} \quad p_{ij} = \frac{m_{ij}}{gf(i)}$$

Measures the lack of balance between terms and documents. The more balance in the frequencies, the higher the entropy.

3.1.5 Similarity measures

Similarity measures are calculated computing the cosine between the two vectors representing the semantic context [6, 13]. Cosine is the similarity measure which is closest to humans in semantic decisions made in vector semantic models. However, other similarity measures are available and have sometimes been tested for similarity purposes, e.g. dot product.

3.1.5.1 Similarity measures in *The Stolen bike*

Critical keywords for *The stolen bike* story were selected considering ASCT [29] coding and the story-based corpus lexicon. In order to gather this lexicon, the corpus was divided into two halves, secure and insecure stories. Both were tokenised and the most frequent story related content-words were selected as security representative keywords. The most frequent insecure tokens were mainly function words and were not very representative of insecurity for this specific story context. Therefore, in *The stolen bike* story-line, non-semantic approaches were found to be more informative for the detection of insecure story-lines.

In *The stolen bike* story, the *help seeking to the parents* was representative of secure narratives. In case the bike was not there, the parents would *help look for the bike until it was found*. Once the bike was found, the parents would assertively ask the robber to *get off the bike*. Often, secure stories would end up *recovering the bike* with a happy ending. Then, secure content keywords such as *bajar, quita (get off), da (gives), buscar (search), encontrar (find), recuperar (recover)* were selected.

Finally, all those term vectors and term combination vectors would be compared to the different stories to obtain similarity measures (cos) between the keywords and each of the stories.

3.1.5.2 Similarity measures in *The present*

The main keywords for *The present* were also selected considering ASCT [29] coding and the story-based corpus lexicon. In order to gather this lexicon, for the hand-coding procedure the corpus was divided into two halves, secure and insecure stories and the most frequent story related content-words were selected. For *the present* story-line, both secure and insecure speech keywords were selected.

A present, both to the eyes of children and parents in the story [29], can be *pretty* or *special* representing positive emotions, or can be *unattractive, uninteresting, useless* or *not good enough*.

■ **Table 1** Selected LIWC measures.

LIWC2007	Description	Dictionary examples
Word count (WC)	Narrative length	–
Words per sentence (WPS)	Cognitive complexity	–
Negation (negate)	Inhibition	<i>No, not, never</i>
Affective processes (affect)	Emotional narratives	<i>Happy, cried, abandon</i>
Positive emotions (posemo)	Positive emotional narratives	<i>Love, nice, sweet</i>
Negative emotions (negemo)	Negative emotional narratives	<i>Hurt, ugly, nasty</i>

Therefore, keywords were labeled as indicators of secure attachment –bonito (pretty), bien (well), contenta/o (happy), gracias (thank you), gusta (like), beso (kiss), abrazo (hug), jugando (playing), bueno (good), etc.

The same procedure was followed to select keywords for insecure attachment. In this case, indicators of insecure attachment such as mal (bad), triste (sad), enfadado/a (angry), castiga (punish), rompe (breaks), feo (ugly) or guardar (hide) were selected.

3.2 Linguistic Inquiry and Word Count

Linguistic Inquiry and Word Count (LIWC) is a text analysis program that counts psychologically meaningful words. Emotionally relevant features are computed in comparison to relevant dictionaries (word-lists). LIWC is able to detect meaningful measures in categories such as attentional focus, emotion, social relationships, thinking styles, and individual differences [27, 20].

LIWC contains word collections for each category. The studied categories are: Linguistic processes, psychological processes, cognitive processes and personal concerns. Those categories contain approximately 80 indicative measures for each psychological phenomenon [27]. Emotions can be detected by means of the identification of emotional features present in language. For example, deceptive language [18] or depression [26].

LIWC measures emotion indicators which are very relevant for the secure versus insecure attachment categorization. The length of the story can be an indicative characteristic of a secure versus an insecure emotional attachment relationship. A long story could mean a lot of explaining and tends to be related to insecure attachment relationships, while secure attachment relationship tend to be represented by a short story with clear facts [29]. Therefore, narrative length (WC) might be representative or indicative of the security level in the attachment relation. However, longer sentences and clear statements have been found to reflect a more complex secure relationship [29]. Sentence length will be represented in (WPS) and is indicative of complex sentences and a greater cognitive complexity [27]. Negative statements (negate) such as *I don't know* are common in an insecure attachment story-line [29]. Finally, negative emotional interaction, aggression and a bad ending (negemo, affect) are indicative of an insecure emotional attachment relationship[29]. These are five LIWC measures which seem to be very related to the ASCT coding and *The stolen bike* story-line. In addition, in the ASCT *The present* story coding positive emotions (posemo) are coded as indicative of a secure relationship. In the same way the absence of negative emotions (negemo) is indicative of a secure relationship. A summary of these measures is listed in Table 1.

3.3 Computational Analysis of Pauses and Response Eliciting Questions

Pauses are common during narratives and the flow of the story is recovered by posing questions to redirect attention to the story.

3.3.1 Pauses

Pauses are considered in speech in terms of cognitive processing, affective-state, and social interaction [24]. Pauses, far from being empty of meaning, gather a great amount of information that needs to be considered. In the context of the “Attachment Story Completion Task” [29], the pause is a latency produced by the effect of the stimulus story-context in the narrative response of the child. Therefore, long pauses will have an affective nature and its length is most likely to be related to insecure affective relationships.

None of the previous resources take pauses into consideration, therefore a program was developed ad hoc to measure the presence of pauses in the narrative speech of children.

3.3.2 Response Eliciting Questions (REQ)

In addition to the previously described pauses, another important linguistic cue is questions. Once the pauses break the narrative flow or if the attention is directed to unrelated matters, posing Response Eliciting Questions (REQ) is required in the context of the “Attachment Story Completion Task” [29] to encourage the narrative process. In *The stolen bike* story, the need for more specific response eliciting questions is representative of insecure attachment relationships [29].

None of the previous resources take REQ into consideration, therefore a program was developed ad hoc to measure the presence of those questions in children narrative speech interaction.

4 Attachment Security Detection

Computational analysis of semantic indicators, linguistic cues and pauses were compared to ASCT coded judgements to observe whether the measures were significantly related and if these indicators were capable of significantly discriminating between secure and insecure attachment.

4.1 LSA vector representation of Semantic Attachment Security

LSA semantic spaces were created for both corpora and the resulting semantic representations were analysed separately for each story context.

4.1.1 Corpus one: *The stolen bike*

In *The stolen bike* story expressions such as “*baja / baje de la bicileta*” (“*get/got off the bike*”) or “*recupero / recupera / recuperan la bicileta*” (“*recover / recovered the bike*”) are representative of secure attachment. Vectors of content-words such as *bajar*, *quita* (*get off*), *da* (*gives*), *buscar* (*search*), *encontrar* (*find*), *recuperar* (*recover*) were selected and the combination of vectors were selected as detectors. Those vectors were compared to the different stories to observe whether the resulting similarity measures (cos) between the vectors and each of the stories were related to security coding (see Table 2).

■ **Table 2** LSA Semantic Indicator Relatedness to Expert Measures in *The stolen bike*.

LSA Term Vector	Spearman's ρ	Significance
Baja / Baje / Recupera /Recuperado / Recuperan (Get off / Recover)	0.38	$p < 0.01$
Baja / Recuperan (Get off / Recover)	0.36	$p < 0.01$
Baja / Baje (Get /got off)	0.29	$p < 0.01$
Baja / (get off)	0.31	$p < 0.01$
Recupera / Recuperado / Recuperan (Recover / Recovered)	0.33	$p < 0.01$
Recuperan (Recover)	0.22	$p < 0.01$

■ **Table 3** LSA Security Indicator Relatedness to Expert Measures in *The present*.

LSA Term Vector	Spearman's ρ	Significance
Contenta (Happy - she case)	0.19	$p < 0.05$
Contenta / contento / contentos (Happy - she/he/we cases)	0.24	$p < 0.01$

All the measures show significant relatedness to secure attachment expressions. However, it is clear that the first vector, “*baja / baje / recupera / recuperado / recuperan (Get - got off / recover / recovered)*” is best related to the expert coding. Therefore, this vector will be selected as the most salient semantic detector.

Next, we will observe if there are differences between security categories. A Kruskal-Wallis test shows that there are significant differences in terms of the studied three categories (secure, secure-insecure and insecure), $H(3) = 14.13$ and $p < 0.05$. Post hoc test using Man Whitney shows that there are significant differences between secure and insecure attachment measures, $U = 1988$; $p < 0.01$ and Cohen's $d = 0.54$.

4.1.2 Corpus two: *The present*

In *The present* there were indicators for both security and insecurity.

Representative of **security** were keywords such as bonito (pretty), bien (well), contenta/o (happy), gracias (thank you), gusta (like), beso (kiss), abrazo (hug), jugando (playing), bueno (good), etc. However, only *contenta* or the combined *contenta*, *contentos*, *contento* vectors produced significant associations (see Table 3).

Thus, secure attachment is best represented by “*Contenta/o-s (Happy)*”, $\rho = 0.24$ and $p < 0.01$. A Kruskal-Wallis test shows that as security indicator significantly differentiates in terms of the studied three categories (secure, secure-insecure and insecure), $H(3) = 12.65$ and $p < 0.01$. Post hoc test using Man Whitney shows that there are significant differences between secure and insecure attachment measures, $U = 1512$; $p < 0.01$ and Cohen's $d = 0.71$.

Representative of **insecurity** were mal (bad), triste (sad), enfadado/a (angry), castiga (punish), rompe (breaks), feo (ugly) or guardar (hide). “*Castiga (Punish)*”, “*Enfadado (Angry)*”, “*Mal (Bad)*” or the combined vectors “*Castiga / enfadado / mal (Punish / angry / bad)*” offer significant associations to expert judgments (see Table 4).

■ **Table 4** LSA Insecurity Indicator Relatedness to Expert Measures in *The present*.

LSA Term Vector	Spearman's ρ	Significance
Castiga (Punish)	-0.2	$p < 0.01$
Enfadado (Angry)	-0.17	$p < 0.05$
Mal (Bad)	-0.15	$p < 0.05$
Castiga / enfadado / mal (Punish / angry / bad)	-0.17	$p < 0.05$

■ **Table 5** LIWC2007 Indicator Relatedness to Expert Measures in *The stolen bike*.

LIWC2007	Spearman's ρ	Significance
WC	-0.2	$p < 0.01$
WPS	0.42	$p < 0.01$
negate	-0.51	$p < 0.01$
affect	-0.2	$p < 0.01$
negemo	-0.37	$p < 0.01$

Nevertheless, none of the insecurity indicators produce significant differences in secure, secure/insecure or insecure category discrimination.

4.2 LIWC indicators

In addition to distributed semantic indicators, there are other lexical and linguistic cues which are relevant as attachment security indicators: (1) the length of the story (WC), (2) the length of the sentence (WPS), (3) negation (negate), affective expressions (affect), positive emotions (posemo) and negative emotions (negemo).

4.2.1 Corpus one: *The stolen bike*

The indicators were compared to expert judgments to analyse which of them are most related (see Table 5).

All the indicators except posemo show significant relatedness to attachment expressions. Positive emotion was not significant, which is expected in the *The stolen bike* story line because it is created to evoke an external threat. Therefore, most of the selected LIWC 2007 indicators are representative of insecure attachment. For instance, negation is a good indicator for insecure attachment ($\rho = -0.51$ and $p < 0.01$). A high amount of negation implies a low attachment security measure (or insecure attachment). In addition, the presence of negative emotion expressions ($\rho = -0.37$ and $p < 0.01$), affective lexicon ($\rho = -0.2$ and $p < 0.01$) and the length of the story ($\rho = -0.2$ and $p < 0.01$) are also representative of insecure attachment. The higher the amount of negative emotion expressions, affective lexicon and longer stories, the more insecure the attachment measure is.

However, one of the LIWC 2007 indicators, sentence length (WPS), is well associated to secure attachment ($\rho = 0.42$ and $p < 0.01$). Long sentences are related to secure attachment.

All the LIWC 2007 indicators were tested but only those ASCT related measures showed to be significantly related.

Attachment security was significantly different across the studied narrative representations. Kruskal-Wallis test shows that there are significant differences for length of narratives, $H(3) = 2.77$ and $p < 0.05$ and post hoc test using Man Whitney show that there are

■ **Table 6** LIWC2007 Indicator Relatedness to Expert Measures in *The present*.

LIWC2007	Spearman's ρ	Significance
WC	-0.47	$p < 0.01$
WPS	0.11	$p = 0.13$
negate	-0.24	$p < 0.01$
affect	0.14	$p = 0.55$
posemo	0.15	$p = 0.04$
negemo	-0.11	$p = 0.1$

significant differences between the secure and the insecure category, $U = 2254$; $p < 0.01$ and Cohen's $d = 0.42$. The cognitive complexity (WPS) also shows differences in terms of the three security categories, $H(3) = 30.56$ and $p < 0.05$ and post hoc test reflect significant differences between the secure and the insecure category, $U = 1477$; $p < 0.01$ and Cohen's $d = 0.84$. The same was tested for the inhibition (negate) represented in negative expressions, $H(3) = 14.36$ and $p < 0.05$ with post hoc significant differences between the secure and the insecure category, $U = 1647.5$; $p < 0.01$ and Cohen's $d = 0.92$. Another significant difference was found for emotional narratives (affect) $H(3) = 8.6$ and $p < 0.05$ with post hoc significant differences between the secure and the insecure category, $U = 2169.5$; $p < 0.01$ and Cohen's $d = 0.5$. Finally, significant differences were detected in terms of negative emotion expressions (negemo) $H(3) = 24.87$ and $p < 0.05$ with post hoc significant differences between the secure and the insecure category, $U = 1906.5$; $p < 0.01$ and Cohen's $d = 0.78$.

Consequently, all the studied LIWC measures were capable of producing security discrimination for ASCT coding.

4.2.2 Corpus two: *The present*

The indicators were compared to expert judgments to analyse which of them are most related (see Table 6).

Narrative length (WC) ($\rho = -0.47$ and $p < 0.01$), negation ($\rho = -0.24$ and $p < 0.01$) and positive emotion ($\rho = -0.15$ and $p = 0.04$) show association with expert judgements, whilst sentence length (WPS), affective expressions and negative emotion do not show to be associated. In the same way as in corpus one, short stories are representative of a secure attachment relationship. Similarly negative statements are associated with insecure relationships.

The same procedure was applied to observe if the studied indicators were capable of discriminating between attachment security categories. Kruskal-Wallis test shows that there are significant differences for length of narratives, $H(3) = 35.33$ and $p < 0.01$ and post hoc test using Man Whitney shows that there are significant differences between the secure and the insecure category, $U = 1005.5$; $p < 0.01$ and Cohen's $d = 0.94$. The cognitive complexity (WPS) did not show differences in terms of the three security categories, $H(3) = 5.17$ and $p = 0.07$. The same was tested for the inhibition (negate) represented in negative expressions, $H(3) = 16.36$ and $p < 0.01$ with post hoc significant differences between the secure and the insecure category, $U = 1390.5$; $p < 0.01$ and Cohen's $d = 0.83$. There were also significant differences in the case of emotional narratives (affect) $H(3) = 11.46$ and $p < 0.01$ with post hoc significant differences between the secure and the insecure category, $U = 1744.5$; $p = 0.03$ and Cohen's $d = 0.41$. There were not significant differences in terms of negative emotion expressions (negemo) $H(3) = 5.8$ and $p = 0.055$. However, positive

■ **Table 7** Pauses and Response Eliciting Questions (REQ) in *The stolen bike*.

Indicators	Spearman's ρ	Significance
REQ	-0.47	$p < 0.01$
Pauses	-0.49	$p < 0.01$

■ **Table 8** Pauses and Response Eliciting Questions (REQ) in *The present*.

Indicators	Spearman's ρ	Significance
REQ	-0.56	$p < 0.01$
Pauses	-0.45	$p < 0.01$

emotion expressions (posemo) did show capability of significantly detecting attachment security differences $H(3) = 12.96$ and $p < 0.01$ with post hoc significant differences between the secure and the insecure category, $U = 1704$; $p = 0.02$ and Cohen's $d = 0.44$.

4.3 Computational Analysis of Pauses and Response Eliciting Questions

In addition to the observed indicators, there are other narrative processes which are relevant as attachment security indicators: such as pauses (see Section 3.3.1) and response eliciting questions (see Section 3.3.2). Two programs were developed ad hoc to measure the presence of pauses and Response Eliciting Questions (REQ) produced by the expert to encourage the narrative process.

4.3.1 Corpus one: *The stolen bike*

Both indicators were compared to expert judgments to analyse which of them are most related (see Table 7).

Both narrative measures, pauses and response eliciting questions, are significantly related to attachment security representation ($\rho = -0.49$; $p < 0.01$ and $\rho = -0.47$; $p < 0.01$). The higher the amount of pauses and questions, the higher the tendency to insecurity.

Both measures detect significant differences across the studied narrative representations. Kruskal-Wallis test shows that there are significant differences for pauses, $H(3) = 33.66$ and $p < 0.05$ and post hoc test using Man Whitney shows that there are significant differences between the secure and the insecure category, $U = 1489.5$; $p < 0.01$ and Cohen's $d = 0.6$. There are also significant differences produced by the REQ, $H(3) = 33.67$ and $p < 0.05$ and post hoc test using Man Whitney shows that there are significant differences between the secure and the insecure category, $U = 1368$; $p < 0.01$ and Cohen's $d = 0.8$.

4.3.2 Corpus two: *The present*

Both indicators were compared to expert judgments to analyse which of them is most related (see Table 8).

Both narrative measures, pauses and response eliciting questions, are significantly related to attachment security representation ($\rho = -0.45$; $p < 0.01$ and $\rho = -0.56$; $p < 0.01$). The higher the amount of pauses and questions, the higher the tendency to insecurity.

■ **Table 9** Effect Sizes (Cohen's d) for the Computational Indicators of Children-parent Attachment Relationships.

Indicator	<i>The Stolen Bike</i>	<i>The Present</i>
LSA vector	0.54	0.71
WC	0.42	0.94
WPS	0.84	0.33
negate	0.92	0.83
affect	0.5	0.41
negemo	0.78	0.27
posemo	–	0.44
REQ	0.8	1.25
Pauses	0.6	0.64

Both measures detect significant differences across the studied narrative representations. Kruskal-Wallis test shows that there are significant differences for pauses, $H(3) = 30.78$ and $p < 0.01$ and post hoc test using Man Whitney show that there are significant differences between the secure and the insecure category, $U = 1019.5$; $p < 0.01$ and Cohen's $d = 0.64$. There are also significant differences produced by the REQ, $H(3) = 47.69$ and $p < 0.01$ and post hoc test using Man Whitney shows that there are significant differences between the secure and the insecure category, $U = 775.5$; $p < 0.01$ and Cohen's $d = 1.25$.

4.4 Comparison of the effect sizes obtained in the studied corpora

In conclusion, the LSA semantic vector, Word Count (WC), Words per Sentence (WPS), negation (negate), affective processes (affect), positive emotions (posemo), negative emotions (negemo), pauses and Response Eliciting Questions (REQ) are computational indicators of the level of the children-parent attachments security observed in narratives. These indicators have been compared to two corpora of unfinished children stories produced to elicit different emotions. *The Stolen Bike* was produced to evoke an external threat, whilst *The Present* was created to elicit a potentially positive emotional interaction between parents and children. Effect sizes in security vs. insecurity discrimination can be observed in Table 9.

5 Discussion

The aim of this paper has been to analyse whether the exploration of 3 to 6-year-old children-parent attachment representation through unfinished stories was feasible by computational means. The current study was run with a sample of 184 stories in *The Stolen Bike* corpus and 170 stories in *The Present*, which are a larger sample than most of the previous studies in this specific theme. Both story lines elicit different affective states: an external threat in the case of the *The Stolen Bike* and a positive interaction in *The Present*.

The studied computational frameworks were capable of producing significant associations in relation to expert ASCT judgements.

LSA was capable of capturing the semantics behind secure affective expressions for both corpora. LSA significantly discriminates secure and insecure stories producing a medium effect size in *The Stolen Bike* (Cohen's $d = 0.54$) and in *The Present* (Cohen's $d = 0.71$). Therefore, once ASCT coding is considered, LSA produced consistent medium effect sizes in different corpora.

LIWC was also capable of capturing the linguistic cues which reflect secure and insecure affective expressions. However, there were differences in the two stories. In *The Stolen Bike* the story length (WC), sentence length (WPS), negative expressions (negate), affective expressions (affect) and negative emotions (negemo) were associated with human judgments. But in *The Present* only story length (WC), negative expressions (negate) and positive emotions (posemo) were related to expert judgments. The fact of negative emotion (negemo) being indicative only in *The Stolen Bike* and positive emotion being indicative only in *The Present* is due to the different affective states elicited by each story type. An external threat in the case of *The Stolen Bike* and a positive interaction in *The Present*. Hence, *The Present* involves a positive and quality interaction, while in *The Stolen Bike* negative emotions are maximised.

LIWC was also able to discriminate secure and insecure attachment relationships in both story lines with some indicators. Thus, the story length (WC) shows a small effect size in *The Stolen Bike* (Cohen's $d = 0.54$) and a large effect size in *The Present* (Cohen's $d = 0.71$). Sentence length (WPS) produces a large effect size (Cohen's $d = 0.84$) in *The Stolen Bike* and a small effect size in *The Present* (Cohen's $d = 0.33$). This effect might be due to the fact that in the *The Stolen Bike* story line sentence and cognitive complexity are more relevant in ASCT coding criteria. Negations (negate) produces a large effect size in both *The Stolen Bike* (Cohen's $d = 0.92$) and *The Present* (Cohen's $d = 0.83$). When an insecure relationship is present children have difficulties to answer, presenting avoidance and negation. The affective processes (affect) produce a medium effect size in *The Stolen Bike* (Cohen's $d = 0.5$) and a small effect size in *The Present* (Cohen's $d = 0.41$). The negative emotions (negemo) produce a medium effect size in *The Stolen Bike* (Cohen's $d = 0.78$) and a small effect size in *The Present* (Cohen's $d = 0.27$). The positive emotions (posemo) has a small effect size in *The Present* and no effect in *The Stolen Bike*.

Finally, the additional speech cues included in this study, response eliciting questions (REQ) and pauses, produced consistent medium and large effect sizes in both corpora. REQ produced large effect sizes in *The Stolen Bike* (Cohen's $d = 0.8$) and *The Present* (Cohen's $d = 1.25$). The measure of pauses in the narrative flow produced medium effect sizes in *The Stolen Bike* (Cohen's $d = 0.6$) and *The Present* (Cohen's $d = 0.64$).

In summary, LSA, REQ and pauses produced consistent effect sizes across the corpora. In the case of LIWC, only negative expressions produced consistent large effect sizes across the corpora. Other indicators varied in effect sizes, affected by corpus characteristics. *The Stolen Bike* story being more representative of insecure attachment relationships than *The Present*. Similarly, semantic information was more important to detect secure attachment relationships than insecure attachment relationships. LIWC measures were significantly related mainly to insecure story detection in this specific story, but also provided strong indicatives for security (WC).

Therefore, the study shows that it is possible to explore attachment relations by computational means. Computational modelling reduces time and eases classification in unfinished story classification. However, an in-depth study is required to further explore and expand the possibilities of this approach on a wider dimension exploring the predictive capabilities of the different indicators.

Future lines include the extension of the current corpora adding other incomplete stories included in [22] for a more in-depth analysis of attachment relationships. The corpus-based computational narrative analysis would also allow to further study theoretical questions in attachment such as representational differences for mothers and fathers.

The ability to detect children affective states computationally based on story transcriptions

opens the possibility of computationally detecting affection based on narratives in a wide variety of contexts. Applying computational means to the study of cognitive affective phenomena has already shown to be very promising and full of ongoing challenges [21].

References

- 1 M. D. S. Ainsworth. Some consideration regarding theory and assessment relevant to attachments beyond infancy. In M. T. Greenberg, D. Cicchetti, and E. M. Cummings, editors, *Attachment in the preschool years: Theory, research and intervention*, pages 463–488. University of Chicago Press, Chicago, IL, 1973.
- 2 J. Bowlby. *Attachment and Loss. Vol. I Attachment*. Hogarth Press, London, 1969.
- 3 I. Bretherton. In pursuit of the internal working model construct and its relevance to attachment relationship. In K.E. Grossmann and E. Waters, editors, *Attachment from infancy to adulthood. The major longitudinal Studies*, pages 13–47. Guilford Press, London, 2005.
- 4 J. Cassidy. Emotion regulation: Influences on attachment relationships. *Monographs of the Society for Research in Child Development*, 59:228–249, 1994.
- 5 D. Cicchetti, E. M. Cummings, M. T. Greenberg, and R. Marvin. An organizational perspective on attachment beyond infancy. In M. T. Greenberg, D. Cicchetti, and E. M. Cummings, editors, *Attachment in the preschool years: Theory, research and intervention*, pages 3–49. University of Chicago Press, Chicago, IL, 1990.
- 6 S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by Latent Semantic Analysis. *Journal of the American Society of Information Science*, pages 391–407, 1990.
- 7 S T Dumais. Latent Semantic Indexing (LSI): TREC-3 report. In D Harman, editor, *The 3rd Text Retrieval Conference (TREC-3)*, pages 219–230. NIST Special Publication, 1995.
- 8 P. W. Foltz, W. Kintsch, and T. K. Landauer. The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes*, 25:285–307, 1998.
- 9 Thomas L. Griffiths, Mark Steyvers, and Joshua B. Tenenbaum. Topics in semantic representation. *Psychological Review*, 114(2):211–244, 2007.
- 10 Walter Kintsch. Metaphor comprehension: A computational theory. *Psychonomic Bulletin and Review*, 7(2):257–266, 2000.
- 11 Walter Kintsch. Predication. *Cognitive Science*, 25:173–202, 2001.
- 12 T K Landauer. Applications of Latent Semantic Analysis. In *Proceeding of the 24th Annual Meeting of the Cognitive Science Society*, 2002.
- 13 T. K. Landauer and S. T. Dumais. A solution to Plato’s problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240, 1997.
- 14 T. K. Landauer, D. Laham, and P. W. Foltz. Introduction to Latent Semantic Analysis. *Discourse Processes*, 25:259–284, 1998.
- 15 F. López. *Amores y desamores. Procesos de vinculación y desvinculación sexuales y afectivas*. Biblioteca Nueva, Madrid, 2010.
- 16 Max Lowrese. Embodied relations are encoded in language. *Psychonomic Bulletin and Review*, 15(4):838–844, 2008.
- 17 K. Lund and C. Burgess. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods*, 28(2):203–208, 1996.
- 18 M.L. Newman, J.W. Pennebaker, and D.S. Berry J.M. Richards. Lying words: Predicting deception from linguistic styles. *Journal of Language and Social Psychology Bulletin*, 29(5):665–675, 2003.
- 19 D. Oppenheim. The attachment doll-play interview for preschoolers. *International Journal of Behavioral Development*, 20:681–697, 1997.

- 20 James W. Pennebaker, Roger J Booth, and Martha E. Francis. *Linguistic Inquiry and Word Count: LIWC2007*. The University of Texas and The University of Auckland, Austin, Texas (USA) and Auckland, New Zealand, 2007.
- 21 R. W. Picard. Affective computing: Challenges. *International Journal of Human-Computer Studies*, 59(1-2):55–64, 2003.
- 22 Nerea Portu-Zapirain. Attachment relationships with fathers and mothers during early childhood. *Psychology*, 4(3):254–260, 2013.
- 23 Brian Riordan and Michael N. Jones. Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, 3(2):303–345, 2011.
- 24 S. R. Rochester. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, 2(1):51–81, 1973.
- 25 Maite Román. Methods of assessing attachment in infancy and childhood: From observation of behaviours to exploration of mental representations. *Acción Psicológica*, 8(2):27–38, 2011.
- 26 Stephanie Rude, Eva Maria Gortner, and James Pennebaker. Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18(8):1121–1133, 2004.
- 27 Yla R. Tausczik and James Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.
- 28 R. A. Thompson and S. Meyer. The socialization of emotion regulation in the family. In J. Gross, editor, *Handbook of emotion regulation*, pages 249–268. The Guilford Press, New York, 2007.
- 29 K. Verschueren and A. Marcoen. Attachment story completion task classification system. Unpublished manual, Belgium: University of Louvain, 1994.
- 30 K. Verschueren and A. Marcoen. Representation of self and socioemotional competence in kindergartners: Differential and combined effects of attachment to mother and to father. *Child Development*, 70:183–201, 1999.
- 31 P. Wiemer-Hastings and A. Graesser. Select a kibitzer: A computer tool that gives meaningful feedback on student composition. *International Journal of Artificial Intelligence in Education*, 8(2):149–169, 2000.