Communicative Sequences and Survival Analysis

Carl Vogel

Abstract Two new methods of analyzing dialogue interactions are outlined. One method depends on abstract representations of dialogue events as symbols in a formal language. This method invites analysis of the expressivity requirements of dialogue grammar, as well as distribution analysis of dialogue event symbol sequences. The method is presented in relation to a temporal construction of regular languages, one which supports increasingly fine granularity of temporal analysis. The other method proposed is also temporally oriented. It also depends on dialogue events and dialogue states, and proposes to analyze causal relations among dialogue events through survival analysis. These methods are suggested as additions to the extant repertoire of approaches to understanding the structure and temporal flow of natural dialogue. Additional methods of analysis of natural dialogue may contribute to deeper understanding of the phenomena. With deeper undertanding of natural dialogue one may hope to more fully inform the construction of believable artificial systems that are intended to engage in dialogue with a manner close to human interaction in dialogue.

1 Introduction

Artificial agents with socially believable communication strategies are most naturally informed by human communication behaviors. Recent research into patterns of human communication has examined quantification of engagement. Degree of repetition is a useful proxy measure of engagement, and within a number of studies, comparison of levels of self-repetition and allo-repetition between turns in actual dialogue and randomized counterparts of dialogue provides a means to state when repetition in dialogue exceeds that which one might expect in random base-

Carl Vogel
Centre for Computing and Language Studies, School of Computer Science and Statistics, O'Reilly Institute, Trinity College Dublin, The University of Dublin, Dublin 2, Ireland, e-mail: vogel@tcd.ie

lines [15, 20, 19]. Such repetition effects have been shown to correlate with task-oriented success as proxy measures of communication success in dialogue [17, 16]. This article examines idealizations alternative specifications of baselines for believable social interaction. Two families of approaches are considered: firstly, sequence analysis, drawing on formal language theoretic idealizations of dialogue; secondly, survival analysis. Both of these directions embrace the integrative approach to believable artificial communication strategies recently argued for in the literature [7].

2 Dialogue Symbol Sequences

It is possible to analyze dialogue interactions via idealizations of the sort of content that is expressed at any particular moment. This may be at the rich level of description afforded by conversational analysis [18] or at more abstract levels [14, 2]. For example, the temporal flow of patterns of silence and non-silence are useful in explicating the structure of discussions and their social dynamics [3]. Consider the regular grammar in (1).[1] The terminal symbol s^i represents a silence of speaker i,[2] and the terminal symbol v^i represents a vocalization of speaker i. The set of productions is given in (2)–(8). The superscripts n and m on the nonterminal and terminal symbols vary over the set $\{1,2\}$, for two speakers; thus the production rules with these variable superscripts are schema that abbreviate the enumeration of all of the (finite number of) possible instantiations of those variables.

$$\langle S, \{P^1, P^2\}, \{s^1, s^2, v^1, v^2\}, \{2,3,4,5,6,7,8\}\rangle \tag{1}$$

$$P^n \rightarrow v^n \tag{2}$$

$$P^n \rightarrow s^n \tag{3}$$

$$P^n \rightarrow v^n P^m \tag{4}$$

$$P^n \rightarrow s^n P^m \tag{5}$$

$$S \rightarrow P^1 \tag{6}$$

$$S \rightarrow P^2 \tag{7}$$

$$P^n \rightarrow S \tag{8}$$

The language generated is as in (9).

$$\{w | w \in (s^1 + s^2 + v^1 + v^2)^+\} \tag{9}$$

[1] In this structure, the first element, S, is a special non-terminal symbol, the start symbol; the next element the remaining non-terminal symbols; the third element is a set of terminal symbols; the final element is a set of productions, here pointers to the productions. Each set in this structure is finite. The language generated is infinite.

[2] The representation s^i is adopted as a positive representation of silence, thinking of this as distinct from the usual representation of the empty string (ε). Shortly, representation of distinct sorts of silence will be introduced.

Notice that this is exactly Σ^+. This is trivial in the sense that any non-empty sequence of terminal symbols in the alphabet of the grammar is licensed as well-formed. However, this is a reasonable model at this level of granularity of description of dialogue contributions — the extent to which the assertion that any sequence of vocalizations and silences is well formed is wrong is the extent to which one pays attention to subcategories of vocalization and silence [18].

Nonetheless, this structure is not satisfying. Although it does capture contributions of the two speakers (and generalizes with additional symbols to any number of speakers), it does not capture the simultaneity of contributions of each. It is adequate to think of discrete moments of time (generalizing from finite state approaches to temporality and the semantics of aspect [8]). Define a grammar as in (10) for the sequence of vocalizations and silences of each dialogue participant, *i*.

$$\langle S, \{P^i\}, \{v^i, s^i\}, \{11, 12, 13, 14, 15\}\rangle \tag{10}$$
$$P^i \rightarrow v^i \tag{11}$$
$$P^i \rightarrow s^i \tag{12}$$
$$P^i \rightarrow v^i P^i \tag{13}$$
$$P^i \rightarrow s^i P^i \tag{14}$$
$$S \rightarrow P^i \tag{15}$$

The language generated is given in (16).

$$L^i = \{w | w \in (s^i + v^i)^+\} \tag{16}$$

Given L^m and L^n, instances of (10), define the *superposition* (&) of L^m and L^n ($L^m \& L^n$) as in (17) [8].

$$L^m \& L^n = \bigcup_{k \geq 1} \{(\alpha_1^m \cup \alpha_1^n) \ldots (\alpha_k^m \cup \alpha_k^n) | \alpha_1^m \ldots \alpha_k^m \in L^m, \alpha_1^n \ldots \alpha_k^n \in L^n\} \tag{17}$$

This depends on a notion of component sequences, for example, as in (18), where each component is a set of symbols. Components may stand as string symbols in their own right. The grammar that generates this can be obtained from the grammar (10) by embedding each terminal symbol in the production rules inside a set.

$$L^i = \{w | w \in (\{s^i\} + \{v^i\})^+\} \tag{18}$$

In turn, this permits a definition of componentwise union ($\cup$) of strings of equal length (k), where Σ_i is the terminal vocabulary of L^i, and $\aleph^i \subseteq \{\beta | \beta \in \Sigma_i\}$, and $\sigma^i = \langle \aleph_1^i, \ldots, \aleph_k^i \rangle$ as in (19).

$$\sigma^m \cup \sigma^n = \langle \aleph_1^m \cup \aleph_1^n, \ldots, \aleph_k^m \cup \aleph_k^n \rangle \tag{19}$$

Each component, a composite symbol, of a string σ represents a moment of time, and the componentwise union of strings corresponds to the contribution of each participant at the same moment in time. Thus, given the sequence of contributions (20) and (21), the componentwise union involves two moments in which the speakers alternate silence and speaking, followed by one moment of overlapping vocalization (22).

$$\sigma^1 = \langle \{v^1\}, \{s^1\}, \{v^1\} \rangle \in L^1 \tag{20}$$
$$\sigma^2 = \langle \{s^2\}, \{v^2\}, \{v^2\} \rangle \in L^2 \tag{21}$$
$$\sigma^1 \cup \sigma^2 = \langle \{v^1, s^2\}, \{s^1, v^2\}, \{v^1, v^2\} \rangle \tag{22}$$

On this model, all symbols within a component are understood as simultaneous; each component of a string occupies the same duration as each other unit in the string. At this point, it is useful to note that there are at least two ways in which this representation may be understood. One is that the symbol v^i stands for the vocalization *qua* vocalization of the utterance (linguistic or otherwise) that it names, wherein its duration is the composite duration of the duration of each constituent vocalization, and each constituent vocalization is a distinct proper part of v^i. Another interpretation is that v^i is the proposition that a total vocalization of some sort is happening.[3] On this interpretation, because the proposition is stative, each proper sub-interval of the duration of v^i is also an interval in which the total vocalization is happening, and v^i holds true, even if the total vocalization is incomplete over the sub-interval. This property does not hold as cleanly for the first interpretation, as at a proper sub-interval of the duration of v^i, v^i itself cannot be correctly said to have been vocalized; rather, a proper part of v^i is vocalized over the sub-interval. That is, a vocalization of "Sam has baked a loaf of bread" occupies some total duration, t. However, a proper sub-interval of that duration is not that of a vocalization of "Sam has baked a loaf of bread", but perhaps of something like "Sam has". On the other hand, the proposition that "Sam has baked a loaf of bread" is being uttered holds true during both $\Delta = \delta$("Sam has baked a loaf of bread") and $\Delta' = \delta$("Sam has"), where Δ' is a sub-interval of Δ.[4] The same consideration applies to the silence symbol (or additional symbols for dialogue events).

In order to accommodate varying temporal granularity of strings it is necessary to more clearly associate temporal information with strings. This may be done by associating a clock, C, with the grammar, as in (23).

$$\langle S, N, T, P, C \rangle \tag{23}$$

The clock is used to determine a start time and a constant "frame rate" for a grammar, using the motion picture metaphor that has been used in finite state temporality [8]). Using this analogy, a string corresponds to a film clip, each terminal symbol amounting to the observations recorded within a corresponding frame of the clip.

[3] On the first interpretation, v^k may represent an utterance such as, "the cat is on the mat." On the second interpretation, v^k may represent the proposition, "'the cat is on the mat' is being uttered". In the first case, v^k stands for the vocalization as a vocalization, and in the second case it stands for the proposition that the vocalization is under way.

[4] During $\Delta^\dagger = \delta$("Sam has"), $\Delta^\dagger$ a proper sub-interval of $\Delta^\ddagger = \delta$("Sam has water"), the proposition that a vocalization of "Sam has baked a loaf of bread" is happening is not true.

Each string has a start time, which is at whatever offset from the clock's zero, and each component (each set that contains a terminal symbol), has the same duration, since the clock is assumed to set the start time and frame rate. This model makes it possible to devise a new sort of component union, one that depends not on strings having the same number of components, as in (19), but rather on strings having the same starting time and ending time. Two strings with the same start time and end time and which also share the same frame rate will have the same number of components, and their temporal component union will be the same as with (19), except for the additional temporal information.

Consider the three putative strings depicted in (24), and assume that each of σ^1, σ^2 and σ^3 share temporal endpoints. Firstly, notice that the putative string σ^3 is ruled out as a possible string, since it is depicted as having variable frame durations, contrary to the assumptions described above. In contrast, σ^1 and σ^2 are shown as having constant frame rates within them.

$$\begin{array}{l} \sigma^1 = \langle\{\ v^1\ \}\{\ s^1\ \}\{\ v^1\ \}\{\ \ v^1\ \ \}\{\ s^1\ \}\{\ v^1\ \}\rangle \\ \sigma^2 = \langle\{\qquad v^2 \qquad\}\{\qquad v^2 \qquad\}\{\qquad v^2 \qquad\}\rangle \\ \sigma^3 = \langle\{\qquad s^3 \qquad\}\{\ \ v^3\ \ \}\{\ s^3\ \}\{\qquad s^3 \qquad\}\rangle \end{array} \tag{24}$$

Now, the question is what a temporal componentwise union of σ^1 and σ^2 can be. Since the terminal symbols are interpreted as propositions in the way described above, a proposition that holds of a duration also holds during sub-intervals of that duration. On the alternative interpretation of the symbols described above, this construction would not be accurate in the start and stop times of symbols distributed across shorter constituent durations.

$$\sigma^1 \overset{t}{\cup} \sigma^2 = \langle\{v^2,v^1\}\{v^2,s^1\}\{v^2,v^1\}\{v^2,v^1\}\{v^2,s^1\}\{v^2,v^1\}\rangle \tag{25}$$

A specification of temporal compontentwise union is given below (26) for the basic case in which σ^m and σ^n are co-terminus, but σ^n consists of a single component ($\sigma^n = \langle\aleph^n\rangle$). In this case, there is a component for each component of σ^m, and the constituent durations are determined by the components of σ^m.

$$\sigma^m \overset{t}{\cup} \langle\aleph^n\rangle = \langle\aleph_1^m \cup \aleph^n, \ldots, \aleph_k^m \cup \aleph^n\rangle \tag{26}$$

Let δ be a measure of the duration of a symbol, $\alpha \in \Sigma$, or of a component χ in a string, χ^n, of length n. As specified above, the clock applies to all strings of a language (27).

$$L_\Delta = \{w|w \in (\{\})^*, \delta(\{\}) = \Delta\} \tag{27}$$

If $\sigma = \Sigma^n$, then $\delta(\sigma) = n \cdot \delta(\alpha)$. The temporal componentwise union allows strings from languages with distinct frame rates to be joined. The distinction between vocalization and silence may be further articulated (e.g. listening silence, ignoring silence, etc.), but also other dimensions may be addressed. The intuition is that there is a language of this sort for each dimension of communication in which each agent may engage (e.g. lexical, hand-gesture, facial expression, intonation, laughter, coughs, etc.). Each of those dimensions for each agent brings its own grammar [13, 4]. For each of these dimensions, a separate grammar introduces terminal symbols for the sorts of vocalizations or other form of expression sanctioned in that dimension. For example, the grammar of (28) introduces terminal symbols for (largely involuntary) bodily functions that accompany communication: *a* for sneezes, *b* for burps, *c* for coughs.[5] Of course, these same functions may be dissembled as involuntary, in which case they have a deeper communicative function. However, even in the case of involuntary actions of these sorts, there is a frequent and natural tendency to try to suppress these to "least worst" (if not "most appropriate") moments of interaction for them to happen. As such, they convey information about their agent's evaluation of the unfolding interaction.

$$\langle S, \{P^i\}, \{v^i, s^i\}, \{29, 30, 31, 32, 33, 34, 35, \}, C\rangle \tag{28}$$
$$P^i \rightarrow \{a^i\} \tag{29}$$
$$P^i \rightarrow \{b^i\} \tag{30}$$
$$P^i \rightarrow \{c^i\} \tag{31}$$
$$P^i \rightarrow \{a^i\}P^i \tag{32}$$
$$P^i \rightarrow \{b^i\}P^i \tag{33}$$
$$P^i \rightarrow \{c^i\}P^i \tag{34}$$
$$S \rightarrow P^i \tag{35}$$

One of the goals of this mode of analysis is to support reasoning about the nature, including computational complexity, of dialogue interactions. In modeling dialogue interactions with regular grammars,[6] the least expressive computational framework for capturing infinite sequence possibilities is invoked. Where Σ is the overall set of terminal symbols from the superposition of the grammars for each dimension, this does not entail the assertion that the set of interactions is best described by Σ^* or Σ^+, that any sequence of moves at all is well formed (although this may be the case, and is not ruled out, *a priori*. Rather, the potential for encoding that some sequences are possible and others are not (as has been suggested [18]) is available up to the level of expressivity of regular grammars. It has been argued that some fragments of of natural dialogue interactions are inherently more expressive than regular, and

[5] The symbol *a* is meant to be suggestive of the English onomatopoeic expression, "achoo".

[6] I have not proven that the addition of clocks as proposed here does not increase the expressivity of the framework beyond the expressivity of regular languages. The intuition behind the argument that regular grammars with clocks of this sort remain regular is that the "grouping" of symbols within sets, in an initial grammar, only ever includes a single terminal symbol, never a non-terminal symbol; thus, the effect of bracket matching on either side of a recursive use of a non-terminal symbol (such as is the prototype of a context-free grammar rule) does not occur. The information encoded by the clock for a language may be captured in a constructed regular language without a clock by adding a terminal symbol for the starting time, and another symbol for subsequent "ticks" of the clock. Also, recall that regular languages are closed under both intersection and union [12].

beyond context-free [14]. However, it is not fully clear that the extant arguments are not equivalent to the fallacies of putative expressivity that have been pointed out in other areas of natural language [10]. If regular expressivity is adequate, then, given the propositional interpretation of the terminal symbols used here, then the current framework has the same advantage open to finite state temporality in modeling entailment with the language inclusion problem, whether one set of sequences of propositions is a subset of another set of sequences of propositions, [9], since inclusion is decidable for regular languages [12]. Nonetheless, even if the abstract language of dialogue interactions is provably of greater than regular expressivity,[7] it is useful to think about dialogue interactions in these terms. For example, recent work has explored distributions of symbol sequences effectively involving temporal componentwise unions of social and linguistic signals, in order to study the discourse marking effects of social signals [2]: it was found that sequences inclusive of a topic change symbol significantly more frequently also contain overlap symbols than symbols representing long pauses.[8] Languages specified in the way described here enumerate sequences that represent possible dialogues in which all sequences are as likely as every other sequence. A point, however, is that one may conduct distributional analysis of the grammatical sequences [2] in relation to the baseline model of a uniform distribution. It is then possible to articulate more fine-grained relations about the proximity of some symbols to other symbols in the categories of sequences that dominate the actual distributions. This is orthogonal to being able to study the difference between allowable sequences and ill-formed sequences, towards understanding the expressive requirements of dialogue grammar.

3 Survival analysis

The preceding discussion attached starting points and frame rates to strings of components of symbols, each symbol interpreted as the proposition that a dialogue move of some sort or other has happened in the dialogue modelled. It has been noted that it is of interest to examine these sequences in order to learn about relationships they embed – for example, the greater proximity of some categories of symbols to other reference point symbols (such as topic change). In the mode of operation described above the frame rates are taken to be constant throughout a string, but through temporal componentwise union, one may obtain strings with a "faster" frame rate, a finer level of granularity of temporal analysis. Study of dialogue phenomena also benefits from scrutiny without the assumption of fixed frame rates. Analysis drawn from survival analysis may be relevant, and again it is necessary to establish relevant baselines so that actual data may be compared to the baselines.

The assertion that survival analysis may be appropriate to the study of dialogue is drawn from the observation that if an interaction qualifies as a well formed dialogue,

[7] At greater levels of expressivity, context-free and higher, the inclusion problem is not decidable [12].

[8] In the present work, an overlap is equivalent to a component like $\{v^1, v^2\}$.

then one can be as certain of some facts as of death and taxation: utterances will happen; silences will happen; gestures will happen; overlaps might happen; laughter might happen; coughs might happen. Not all events are certain, but some are. However, even the propositions related to uncertain events are temporally bounded – an overlap might happen, but if it does, it will eventually end. This suggests an alternative way of exploring the relationships among dialogue events via timed symbol sequences that represent the containing dialogues. In the simplest version of the idea, one examines the "survival" of symbols of interest in relation to alternative treatments: other symbols that precede "death". That is, topics will happen during a dialogue and will inevitably terminate within the dialogue. The question is whether survival analysis may be used to quantify which other symbols' appearance in immediate or relatively prior positions in the sequences that include topic termination are more or less likely to hasten, correlate with or otherwise predict topic termination.

3.1 General method

To engage in this sort of analysis a coding of data may be conducted using standard statistical tools for analysis of "time to events". Again, actual data and effects may be compared with effects that obtain in data constructed using random generators. Suppose that one is interested in whether the occurrence of laughter correlates with topic change, then it is possible to measure the survival of topics in relation to whether laughter occurs within some interval from the end of a topic. For each topic the duration of the topic is recorded. Using the same starting point measure as for the duration of the topic, the duration of time without laughter may be recorded. For some d either that varies with the topic length (e.g. half the duration of the duration of the topic) or a constant (e.g. 15 seconds), a factor related to laughter being present close enough to the topic termination to be related to it may be created (36).[9] If d is equal to the topic duration, then LM is a binary indication of whether laughter occurs within the topic or not.

$$LM = ((T - \neg L) > 0) \wedge ((T - \neg L) \leq T \cdot d) \tag{36}$$

The terms of (36) are: LM, a binary indication of whether laughter matters for the topic;[10] T, the time from the onset to the termination of the topic; $\neg L$, the time from the onset of the topic without laughter;[11] d, described above. Similar factors may be constructed from other signals that one might wish to correlate with topic

[9] In order to subtract away a constant c from T using (36), it is sufficient to specify d as $\frac{T-c}{T}$; to test whether $T - \neg L$ is less than some constant c, it suffices to specify d as $\frac{c}{T}$.

[10] That is, this is true if laughter is close enough to the change of topic to be reasonably hypothesized as a discourse marker.

[11] Non-laughter, which terminates with the onset of laughter, is constructed as a dialogue event, so that one may examine the survival rate of non-laughter.

change (e.g. coughing, overlap, etc.). One may then apply *LM* and its counterparts for the other possible dialog events in an analysis of the "survival" of topics with and without laughter, for example, close enough to the topic termination (for whatever measure of d that is appropriate) to plausibly be causally related.[12] The interest is in whether there is a difference in the survival of topics between topics with instances of laughter (or other dialogue events) commencing d-close to the topic end and topics without such laughter (or other dialogue events). Equivalently, one may study the survival of laughter conditioned upon factors derived from other dialogue events.

3.2 An example

Recent work in the analysis of laughter as a discourse marker conveying information about topic change explored the likelihood of laughter in the windows of time before and after topic changes. Whether the windows were defined relative to the overall duration of the topic (for example, considering the first quarter of a topic's duration or the topic's final quarter) or defined in absolute temporal terms (e.g. the first 15 seconds or last 15 seconds of a topic), the same effects emerged during the analysis of more than one multi-modal dialogue corpus: instances of laughter were less likely at the start of a new topic than at a topic's end [1, 11].[13] This section develops an example application of the method described above to the TableTalk dataset analyzed within the recent works just mentioned [5].

In this dataset, free conversation among five participants over three sessions are recorded. The conversation setting was informal, within the Advanced Telecommunication Research Labs in Japan. The participants included three women (Australian, Finnish and Japanese) and two men (Belgian and British). Topic annotations (116 distinct topics) and annotations of laughter (719 instances) are provided with the corpus, which spans 3 hours and 30 minutes of chat.

Initially, consider $\neg L$ to be the time within a topic before the onset of its last moment of laughter, and take d to be 0.10, so that $T \cdot d$ (in (36)) refers to 10% of the topic's duration: where *LM* is true, the topic is one in which the last moment of laughter occurs in the final 10% of the topic's duration. The plot in the in the left-hand side of Figure 1 depicts the time (in seconds) to the end of topic for topics where the last laugh occurs in the final 10% of its duration in comparison with the topic durations without laughter so placed: topics for which the laughter oc-

[12] Interestingly, given that dialogue participants may be imputed to all have a sense of when discussion of a topic is coming to an end, if this sensation does give release to laughter, or other signals, that correlate with topic end, then there is a case to be made for temporally inverted causation: the topic end may cause the laughter which precedes it.

[13] Laughter, it turns out, is not a precise signal of topic ending, but if laughter is present, it serves as a very good signal that the topic of discussion has not just started.

curs in the final 10% are significantly longer in duration than the complement.[14] This extends to contrasts involving the existence of final laughter in shorter durations relative to the topic duration as well: the final 5% ($p < 0.005$), final 2.5% ($p < 0.005$), final 1.25% ($p < 0.005$). The direction switches for absolute duration measurements. The plot in the right-hand side of Figure 1 compares topic durations for topics that have final laughter in the last ten seconds with topic durations for the complement. The former are significantly ($p < 0.05$) shorter than the latter. The comparisons related to absolute durations slightly less than 10 seconds and slightly more than 10 seconds exhibit the same trend, but without statistical significance. A fact common to both of the relative and absolute examinations is that ten topics did not contain laughter, and as a group those topics were significantly shorter in duration than topics that contained laughter. However, focusing analysis on just the topics containing laughter does not disrupt the overall pattern of results – only the test of topic durations with final laughter in the final 5% loses significance, but the same direction and approximate strength of effect holds in each other case.

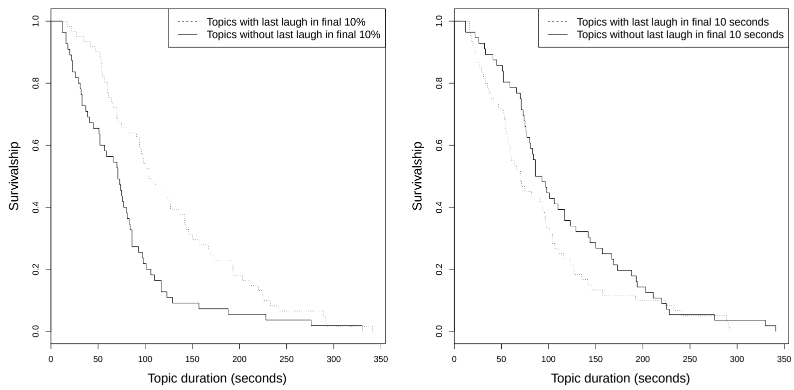
There is a significant ($P < 0.0001$) positive correlation (0.51) between laughter counts and topic duration, but this correlation is weaker, yet still strong (0.48), and still significant ($P < 0.0001$) among only topics containing laughter. Thus, the analysis of laughter in the final d percentages of topic durations may be revealing little more than that topics containing more laughter in the final d percent contain more laughter overall. Examining topics that have laughter in the final d percent ($d < 10$, for the percentages indicated above), or in the final d seconds, there are significantly more instances of laughter overall for the topic than for topics that do not have laughter in those topic-final moments (Wilcox test, $p < 0.005$). Thus, the greater presence of laughter overall during the topics may explain the plot on the left in Figure 1, but not the data plotted on the right, where it is shown that topics with laughter in the final ten seconds are shorter in duration than the complement.

3.3 Discussion

The differences noted above in the time to the end of the topics for topics that have instances of laughter in terminating sequences defined in absolute terms in relation to those with terminating sequences defined relative to the overall topic length is intriguing. For some types of events, their causal impact is most likely to be evident within a temporal window defined in absolute terms (a falling is likely to have been caused by a tripping within a few seconds prior to the falling rather than by a tripping within the final few percent of the walk). For others, such causal relations are more evident in relational definitions (fatal lung cancer is more likely to have been caused by having taken up intensely smoking cigars in the final ten percent of a life than in the final three years of life without reference to the total life span). The re-

[14] Using the one-tailed asymptotic log-rank test $p < 0.005$. See `surv_test` within the `coin` package within R (http://cran.r-project.org/web/packages/coin/index.html – last verified, July 2015) created by Torsten Hothorn, Kurt Hornik, Mark A. van de Wiel and Achim Zeileis.

Fig. 1 Time in seconds to end of topic for topics with and without terminating laughter



sults noted above show that the time to the end of topics is extended by laughter (the plot on the left in Figure 1) but also suggest that the presence of laughter creates a window in which it is also relatively easy for topic to change sooner, hastening time to the end of the topic (the plot on the right in Figure 1).[15] These results are far from conclusive, but are suggestive of additional questions to explore in isolating the categories of laughter that hasten or delay topic termination – for example, evenness versus clumpiness in the distribution of laughter across a topic's duration.[16] Of course, by construction, the grouping of topics here does not correspond to treatments as typically analyzed within survival analyses. In fact, it is a purpose of this work to demonstrate that survival analysis may be fruitfully adapted to the purposes of analyzing patterns within sequences of dialogue moves.

4 Conclusions

That believable autonomous systems must be able to demonstrate naturalistic dialogue interactions follows from the fact that dialogue interactions inform diagnoses of pathological human conditions, such as schizophrenia [6]. The present article

[15] It is important to remember that the topic durations in the two representative plots relate to the same topics, but under distinct classifications.

[16] It is also necessary to apply this sort of analysis to additional dialogue corpora in order to understand the robustness of the effects found for the TableTalk corpus.

is programmatic in defining two approaches to studying dialogue interactions as temporal symbol sequences. One approach is anchored in formal language theory analysis and focuses on the generative framework that can yield appropriate symbol sequences, as well as supporting distributional analysis of the sequence sets without reference to grammars directly. The other approach handles temporal information at a more fine-grain level of detail and considers causal relations among propositions related to dialogue events using survival analysis, given temporal distributions of dialogue symbols. The motivation for proposing these paradigms of dialogue analysis is to pursue them as a way of gaining greater insight into the structure and temporal flow of dialogue events. If the nuances of dialogue structure and temporal flow can be better understood for human dialogues from a wide range of genres then there is increased possibility for incorporating this knowledge in automated systems that are intended to have believable interaction possibilities.

Acknowledgements I am grateful to Francesca Bonin, Nick Campbell, Anna Esposito and Emer Gilmartin. They have no responsibility for any wrong-headedness that appears here, though. This research is enhanced by supported from Science Foundation Ireland through the CNGL Programme (Grant 12/CE/I2267) in the ADAPT Centre (www.adaptcentre.ie) at Trinity College Dublin. The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

[1] Bonin F, Campbell N, Vogel C (2014) Time for laughter. Knowledge Based Systems 71:15–24

[2] Bonin F, Vogel C, Campbell N (2014) Social sequence analysis: Temporal sequences in interactional conversations. In: Baranyi ea Péter (ed) 5th IEEE International Conference onf Cognitive Infocommunications, pp 403–406

[3] Bouamrane MM, Luz S (2007) An analytical evaluation of search by content and interaction patterns on multimodal meeting records. Multimedia Systems 13:89–102, DOI10.1007/s00530-007-0087-8

[4] Bunt H, Alexandersson J, Chae JW, Fang AC, Hasida K, Petukhova O, Popescu-Belis A, Traum D (2012) ISO 24617-2: A semantically-based standard for dialogue act annotation. In: Proceedings LREC 2012

[5] Campbell N (2009) An audio-visual approach to measuring discourse synchrony in multimodal conversation data. In: Proceedings of Interspeech 2009

[6] Covington MA, He C, Brown C, Naçi L, McClain JT, Fjordbak BS, Semple J, Brown J (2005) Schizophrenia and the structure of language: The linguists view. Schizophrenia Research 77:85–98

[7] Esposito A, Esposito AM, Vogel C (2015) Needs and challenges in human computer interaction for processing social emotional information. Pattern Recognition Letters To appear. http://dx.doi.org/10.1016/j.patrec.2015.02.013

[8] Fernando T (2004) A finite-state approach to events in natural language semantics. Journal of Logic and Computation 14(1):79–92, DOI 10.1093/logcom/14.1.79

[9] Fernando T (2007) Observing events and situations in time. Linguistics and Philosophy 30(5):527–550, DOI 10.1007/s10988-008-9026-1

[10] Gazdar G, Pullum G (1985) Computationally relevant properties of natural languages and their grammars. Tech. Rep. CSLI-85-24, Stanford: Center for the Study of Language and Information

[11] Gilmartin E, Bonin F, Vogel C, Campbell N (2013) Laugher and topic transition in multiparty conversation. In: Fernández R, Isard A (eds) Proceedings of the SIGDIAL 2013 Conference, Association for Computational Linguistics, Metz, France, pp 304–308, URL http://www.aclweb.org/anthology/W/W13/W13-4045

[12] Hopcroft JE, Ullman JD (1979) Introduction to Automata Theory, Languages, and Computation. Addison-Wesley Publishing Co., Reading MA.

[13] Petukhova V, Bunt H (2009) Towards a multidimensional semantics of discourse markers in spoken dialogue. In: Proceedings of the Eighth International Conference on Computational Semantics, Association for Computational Linguistics, Stroudsburg, PA, USA, IWCS-8 09, pp 157–168

[14] Păun G (1984) Modelling the dialogue by means of formal language theory. In: Vaina L, Hintikka J (eds) Cognitive Constraints on Communication: Representations and Processes, D. Reidel, Dordrecht, Holland, pp 363–372

[15] Ramseyer F, Tschacher W (2010) Nonverbal synchrony or random coincidence? How to tell the difference. In: Esposito A, Campbell N, Vogel C, Hussain A, Nijholt A (eds) Development of Multimodal Interfaces: Active Listening and Synchrony, Springer, LNCS 5967, pp 182–196

[16] Reitter D, Moore J (2007) Predicting success in dialogue. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Association for Computational Linguistics, pp 808–815

[17] Reitter D, Keller F, Moore J (2006) Computational modeling of structural priming in dialogue. In: Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL, Association for Computational Linguistics, p 121124

[18] Schegloff EA (2007) Sequence Organization in Interaction: A Primer in Conversational Analysis, vol 1. Cambridge University Press

[19] Vogel C (2013) Attribution of mutual understanding. Journal of Law & Policy 21(2):377–420

[20] Vogel C, Behan L (2012) Measuring synchrony in dialog transcripts. In: Esposito A, Esposito AM, Vinciarelli A, Hoffmann R, Müller VC (eds) Cognitive Behavioural Systems, Springer, LNCS 7403, pp 73–88