TOWARDS A SIMULATION MODEL OF DIALOGICAL ALIGNMENT

ALEXANDER MEHLER, PETRA WEIß, PETER MENKE, ANDY LÜCKING
Text Technology, Bielefeld University, Universitätsstraße 25,
Bielefeld, D-33615, Germany
`Alexander.Mehler@uni-bielefeld.de`

This paper presents a model of lexical alignment in communication. The aim is to provide a reference model for simulating dialogs in naming game-related simulations of language evolution. We introduce a network model of alignment to shed light on the law-like dynamics of dialogs in contrast to their random counterpart. That way, the paper provides evidence on alignment to be used as reference data in building simulation models of dyadic conversations.

1. Introduction

Simulation models of language evolution mainly start from a simplified notion of dialog. In simulation rounds of such models, a sender is typically selected at random to generate a signal for the listener *without considering any turn taking among the agents*. We call this scenario a *single-turn scenario* and contrast it with *multi-turn scenario*s where turn taking takes place in simulation rounds. By turn taking we refer to the fact that interlocutors continue to change their roles as sender and listener (Sacks, Schegloff, & Jefferson, 1974). Simulation models mostly disregard this dynamics that non-randomly structures communication.

At first sight, turn taking has been modeled in several simulations. Padilha and Carletta (2002), e.g., simulate turn taking with respect to multimodality. However, these simulations focus on discourse-managing strategies. Current research on dialog systems also makes use of simulations to improve strategies of the artificial systems (Scheffler & Young, 2002). Further, there are simulations where interaction among agents is only possible in subsequent rounds, which renders modeling of round-internal turn taking impossible. Steels and Loetzsch (2009), e.g., describe an experiment with robots with the purpose of perspective alignment where agents align to the speaker's usage strategy for ambiguous spatial expressions.

So far, little is known about the impact of *dialogical* communication on the outcome of language evolution. To study this impact, a simulation model is needed that *embeds* a simulation model of multi-turn communication. A central aspect of the dynamics of dialogs is *alignment* (Pickering & Garrod, 2004), which is a largely automatic, resource-saving process of structural coupling among interlocutors on several linguistic, e.g., lexical (Clark & Wilkes-Gibbs, 1986), syntactic

(Branigan, Pickering, & Cleland, 2000) and semantic (Garrod & Anderson, 1987; Garrod et al., 2007) levels that simplifies communication. The basic mechanism underlying alignment is *priming*: linguistic representations are primed by utterances so that (features of) expressions get copied within an agent dyad. Priming operates *intra-* and *inter*personally, that is, on two channels: a horizontal, "monological" one within a speaker and a vertical, dialogical one between speakers. The latter is based on a reciprocal dyadic exchange, which involves role switching, that is, turn taking.

This paper presents a model of *lexical* alignment in the framework of naming games (Jäger, 2006). It approaches reference data for simulation models of a certain class of real dialogs to be embedded into simulation models of language evolution (Kirby & Hurford, 2002). We introduce a network model according to a multi-turn scenario to separate the law-like dynamics of dialog from its random counterpart: Section 2 describes the experiment that we performed to get empirical data; Section 3 and 4 use this data to build a model of dialog lexica that allows for automatically classifying dialogs according to their (non-)alignment.

2. The Naming Game in an Experimental Perspective

Language use in free dialogs is hardly controllable so that experimental paradigms have been developed to elicit semi-spontaneous dialog situations where some degree of control over the topic of conversation is possible: the *referential communication task*, the *maze game* and the *map task* (Krauss & Weinheimer, 1966; Garrod & Anderson, 1987; Anderson et al., 1991). Based thereon, we developed the *Jigsaw Map Game* (JMG; Weiß et al., 2008) that allows for naturalizing experimental dialogs by encouraging face-to-face interaction with participants who mutually perceive and communicate in a multimodal way. In the JMG, parameters like dialog organization are controlled by regulating the game's flow and balancing partner roles. The JMG goes as follows: two participants cooperatively position objects (e.g., cuboids or cones) on a common interaction table according to a predefined arrangement. The arrangement is designed in a way that some objects stand out because of size. They define so called critical objects with two possible names (e.g., *ball* or *bowl*). The cooperative character of the game emerges by the fact that each partner only gets partial information about the final arrangement. This is realized by instruction cards that contain the constellation of three objects at a time: two already positioned and one new object to be placed by the partner in the next step. Guided by these cards, partners communicate in turns which object the other should pick next from a personal object box (*object identification*) and where it has to be placed on the table (*object placement*) until the whole arrangement is completed. This way, it is possible to analyze the names the interlocutors use to identify the objects. In the following sections, we take all our empirical data from JMG-based experiments. More specifically, we recorded and annotated 19 dialogs manifesting this experimentally controlled naming game.

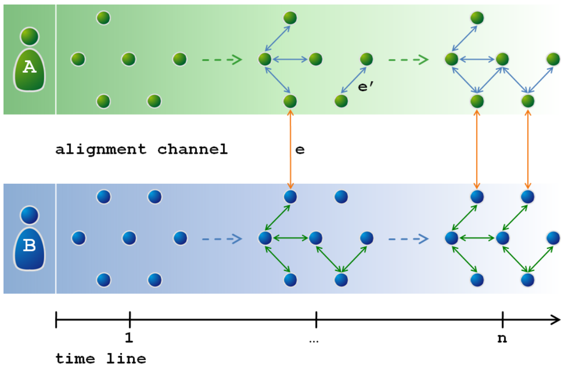

Figure 1. Schematic representation of a bipartite time-aligned network series.

3. A Network Model of Alignment in Communication

Interpersonal alignment is brought forward by the turn taking of interlocutors. Simulation models mainly disregard this process, although it affects the build-up of dialog lexica and the transfer of linguistic representations. We introduce a graph model of such lexica that does not simply count items shared by interacting agents but captures (the strength of) lexical associations and the time course of alignment. This is done in terms of *bipartite Time-Aligned Network (TAN) series* as illustrated in Fig. 1. It presents the gradual build-up of a dialog lexicon by two interlocutors A and B: at the beginning (time point 1), both agents start with a set of unlinked lexical items. For an agent, this set presents the subset of items of his overall lexicon that are actually activated during his conversation. Thus, vertices in Fig. 1 denote lexical items. Henceforth, we represent dialog lexica as labeled graphs $L_t = (V, E_t, \mu_t, \mathcal{L})$ weighted by μ_t and indexed by points in time $t = 1..n$ at which they are built. We assume that vertices are labeled by the surjective function $\mathcal{L} : V \to L_V$ for the set of labels L_V (i.e. lemmata). Each time an interlocutor produces a linguistic output, the series proceeds to the next time point. We assume that L_t is divided into two subgraphs $L_{t_A} = (V_A, E_{t_A}, \mu_{t_A}, \mathcal{L}_A)$ and $L_{t_B} = (V_B, E_{t_B}, \mu_{t_B}, \mathcal{L}_B)$ according to the distribution of L_t over A and B at time t. In Fig. 1 this corresponds to a column of the TAN series. Note that $\mathcal{L}_X : V_X \to L_{V_X}, X \in \{A, B\}$, is the bijective restriction of $\mathcal{L}$ to V_X, while μ_{t_X} is the restriction of μ_t to E_{t_X}.

By continuing their communication, agents gradually span edges in the lexicon. This is done by *inter-* or *intra*personal links. To see that remember that in the JMG, a dialog is divided into rounds each of which corresponds to a focal object or topic (e.g., *ball* or *cone*). Thus, for each time point t of a JMG we can identify the focal topic $x = \mathit{focus}(t)$ at t. Now we induce links in the lexicon as follows:

- If at time t agent $X, X \in \{A, B\}$, uses the item l to express $\mathit{focus}(t)$ that

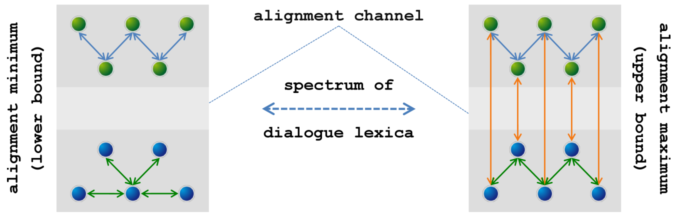

Figure 2. A graph-theoretical representation of turning points of lexical alignment.

has been expressed by $Y \neq X$ in the same round by l, we span an *interpersonal* link between $v_A \in V_A$ and $w_B \in V_B$ where $\mathcal{L}_A(v_A) = \mathcal{L}_B(w_B) = l$, e.g., edge e in Fig. 1. Links of this sort cross the *alignment channel* between A and B. If the link already exists, its weight $\mu_t(e)$ is increased by 1.

- If at time t agent $X \in \{A, B\}$ uses item l to express $focus(t)$, we generate an *intrapersonal* link between $v_X \in V_X$, $\mathcal{L}_X(v_X) = l$, and all other vertices labeled by items that X has used in the same round of the game, i.e., items that are equally used as l (see, e.g., edge e' in Fig. 1). Once more, if the links already exist, their weights (initially set to 1) are augmented by 1.

Note that without turn taking there would be no interpersonal links in this model. The graphs $L_1, \ldots, L_n$ define a time series where each time point corresponds to a network. Since each of them is decomposable into subgraphs L_{t_A} and L_{t_B}, we speak of a bipartition. This model easily allows for representing turning points of alignment as shown in Fig. 2: on the left side, a dialog lexicon occurs without any link across the alignment channel. This happens, e.g., if the interlocutors never use the same item to speak about the same topic. That is, *the interlocutors use different words or the same words differently so that their dialog lexicon is nonaligned.* The right part of Fig. 2 shows the opposite case. Such a situation occurs if *both interlocutors use the same words the same way so that their dialog lexicon is completely aligned.* Obviously, lexica emerging from real dialogs enter into the middle of these extremal points. We assume that this happens law-like so that we can apply complex network theory (Barrat, Barthélemy, & Vespignani, 2008) to analyze the networking of dialog lexica as described in the next section.

3.1. *Quantifying Dialog Lexica*

Bipartite TAN series allow for analyzing intra- and interpersonal links of lexical items, that is, lexicon structure. This is done by means of the following indices:

- The average geodesic distance $\hat{L}(G) = 1/\binom{|V|}{2} \sum_{\{v,w\} \in [V]^2} \delta(v, w)$ where $[V]^2$ is the set of all subsets of 2 elements of V and $\delta(v, w)$ is the geodesic distance of v and w in G. Since at the beginning of lexicon formation vertices tend to be unlinked, this index gives unrealistically small values if $\hat{L}$ is restricted to the largest connected component. Thus, we normalize it to capture all pairs of vertices: $L(G) = 1 - ((\sum_{\{v,w\} \in [V]^2} \delta(v,w))/(n^2(n-1))$, where $\delta(v,w) = n-1, n = |V|$, if v and w are disconnected. L ranges from 0 (completely disconnected graph) to 1 (completely connected graph).

- The cluster coefficient C_{ws} (Watts & Strogatz, 1998) that together with L describes small worlds. Additionally, we compute its weighted extension C_w (Barrat et al., 2008) and its counterpart C_{br} (Bollobás & Riordan, 2003).
- The weighted cluster coefficients $\langle C_w(k) \rangle$ and $\langle C_w^{ns}(k) \rangle$ (Serrano, Boguñá, & Pastor-Satorras, 2006), which, as C_w, evaluate edge weights. Basically, we interpret the weights of links as indicators of the strength of association among the interlinked items: the higher the weight, the stronger their usage-based relation, the higher the association of the one if the other is primed.
- A measure of cohesion $coh(G) = \frac{\sum_{v \in V} d_G(v)}{|V|^2 - |V|}$ that scores the cohesion of G as the ratio of the number of its links in relation to the maximal number of possible links; $d_G(v)$ is the degree of v. We also compute the compactness index $cp(G)$ and $cp_{lcc}(G)$ of hypertext theory (Mehler, 2008).
- An index of modularity that for a given network measures the independence of its candidate modules. As we consider networks with two modules, we use the following variant of the index of Newman and Girvan (2004): $Q(G) = \sum_{i=1}^{2}(e_{ii} - a_i^2)$ where e_{ii} is the number of links within the ith part of the network and a_i^2 is the number of links across the alignment channel.

We hypothesize that aligned dialogs are distinguished from nonaligned ones by topological indices of the TAN series that represent them as tested below.

Our next step is to distinguish our model of the gradual build-up of dialog lexica from a random counterpart. This is done by algorithm that for a given TAN series computes a randomized series of equal order and size. It randomly rewires the last link of the series uniformly at random and then randomly deletes edges till the first link is reached. Then we compute a set of topological indices for each network state of each randomized series and, finally, average those indices over this set. In this way, we get for each topological index of Section 3.1 an expected value under the condition that the networking of the lexicon is at random.

4. Experimentation

In this section, we evaluate topological indices of dialog lexica regarding their relevance for alignment measuring. We start with two dialogs for which we know by human expert annotation that they manifest alignment and nonalignment, respectively. In Fig. 3 we present the value patterns of C_{ws} and $\langle C_w(k) \rangle$ based on three dialog lexicon graphs: the graph of the lexically nonaligned dialog ($-$), the graph of its aligned counterpart ($+$) and of its randomized counterpart averaged over 50 repetitions (rnd). As we focus on naming games, we account for nomi-

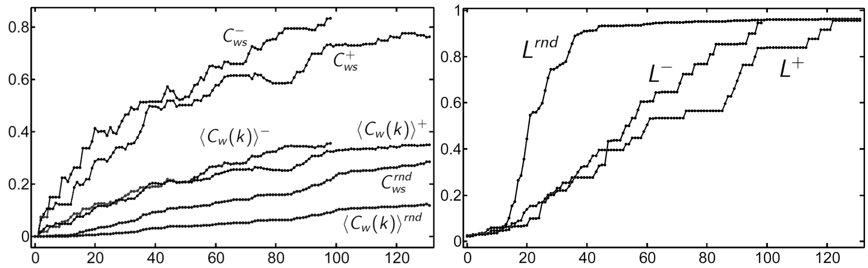

Figure 3. X-axis: time line. Y-axis: cluster values (left) and average geodesic distances (right) of a dialog manifesting nonalignment ($-$), alignment ($+$) and the random model (rnd).

nal units so that each time point in Fig. 3 corresponds to the utterance of a noun (according to the sequence of turns of the interlocutors) and the related update of the dialog lexicon graph. Fig. 3 is in line with our expectations: the randomized dialog lexica have relatively small cluster values. We also observe that the non-aligned dialog is distinguished from its aligned counterpart in terms of weighted and unweighted clustering. These relations are retained irrespective of the cluster coefficient in use. They are also retained if we evaluate the value pattern of Q: in case of nonalignment, modularity is more salient than in case of alignment, while the randomized dialog graphs show an unnaturally small modularity. Regarding geodesic distances L, the picture is somehow different (see Fig. 3): in line with network theory, a small L rapidly emerges in randomized dialog lexica. It also emerges more quickly in the nonaligned dialog compared to its aligned counterpart. In all three cases a small L of about two links characteristically emerges at the end of the dialog – that is, *dialog lexica are tiny small worlds*.

Supposed that topological indices of dialog lexica are really indicative of lexical alignment, they should allow for automatically classifying dialogs with a high F-score (i.e., harmonic mean of precision and recall). This is what we test now. We explore a corpus of 19 manually annotated JMG dialogs. 15 of them have been rated to manifest alignment, 4 have been judged to manifest nonalignment. The classifier should automatically reconstruct this human-based gold standard where a high F-score (up to 1) indicates a good classification, while a small F-score (down to 0) indicates a failure. For each of the lexicon graphs of the 19 dialogs (with an average duration of 11 minutes) we compute the set of indices described in Sec. 3.1. Thus, each dialog is represented by a vector of 10 indices (see Table 1). This allows for applying cluster analysis and to compute F-scores. The results are reported in Table 1. It shows that both baseline scenarios are outperformed: scenario I that assumes an equi-partition among the target classes and scenario II that knows their real sizes. We also see that 5 indices are enough to classify up to 95% of the dialogs correctly – *this shows that classifying dialogs by means of network analysis is feasible*. In Table 1, we also report feature combinations

Table 1. F-scores of dialog classification (above) and sensitivity analysis of the indices (below).

procedure	**F-score**	**remark**
`QNA[Mahalanobis, hierarchical, complete linkage]`	.945	5 out of 10 features
`QNA[Mahalanobis, hierarchical, average linkage]`	.746	all features
`random baseline II`	.703	known partition
`random baseline I`	.634	equi-partition

C_{br}	C_{ws}	L	cp	cp_{lcc}	coh	$\langle C_w(k) \rangle$	$\langle C_w^{ns}(k) \rangle$	C_w	Q	#
$\times$	$\times$		$\times$			$\times$			$\times$	5
$\times$	$\times$	$\times$	$\times$	$\times$	$\times$					6
$\times$		$\times$			$\times$		$\times$	$\times$	$\times$	6
$\times$	$\times$	$\times$			$\times$	$\times$	$\times$	$\times$	$\times$	8
4	3	3	2	1	3	2	2	2	3	#

that provide the same F-score. They have been computed by means of a genetic algorithm for selecting feature subsets. Table 1 shows the features that are most reliably used throughout the different runs. Obviously, the cluster coefficient C_{br} performs best among all candidates considered here. We also see that weighted cluster coefficients are less indicative of alignment. Further, we get the information that the average geodesic distance L, the index of modularity Q, the cluster coefficient C_{ws} and the cohesion measure coh are reliable indicators of alignment vs. nonalignment. Thus, we conclude that the value patterns of these indices can serve as reference frames for building simulation models of the time course of dialogical alignment. By considering their value patterns a simulation model is informed about what naturally happens in dialogs of the sort of the naming game considered here.

5. Conclusion

We presented a network model of lexical alignment in a multi-turn scenario. It allows for distinguishing aligned from nonaligned dialogs. That way, our model provides reference data for simulation models of dialogical communication.

Acknowledgement

Financial support of the German Federal Ministry of Education through the project *Linguistic Networks*, of the German Research Foundation through the EC *Cognitive Interaction Technology* and the SFB *Alignment in Communication* is gratefully acknowledged.

References

Anderson, A.H., Bader, M., Gurman Bard, E., Boyle, E., Doherty, G., & Garrod, S. et al. (1991). The HCRC Map Task Corpus. *Language and Speech*, *34*, 351-366.

Barrat, A., Barthélemy, M., & Vespignani, A. (2008). *Dynamical processes on complex networks.* Cambridge: Cambridge University Press.

Bollobás, B., & Riordan, O. M. (2003). Mathematical results on scale-free random graphs. In *Handbook of graphs and networks.* Weinheim: Wiley-VCH.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, *25*, B13–B25.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representations: Where might graphical symbol systems come from? *Cognitive Science*, *31*, 961–987.

Jäger, G. (2006). Convex meanings and evolutionary stability. In A. Cangelosi, A. D. M. Smith, & K. Smith (Eds.), *Proc. of EVOLANG6* (p. 139-144).

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). London: Springer.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*, 343-346.

Mehler, A. (2008). Structural similarities of complex networks. *AAI*, *22*(7&8), 619–683.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113.

Padilha, E., & Carletta, J. (2002). A simulation of small group discussion. In *Proc. of EDILOG* (pp. 117–124).

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-226.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language*, *50*, 696-735.

Scheffler, K., & Young, S. (2002). Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning. In *Proc. of the 2nd int. conf. on Human Language Technology Research* (pp. 12–19).

Serrano, M. Á., Boguñá, M., & Pastor-Satorras, R. (2006). Correlations in weighted networks. *Physical Review E*, *74*, 055101.

Steels, L., & Loetzsch, M. (2009). Perspective alignment in spatial language. In K. R. Coventry, T. Tenbrink, & J. Bateman (Eds.), *Spatial language in dialogue.* Oxford.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440-442.

Weiß, P., Pfeiffer, T., Schaffranietz, G., & Rickheit, G. (2008). Coordination in dialog: Alignment of object naming in the Jigsaw Map Game. In *Proc. of cognitive science 2007.*