From Neural Activation to Symbolic Alignment: A Network-Based Approach to the Formation of Dialogue Lexica

Alexander Mehler Andy Lücking Peter Menke

***Abstract*— We present a lexical network model, called TiTAN, that captures the formation and the structure of natural language dialogue lexica. The model creates a bridge between neural connectionist networks and symbolic architectures: On the one hand, TiTAN is driven by the neural motor of lexical alignment, namely priming. On the other hand, TiTAN accounts for observed symbolic output of interlocutors, namely uttered words. The TiTAN series update is driven by the dialogue inherent dynamics of turns and incorporates a measure of the structural similarity of graphs. This allows to apply and evaluate the model: TiTAN is tested classifying 55 experimental dialogue data according to their alignment status. The trade-off between precision and recall of the classification results in an F-score of 0.92.**

I. Introduction

PARTICIPATING IN DIALOGICAL COMMUNICATION, the dialogue partners are observed to *align* with each other in their linguistic behaviors – see for instance the findings reported on the phonetic [1], on the lexical [2], and on the syntactic [3], [4] levels of dialogue partners' utterances. This mutual alignment propagates downwards to "deeper" linguistic representations, namely to the level of semantics [5] and situation models [6]. The alignment of situation models facilitates the ease of mutual understanding in dialogue and has been developed by [7] as a mechanistic theory of dialogue, the *Interactive Alignment Model* (IAM). The other, upwards, way round, alignment on the "deeper" linguistic levels leads by the same means to a stronger alignment on the other ones.

On the level of directly observable linguistic behavior, for instance, alignment can be found in the dialogue partners' convergence on commonly used words. In the following dialogue extract, the interlocutors A and B, without discussion, reach an agreement to use the form *lamps* to refer to some light posts in the described scene:[1]

A:	street lights		lamps
B:		street lamps	

A's utterance (*street lights*) is partially corrected by B. The corrected part is subsequently taken up by A. The form *lamps* is then commonly used to refer to the plural referent. As a consequence, the dialogue lexica of the dyad from the example contain three words associated with one and the same collective entity, namely *street lights*, *street lamps*, and *lamps*. It is the latter, shortest form, *lamps*, that becomes routinized, that is, a repeated dialogue lexicon entry used unequivocally to refer to the dialogue-specific topic, in this case, the arrangement of lights in the described scene.

There is disagreement about what psychological mechanisms implement alignment. On the one hand, [7] argue for priming and thereby advocate an automatic process of which the dialogue partners are not consciously aware of. On the other hand, alignment can be the result of conscious cooperation [9] or cognitively demanding accomplishments like maintaining a full common ground [10]. The apparent disagreement, however, is not an either-or: both sides supplement each other and account for different aspects of the same phenomena, ranging from unconscious adaptations to deliberate repairs – see [11] for a pertinent early position paper. The higher-order, conscious alignment phenomena can be (and actually are) captured in terms of mental representations (classical symbolic BDI architecture: belief, desires, intentions), propositions, partner models and the like, and are topics of psychological as well as linguistic research on natural language dialogue.

But how to account for the lower-level, automatic alignment processes? Within formal dialogue theory, automatic alignment has not been acknowledged until very recently: [12, pp. 54f] propose to account for alignment by means of a dialogue routine whose genesis, however, is merely postulated referring to repeated instances of some action pattern. In order to prepare a non-stipulative answer that shows how such routines may come into being and thereby pointing out an interface between symbolic and sub-symbolic modeling, let us recapitulate the key features of the mechanistic alignment process. Alignment in this sense is:

(i) operative on mental representations;
(ii) non-intentional (i.e., not subject to deliberate access);
(iii) persistent in the sense that it brings about routinized, repetitive behavior;
(iv) different and differentiable from non-alignment[2].

In order to understand properly how priming can bring about alignment and in which way(s) it differs from non-alignment[2], we need an explicit, cognitive model of it. In addition to the features itemized in the list above, the model has to be formalized, allowing for computational applications and thereby grounding psycholinguistic accounts of priming.

Support for this work by the German Federal Ministry of Education and Research (*Linguistic Networks* project) and the German Research Foundation (CRC 673 *Alignment in Communication*) is gratefully acknowledged.

Goethe-University Frankfurt am Main (email: mehler@em.uni-frankfurt.de, luecking@em.uni-frankfurt.de).
Bielefeld University (email: peter.menke@uni-bielefeld.de).

[1]The scene is a landmark which is part of a virtual town. The sample dialogue extract occurs at the 10th minute in dialogue 24 from the Bielefeld SaGA direction dialogue corpus [8].

[2]For reasons of simplicity, let us assume that the opposite of alignment is just non-alignment – matters are actually a bit more complex [13].

A network model which has a neural interpretation suits these requirements [14]. We will elaborate on each of the required features in the following.

(i) *Priming affects mental representations.* At first glance this proposition poses a problem. Being a mental phenomenon, a representation is abstract and not part of the causal neural-physical chain. But like any mental representation, linguistic representations are maintained by some brain states or neural processes – the "wetware", to use philosophical parlance that follows the "software"/"hardware" distinction [15]. Representations have a neural substrate that *is* part of the physical world. There is a long-lasting, still ongoing debate about the exact relationship between the mental and the physical, a debate that we cannot contribute to. Notwithstanding the mind-body problem, it is sufficient for our concern expressed as requirement (i) above, if priming can be understood to trigger neural activation. This point is also recognized by [7] in their original formulation of the IAM, where the "alignment channels" are introduced metaphorically and explained by reference to exchange via the physically realized speech signal. The wetware that (somehow) carries representations exhibits a design feature we have to account for: interconnectedness. Neurons are related to their neighboring neurons by synapses, giving rise to a connectionist neural net architecture. Connectionist models have been successfully employed in simulating human information processing of various kinds [16]. In a related vein, a linguistic network model of dialogue lexica has a neural interpretation, *modulo* the mind-body problem. Priming in psychology denotes the effect that the processing of a stimulus s_p is influenced by a previous stimulus s_t. In this case, the prime s_p is primed by the target s_t. Typical priming relations are semantic or associative ones. For instance, the target word *bird* primes *Ostrich* semantically – both flying birds and running Ostriches belong to a common family –, and *worm* associatively, since it is the early bird that gets the worm. In the context of a target word the primed words are more rapidly processed, as tested in numerous response experiments of various kinds (see [17] for an example of semantic priming). Neural networks provide an elegant explication for the psychological notion of priming in terms of spreading activation due to decreased thresholds between coupled information units [18]. In the present paper decreased thresholds are modeled by edges of increased weights between nodes that represent linguistic representations. In our case, the nodes represent lexemes, since our concern is to account for lexical alignment. Let us illustrate a lexical network by means of the simple example from Fig. 1. Suppose a set of lexemes already networked in prior experience (t_1 in Fig. 1). The edges in the lexical network can stand, for example, for sense relations like hyponymy or can be derived from usage-based studies that account for co-occurrences in utterances. Either way, there will be some lexemes that are more strongly connected than other ones. This is expressed by an increased weight of their connecting edges, graphically by means of thickened lines – see nodes

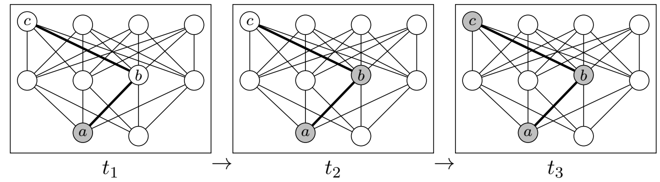


Fig. 1: A simple illustration of spreading activation in a neural network.

a, *b*, and *c* in Fig. 1. Now, if *a* is activated again at a later time t_2, the activation gets disseminated to the nodes most strongly connected to *a* (t_3). Activation of a single node leads to activation of a sub-network. Note that we model representations holistically by single nodes within the network. It would not make a difference in principle if we give up this simplification and decompose representations into a distributed activation pattern of information units. The difference would just consist in a loss of clarity and simplicity of presentation. With this remark in mind, priming can actually be seen to operate on a neural level which in turn maintains mental representations.

(ii) *Priming is non-intentional*. The distribution of weights connecting the nodes within the lexical network is independent of any other source, most notably, independent from higher cognitive modules responsible for explicit common ground, partner models, or BDIs. Furthermore, the priming mechanism is neither triggered nor covered by a supervising subject or processing unit. That is, it works automatically without any conscious effort.

(iii) *Priming leads to routines*. In terms of a neural network, routines can be understood as weight distributions. For instance, the correction in the street lamps example given above plus the coordinated re-use of the form *lamps* lead to a strong association between the respective discourse topic and the word form. That is, if one of the relata is activated during the course of communication, the other is triggered, too. Furthermore, every time the connection is invoked, it is intensified. Thus, routines "emerge" as stable sub-configurations within (neural) networks.[3]

(iv) *Alignment by priming differs and is differentiable from non-alignment.* This proposition sounds trivial at first sight and it certainly would be regarded so in the context of psycholinguistic alignment research. However, matters change if we come to formal models of priming and alignment. Such models should prove that they are valid, that is, that they capture the relevant features of the phenomena they pretend to model. If one can show that the model is able to distinguish aligned from not so aligned dyads and/or that the model is not influenced by random variables, these would count as a desired proof of validation.

The foregoing list of characteristics does not provide much of new insight. With the exception of the technical one (iv) they are already recognized and empirically investigated –

[3]The "alignment pattern" installed within a dyad may then give rise to a "communication ontology" in that dyad [14].

again, see [7] for an overview. Nonetheless, we stress these features since they provide test cases for computational models of alignment in dialogue. The network model presented in section II not only fulfills the content-driven requirements (i) to (iii), but also accounts for (iv). We already exposed the model to the random baseline test [19]. In section IV we evaluate it against annotated data gained in an experimental setting (the data are described in section III).

II. TiTAN

In this section, we present our model of dialogue networks, that is, *Two-Level Time-Aligned Networks* (TiTAN), which we use to model lexical alignment in dyadic conversations. This model was introduced in [19] in the area of task-driven conversations and further applied in the area of more spontaneous conversations [20]. The architecture of a *Two-Level Time-Aligned Network* is as follows: an encompassing graph G models a whole two-person (dyadic) dialogue. G is partitioned into two levels, A and B, representing each interlocutors' dialogue lexicon. A TiTAN is serialized according to the contributions of the interlocutors manifested and organized as turns. Formally, TiTANs are graphs $G_{AB} = (V,E)$ in which the vertex set V is partitioned into non-empty disjunctive subsets V_A and V_B, containing, in the application of this paper, the lemmas of the words used by the interlocutors A and B, respectively. The edge set E is partitioned into three non-empty disjunctive subsets E_{AB}, E_A, E_B so that all edges $\{v,w\} \in E_{AB}$ end at vertices $v \in A, w \in B$, while all other edges $\{x,y\} \in E_X$ end at vertices $x,y \in V_X$, whereas X is the meta variable $X \in \{A,B\}$. E_X captures intra-lexical relations, while E_{AB} links shared lexical items of the interlocutors. The sub-graphs $G_A = (V_A,E_A)$ and $G_B = (V_B,E_B)$ are called the A- and B-level of the two-level graph $G = (V,E)$ and are denoted by the projection function $\pi_A(G_{AB}) = G_A$, $\pi_B(G_{AB}) = G_B$ and $\pi_{AB}(G_{AB}) = (V,E_{AB})$. In terms of our application area, level A represents the dialogue lexicon of interlocutor A, level B the dialogue lexicon of interlocutor B. The graph G provides a unified model of the overall dyadic dialogue lexicon.

The networks defined so far are expressive enough to model linguistic representations and relationships obtaining among them. They do not distinguish between frequently-used and seldom-used representations. This asymmetry is accounted for by assigning weights, which complete the TiTAN model. Dialogue lexica are modeled as weighted labeled graphs $L_t = (V,E_t,\nu_t,L)$ that are indexed by the point in time $t \in \mathbb{N}$ at which they are built. Recall that we derive t from the dialogue turns the interlocutors produce, that is, from a dialog-inherent time-related ordering. The vertices $v \in V$ are labeled by the surjective function $L : V \mapsto L_V$ for the set of labels L_V. In our case, L_V consists of lemmas. As a two-level graph, L_t is divided into two sub-graphs $\pi_A(L_t) = L_{A_t} = (V_A,E_{A_t},\nu_{A_t},L_A)$ and $\pi_B(L_t) = L_{B_t} = (V_B,E_{B_t},\nu_{B_t},L_B)$ according to the distribution of L_t over the agents A and B at time t.

So much for the graph model. The construction of a TiTAN proceeds as follows [19, p. 1453]:

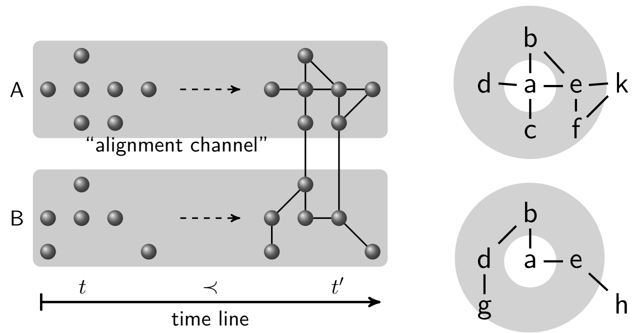
- *Intra-personal links:* if at time t, agent $X \in \{A,B\}$ uses a form belonging to lemma l to express the current turn's topic $T = T(t)$, we generate intra-personal links between $v_X \in V_X, L_X(v_X) = l$, and all other vertices labeled by items that X has used in the same round or in the preceding round on the same topic T. If the links already exist, their weights are incremented by 1.
- *Interpersonal links:* if at time t, agent $X \in \{A,B\}$ uses $l \in V_t$ to express topic $T = T(t)$ that has been expressed by $Y \neq X$ in the same or preceding round on the same topic by the same item, we span an interpersonal link $e = \{v_A,w_B\} \in E_t$ between $v_A \in V_A$ and $w_B \in V_B$ for which $L_A(v_A) = L_B(w_B) = l$, given that e does not already exist. Otherwise, its weight $\nu_t(e)$ is increased by 1.

Note that the construction procedure implements priming as a short-term effect. Short-term priming has to be distinguished from long-term priming, or adaptation [21], [22]. Though [7] do not distinguish these two kinds of priming explicitly, the IAM most likely rests on a short-term mechanism, since its basic setting is ephemeral dialogue. A TiTAN series reflects this by the turn restriction of its construction. The construction procedure could easily be modified in order to account for more delayed priming effects.

Figure 2(a) provides a schematic visualization of a TiTAN series. The starting points are the completely unlinked lexica of lemmas of interlocutor A and B. Following the aforementioned construction procedure, the lexica are networked turn-wise by adding appropriate intra- and interpersonal links (edges). A TiTAN, thus, models for each turn the degree of structural coupling of the dialogue lexica of both interlocutors A and B.

The degree of structural alignment between dialogue lexica is assessed by comparing the neighborhoods of corresponding vertices from the lexica of interlocutors A and B. Remember that the vertices are labeled by lemmas. If we take the pairs of vertices from A and B denoting critical objects and collect the vertices that are related to them with distance $1,\ldots,k$, we can determine the intersection of equally labeled vertices for each neighborhood circle. In order to illustrate the starting point for the approach to measure the structural similarity between graphs, the first neighborhood circle is visualized in Fig. 2(b). In the example, you can also observe that intersection of spheres of equal distance does not take indirectly related vertices into account, vertices exemplified by d in Fig. 2(b). Neighborhood circles have to be generalized to a comparison of spheres of unequal rank. In the didactic example, vertex d is detected by comparing A's 1-sphere with B's 2-sphere. The Mutual Information measure of [19] takes care of this and other logically possible cases. Since a discussion of the graph-theoretically measure for mutual information would lead us astray in this paper, we refer the reader to [19, Sec. 4.2].

The measure derived from the mutual information between the partitions of a bipartite graph as illustrated above has the following properties: if both interlocutors use the critical



(a) Illustration of a *Two-level Time-Aligned Network* series. Initially, the lexica of interlocutors A and B are unlinked (left of dashed arrows). Dialogue turn by dialogue turn, the interlocutors' lexica are networked such that a dialogue lexicon emerges that is spanned by intra- and interpersonal links across the alignment channel (right of dashed arrows).

(b) How similar are the lexica reached in (a) on the left? The answer is based on spheres of a certain distance around paired prime vertices – a in the example above.

Fig. 2: (Co-)Activation of representations within the dialogue networks of interlocutors A and B: A TiTAN series illustration (a) and the point of departure for a structural graph similarity measurement (b). The first neighborhood circles, the *1-spheres*, for the a-labeled vertices from (b) are: $\{b,c,d,e\} \cap \{b,e\} = \{b,e\}$. Observe further, that vertex d, which is part of A's 1-sphere, is an element of B's 2-sphere.

words in exactly the same way, the intersection for each neighborhood pair equals their union. The similarity measure is normalized so that it returns its maximum of 1 in this case. If the interlocutors use the respective words in a completely unrelated manner, the intersection will be the empty set and denominated by the minimum value 0.

Quantifying the degree of coupling in between the two extrema would provide a reference point for assessing alignment. [19] provide such a measure in terms of a *Mutual Information-based Distance Measurement* for graphs. They draw on [23] who proved the following relationship between the mutual information $I(X;Y)$ of two random variables and their distance $D(X;Y)$:

$$D(X,Y) = 1 - \frac{I(X;Y)}{\max(H(X),H(Y))} \in [0,1] \tag{1}$$

Informally, if X and Y are statistically independent (i.e., knowing X does not provide (much) information about Y, and vice versa), then X and Y are highly distant (and conversely). The distance measure $D(X,Y)$ is a cognate measure of the graph distance measure based on graph union proposed by [24]:

$$d_W(G_1,G_2) = 1 - \frac{|mcs(G_1,G_2)|}{|G_1| + |G_2| - |mcs(G_1,G_2)|} \tag{2}$$

$D(X,Y)$ as well as d_W assess the (dis-)similarity of graph structure in terms of graph distances. In this line of research, [19] developed a graph-theoretical measurement of the similarity of dialogue lexica modeled by bipartite graphs. In this paper, we reuse this model to classify 55 dyadic conversations. These conversations are collected in the so called JMG corpus, which is described in the next section.

TABLE I: Corpus size in terms of the number of annotations on various levels.

LEVEL	GLOBAL #	AVG # PER DIALOG
dialogues	64	—
words	93,120	1,501.935
turns	28,380	457.742
events	1,731	27.919
event phases	5,153	83.113
lexops	4,415	71.210
repairs	3,327	53.661

III. Statistics of the JMG Corpus

The *Jigsaw Map Game Corpus* (JMG Corpus, [25], [13]) consists of 64 richly annotated experimental dialogues. Interlocutors of these dialogues had to collaborate in arranging a set of objects on a table according to target arrangements. These were given to the participants as cards with photographs of partial object constellations. In each round, one participant described the constellation while the other had to reproduce it on the basis of the description by inserting one additional object. After each round, roles were switched.

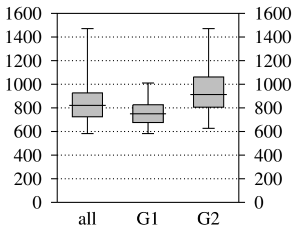
Experiments always consisted of a sequence of two games: In the first game (G1), a confederate played with a naïve participant. On this occasion, the confederate could use special words for certain critical objects (depending on the experimental requirements). Afterwards, the now-somewhat-experienced participant played a second game (G2) with another naïve participant unacquainted with the game.

Audio and video recordings were made during those experiments (resulting in ∼18 hours of material). These have been transcribed on the levels of turns and words, including temporal information. In addition, the following data has been annotated on the basis of the transcripts:

1) Rounds and subordinate phases of the JMG (coded as the running number of the current card, a description of its contents and the name of the subordinate phase),
2) morpho-syntactic information about the structure of nouns denoting critical objects, and
3) major elements of repairs occurring during dialogue.

Table I contains information on the amount of annotations that have been created for all these levels. The following example (from dialogue 1.1) shows one round of the game, separated into three phases, namely *object identification* (ID), *localization* (LOC) and *adjustment and confirmation* (ADJ). In this round, the constructing participant does not speak at all:

> [So, du müsstest einen grünen, ähm, Buntstift haben.]$_{\mathrm{ID}}$ [Und der müsste von dir aus gesehen auf die rechte Seite von der Spielfigur zu dir hin.]$_{\mathrm{LOC}}$ [Ja.]$_{\mathrm{ADJ}}$
> *[Well, you should have a green, ehm, pencil.]*$_{\mathrm{ID}}$ *[And it should go, from your perspective, to the*



	median (s)	(mm:ss)
all	821.1	13:41
G1	750.0	12:30
G2	913.0	15:13

Fig. 3: Average dialogue lengths, for all dialogues and for each of the two games (G1, G2) of a sequence.

(a)

(b)

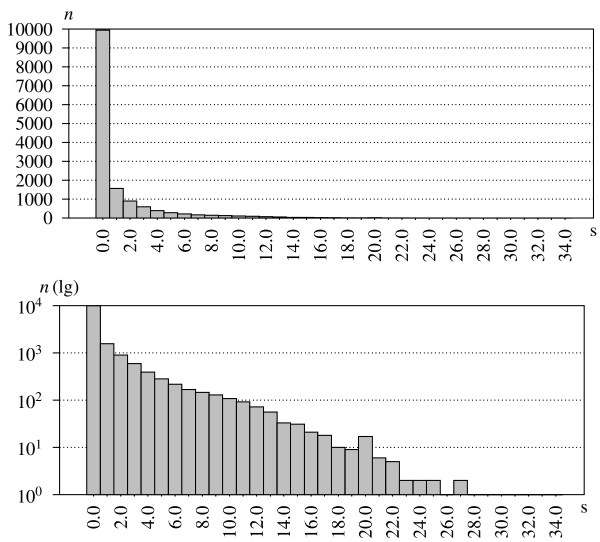


Fig. 4: Histograms of the turn length distribution for all dialogues, shown (a) on a linear scale and (b) on a semi-logarithmic scale.

right-hand side of the token, towards you.]$_{\text{LOC}}$
[Yes.]$_{\text{ADJ}}$

A. Characteristics of the Dialogues

The 64 dialogues have an average duration of 13:41 minutes (cf. Fig. 3). There is a notable difference between the length distributions of dialogues from different games in the experimental sequence: G1 dialogues are significantly shorter than G2 dialogues ($t = -4.852$, $p < 10^{-5}$). A probable explanation would be that G1 dialogues always involved one of several confederates who had profound insight into the experimental design, and especially in the set of critical objects. They also had special instructions related to the choice of lexical elements and syntactic constructions. This knowledge and behavior obviously resulted in turn lengths that were both shorter and less variant than those of naïve speakers.

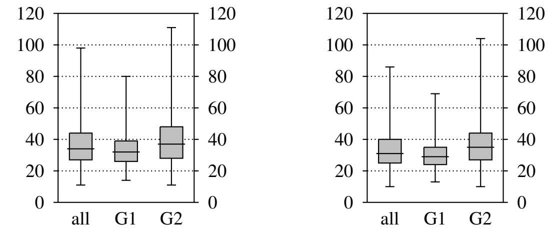
B. Characteristics of Turn Lengths

The two histograms in Fig. 4 show that the distribution of turn lengths is monotonically decreasing (to be more precise, it strongly suggests a power law distribution). While there are 9,940 turns with a duration of less than one second, the number of turns with a duration of 1–2 seconds drops to 1,566.

The apparent cause for this distribution lies within the concept of turns used in the model. Due to the amount of data annotated, any continuous sequence of utterances of one interlocutor that is bordered by either pauses or utterances of the other interlocutor is considered a turn. Since the interlocutors collaborate in a task that involves descriptions and movements of objects, they produce short confirmations and feedback signals frequently, which leads to the high number of very short turns. It can be assumed that the distribution was likely to change as soon as a more complex and detailed turn model (as used in various Conversation Analysis theories [26]) were adopted.

The second diagram in Fig. 4 reveals another oddity: Unlike most of the rest of the data, turns with a duration of ∼20s deviate slightly from the power-law-like distribution.

In contrast to dialogue lengths and round lengths, the distribution of turn lengths from both games G1 and G2 do *not* differ significantly from each other ($t = -.542$, $p = .5881$).



	mean	std. dev.	skewness
word forms	36.984	15.125	1.526
lemmas	33.618	13.099	1.389

Fig. 5: Distribution of the number of word forms (left) and lemmas (right) used per game round.

C. Vocabulary

Per game round, interlocutors used ∼37 word forms and ∼34 lemmas on average. Both distributions are right-skewed (detailed values can be found in Fig. 5). This could be caused by the fact that in some cases, the participants could not succeed immediately at positioning the object – they had to negotiate positions and relations between the target and other objects, and the describer had to present additional information. Naturally, this results in additional utterances, words and lemmas. These incidents increase the number of above-average sized rounds (if you assume that the set of successful rounds roughly corresponds to a normal distribution).

The difference between G1 and G2 dialogues is also observable on the level of word forms and lemmas: G1 dialogues contain very significantly fewer word forms ($t = -7.851$, $p < 10^{-14}$) and lemmas ($t = -8.636$, $p = 0$) per game round than G2 dialogues.

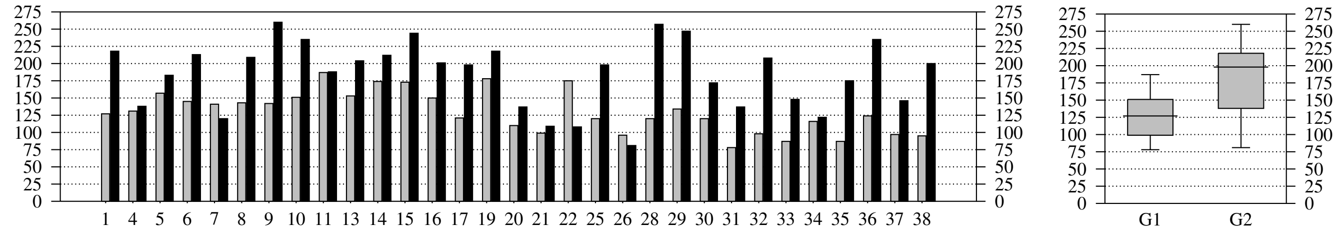


Fig. 6: Sizes of dialogue lexica for all dialogues. On the left: Bar chart, grouped by sequences (light bars represent G1 dialogues, black bars represent G2 dialogues). On the right: box plots of both distributions.

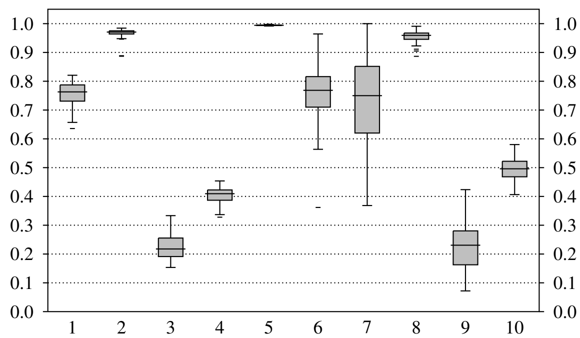


Fig. 7: From left to right, boxplots of 10 network indices as calculated by example of TiTAN series representing the 55 dialogues within the JMG corpus (see Section III): (1): Watts & Strogatz's cluster coefficient [27]; (2): geodesic compactness [28]; (3): network cohesion (graph size in relation to the size of a completely connected graph); (4): weighted cluster coefficient [29]; (5): local graph entropy [30]; (6): Wallis' graph similarity [24]; (7): graph bipartivity [19]; (8): Newman's network modularity [31]; (9): mutual information of two-level graphs [19]; (10): transitivity compactness [32].

Fig. 6 gives an overview of the size of the dialogue lexicon for each experiment.[4] Yet again, G1 and G2 dialogues differ significantly: G2 dialogues have a larger dialogue lexicon than G1 dialogues ($t = -5.349$, $p < 10^{-5}$).

The statistical description of the JMG corpus not only provides a detailed introduction of the data we make use of in the following, it also points to the genuine difference between confederate and naïve dialogues. Notwithstanding these results, the present section also illustrates that descriptive statistics are not appropriate to come closer to (lexical) alignment in dialogue. For that reason, in the next section, we present a classification that uses the JMG corpus to automatically separate aligned from non-aligned dialogue lexica represented as the final states of TiTAN series, the model introduced in section II.

IV. Experimentation

In this section, we evaluate our network model by example of the JMG corpus described in Section III. We test the hypothesis that dialogues can be classified according to their status as manifesting alignment or non-alignment by analyzing the topology of the dialogue lexica of the corresponding interlocutors. More specifically, we examine whether interlocutors are lexically aligned in task-oriented conversation by examining the topology of their dialogue lexicon. In [19], we tested this hypothesis by example of 24 dialogues showing that alignment is reflected by the topology of lexical networks. Now, we test the same hypothesis by means of 55 dialogues (including 47 instances of alignment and 8 instances of non-alignment) from the whole set of the 64 JMG dialogue corpus. 11 dialogues were excluded due to their intermediary status. Using TiTAN series to represent dialogue lexica as two-level lexical networks (see Section II), we apply *Quantitative Network Analysis* (QNA) [33] to test this classification hypothesis. QNA aims to learn to discriminate classes of networks by their distinctive topological features. Starting from a classification of a set of networks, QNA seeks structural features that best distinguish the networks in the sense of this classification. QNA involves three steps of modeling:

1) *Graph modeling*: firstly, each target object (i.e., graph) is represented by a vector of topological indices that model its micro, meso or macro level structure. In the present scenario, these indices operate on the final state networks of the TiTAN series described in Section II, that are the networks spanned for completed dialogues.
2) *Feature selection:* a genetic search is performed to find those topological features that best separate the target classes. This search utilizes the F-measure (i.e., the harmonic mean of precision and recall). This selection may stop at a local maximum in the case that it does not find an optimal feature subset.
3) *Classification:* based on the selected feature vectors, a hierarchical agglomerative clustering is performed together with a subsequent partitioning that knows the number of target classes. Classification and feature selection occur in a single step such that the selection includes a search for the best performing linkage method and distance measure. We use *complete linkage* and the *Mahalanobis distance* to perform this step. All computations are made using MATLAB.

In a nutshell: QNA takes the set of input networks together with the parameter space of linkage methods and distance measures to find out the subset of features that best separate

[4] For this summary, dialogue lexica were modeled as sets of pairs of lemmas and part of speech values.

these networks according to the underlying classification. In this study, we utilize the network model of TiTAN series introduced in [19]. All in all, 103 topological indices were computed per dialogue network to model its topological characteristics. In this context, a simple measure of lexical alignment is the index of bipartivity, that is, the number of equally used words shared among interlocutors in relation to the overall number of lexemes common to them. Table II displays the range of values of this and related topological indices as observed in the network representations of the dialogues in the JMG corpus. Obviously, our graph model captures indices of a wide spectrum of variability: on the one hand, graph entropy and geodesic compactness are of low variance making dialogue lexica hardly distinctive. On the other hand, bipartivity and mutual information (see Section II, Figure 2b and Equation 1) are of high variance. This also shows that lexical overlaps as mapped by indices like bipartivity are bad indicators of alignment. In other words: to use the same words is not sufficient for being lexically aligned and vice versa. Rather, lexical alignment is a latent variable that is reflected by topological characteristics, which are hardly measured by a single topological index. Actually, value ranges as displayed in Table II do not sufficiently indicate separability, since the target classes are not linearly separable. This means representations based on multiple indices are indispensable.

Table II shows the results of several classification experiments based on QNA. The best performing classification uses complete linkage in conjunction with the Mahalanobis distance. With an F-score of .92 it outperforms two random clusterings: the one assuming an equi-partitioning among the target classes, the other knowing the number of objects within the target classes. As the difference between the best performing random clustering and the best performing QNA-based clustering is more than 17%, we can assume that our classification hypothesis about the topological separability of dialogue lexica is *not* falsified. Thus, we can retain it (at least until it is falsified in the future). At the same time, we observe a decreasing F-score (from 1 to .92) compared to the classification reported in [19]. This loss may be due to the increasing number of dialogues manifesting alignment in relation to the more or less constant number of only 8 dialogues that manifest non-alignment. Obviously, the sort of alignment in task-oriented dialogues as collected in the JMG corpus departs from the alignment in more spontaneous dialogues as exemplified by direction giving, which provide more test cases for non-alignment (see [20]). In any event, an F-score of more than 90% demonstrates the classificatory power of QNA based on TiTAN series as presented here.

Table III shows the result of learning three target classes. This classification comprises in addition to the set of 55 instances of aligned and non-aligned conversations a set of 6 borderline cases that were not unambiguously mapped to these two classes by the human annotator. All in all, Table III considers 61 dialogues. It shows an F-score of around 91% for QNA, that is, a loss in only about 1%

TABLE II: Summary of the results of differently parameterized *Quantitative Network Analyses* (QNA). Row no. 7 shows the average F-score of these variants. The last column denotes the number of features output by the genetic search of the best performing subset of features as part of QNA. All features are based on the TiTAN model described in Section II.

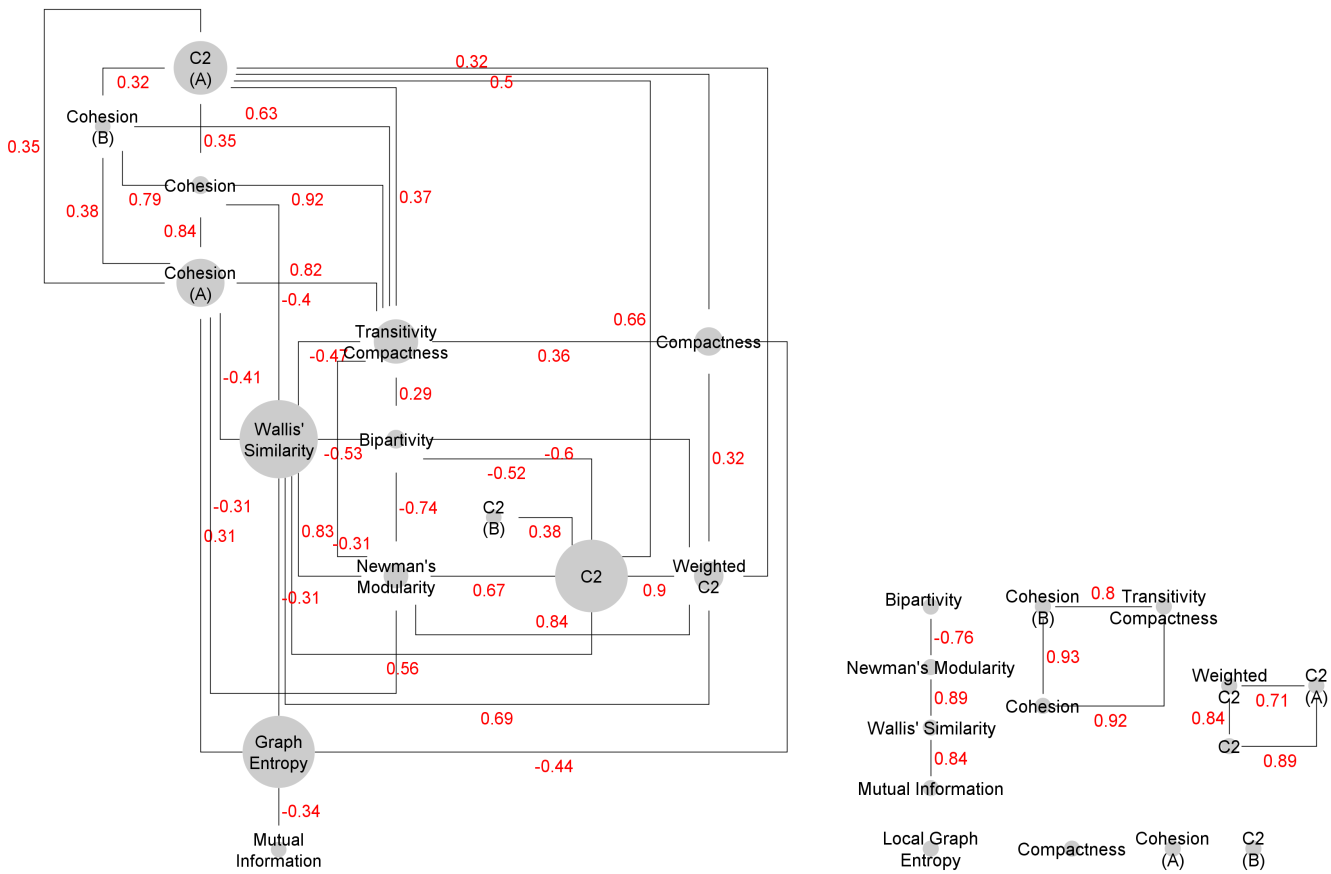
No.	Procedure	F-score	Feature Selection
1.	QNA[Mahalanobis,complete]	.92284	47 / 103
2.	QNA[Mahalanobis,complete]	.90654	46 / 103
3.	QNA[Mahalanobis,Ward]	.89578	49 / 103
4.	QNA[Mahalanobis,Ward]	.88425	48 / 103
5.	QNA[Mahalanobis,Ward]	.88068	43 / 103
6.	QNA[correlation,complete]	.861	49 / 103
7.	average over non-random approaches	.8918	47 / 103
8.	random baseline known-partition	.75073	
9.	random baseline equi-partition	.61788	

compared to Table II. However, we also see that the F-scores of the corresponding baseline scenarios are heavily reduced compared to Table II. Thus, there is a gain of nearly 30% of QNA compared to the best performing random baseline. This is further evidence in support of the feasibility of classifying cognitive-linguistic states by means of complex network theory.

Figure 8 sheds some light on the kind of data sparseness that we face in classifying dialogue data. Although 55 dialogues is, without any doubt, a large data set (as annotating this data is highly labor intensive), the number of instances of non-alignment is too small to measure correlations among the topological indices under consideration (see Figure 8, right). However, when looking at the set of dialogues manifesting alignment, correlations of topological indices can be detected. As shown in Figure 8, there is a high correlation between the cluster coefficient C2 of [27] and its weighted counterpart. There is also a higher correlation between both variants of compactness studied here. However, we also observe that there is no correlation between Wallis' measure of graph similarity and the mutual information presented in equation 1 – although both measures are used to measure graph similarity. This is a hint on the independence of both measures and, thus, additional evidence for their distinctive role in network classification as studied in [19]. Further, correlation analyses of this sort help to test network models in the area of symbolic and distributed neural networks. In this sense, we provide methodical insights in support of graph-theoretical models of neural networks – from the point of view of natural language communication.

V. Conclusion

In this paper, we tested a cognitive model of linguistic networking on the level of lexical units. More specifically, we presented a model of lexical alignment as the result of the dynamics of dialogical communication of two interlocutors. From the point of view of neural network theory, our model



(a) For correlations that are computed for dialogues that manifest lexical alignment among interlocutors, Wallis' graph similarity measure is of highest centrality among the 10 indices (see also Figure 7).

(b) The correlation network of the same indices by example of the set of dialogues manifesting non-alignment.

Fig. 8: Correlation networks of network indices for the dialogues of aligned dyads (a) and not aligned dyads (b). Only significant (positive or negative) correlations are displayed for which the probability p to observe the same or higher correlations by random chance is less than 0.05. The size of the vertices (gray dots) denotes the betweenness centrality of the corresponding index in the correlation network.

TABLE III: Summary of the results of differently parameterized *Quantitative Network Analyses* (QNA) concerning three target classes.

No.	**Procedure**	F**-score**	**Feature Selection**
1.	QNA[Mahalanobis,complete]	.91966	51 / 103
2.	random baseline known-partition	.63146	
3.	random baseline equi-partition	.470798	

focuses on symbolic representations that span linguistic networks generated by the law-like dynamics of dyadic conversations. In order to map the structural peculiarities and similarities of these networks, we applied several topological models of graph theory including the statistics of geodesics, the entropy of graphs and of graph similarity measurement.

We started with presenting various results on the quantitative profile of dialogical data. More specifically, we characterized 64 dialogues as part of the so-called JMG corpus. To the best of our knowledge, this is one of the largest dialogue corpora in the area of research on alignment in communication. We have shown the characteristic length separation between dialogues to which confederates participate and those to which only naïve probands participate. Further, we analyzed the distribution of turn lengths, showing that the majority of turns tend to be very small in the sort of task-oriented communication studied here. Interestingly, we found evidence for a scale-free behavior of this distribution, which hints at a process of preference ordering, subject to the short-term dynamics of dyadic conversations. Note that the conversations covered by the JMG corpus are rather short, at least compared to the amount of data available for written text studies conducted on, say, the English Wikipedia. Nevertheless, they seem to be characterized by the sort of scale-freeness that has been found to be characteristic of diverse phenomena such as linkage in the WWW [34] and collaboration in social networks [35].

Beyond this statistical description of dialogue data, we provided insights into the separability of dialogues concerning their status as instances of aligned and non-aligned communication, respectively. Using the classificatory apparatus of *Quantitative Network Analysis* (QNA) and building upon TiTAN series as a graph model of dyadic conversations, we presented a graph model that sufficiently separates both classes of dialogues in a non-random fashion. An F-score of around 92% clearly indicates the classificatory potential of QNA in the area of alignment phenomena. This finding continues a series of results that have recently shown how purely cognitive phenomena such as alignment can be made an object of machine learning by using network models ([36], [19], [20]) which, finally, bridge to the area of (symbolic representations in) neural networks.

Beyond that, we have shown that well-known models of complex network theory (as, for example, clustering or transitivity, geodesic compactness and modularity) span a sparse correlation network. This indicates both the relative independence of topological indices studied here as well as their non-linear relatedness, which makes classifications of networks a genuine task of machine learning. In a nutshell: there is no single index or pair of indices that clearly separate our target classes of dialogues. Rather, we needed to account for the non-linear interaction of these indices in order to arrive at a representation model of linguistic networks that makes them separable in the sense of sufficiently high F-scores. This methodological result also has implications for psycholinguistic assessments of the degree of alignment in dialogue. Researchers who want to evaluate their empirical data on alignment may not concentrate on just one index or a few indices, they rather also have to look on the relationships between the indices. Thus, quantifying alignment should involve a kind of meta-evaluation. If the network of indices proves to have stable traits (proven in a couple of applications on empirical data), these traits can be used in stating alignment-related hypotheses and, hence, would play a role in generating predictions for empirical research. As the F-scores that we have found clearly outperform two related baseline scenarios, it seems to be reasonable that our model demonstrates the feasibility of classifying alignment data.

In this sense, our results are in support of applying and extending the apparatus of machine learning in order to map phenomena of dialogical communication that, so far, seemed to be out of reach because of their ephemeral, cognitive nature. However, the cooperation of diverse fields such as cognitive modeling, linguistic networks, complex network theory and unsupervised learning seems to provide a methodical access to dialogical data. Notwithstanding this finding, we have to face the labor intensity of preprocessing dialogue data that makes it hard to work with amounts of data as huge as commonly analyzed in related areas of computational linguistics. This problem of data sparseness of research on dialogical communication will certainly last for decades. Thus, analyses even on smaller data sets as provided here gain in importance. Future work will deal with extending this data set as well as our graph model. The aim is to handle both more complex linguistic representations (above the lexical level) as well as multilingualism as commonly observed in conversations of immigrants. Especially the alignment of complex linguistic expressions such as multimodal ensembles ([37], [38]), multi-word units and units on the (sentence) phrasal level will be the focus of future work in this area.

REFERENCES

[1] H. Giles and P. F. Powesland, *Speech Styles and Social Evaluation*. London ; New York: Academic Press, 1975.

[2] H. H. Clark and D. Wilkes-Gibbs, "Referring as a collaborative process," *Cognition*, vol. 22, pp. 1–39, 1986.

[3] H. P. Branigan, M. J. Pickering, and A. A. Cleland, "Syntactic coordination in dialogue," *Cognition*, vol. 25, pp. B13–B25, 2000.

[4] H. P. Branigan, M. J. Pickering, A. J. Stewart, and J. F. McLean, "Syntactic priming in spoken production: Linguistic and temporal interference," *Memory & Cognition*, vol. 28, no. 8, pp. 1297–1302, 2000.

[5] S. Garrod and A. Anderson, "Saying what you mean in dialogue: a study in conceptual and semantic co-ordination," *Cognition*, vol. 27, no. 2, pp. 181–218, 1987.

[6] M. E. Watson, M. J. Pickering, and H. P. Branigan, "An empirical investigation into spatial reference frame taxonomy using dialogue," in *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 2006, pp. 2353–2358.

[7] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *Behavioral and Brain Sciences*, vol. 27, pp. 169–226, 2004.

[8] A. Lücking, K. Bergmann, F. Hahn, S. Kopp, and H. Rieser, "The Bielefeld speech and gesture alignment corpus (SaGA)," in *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*. Malta: 7th International Conference for Language Resources and Evaluation (LREC 2010), 5 2010, pp. 92–98.

[9] H. H. Clark, *Using Language*. Cambridge: Cambridge University Press, 1996.

[10] R. Stalnaker, "Common ground," *Linguistics and Philosophy*, vol. 25, pp. 701–721, 2002.

[11] G. Rickheit, "Alignment und Aushandlung im Dialog," *Zeitschrift für Psychologie*, vol. 213, no. 3, pp. 159–166, 2005.

[12] M. Poesio and H. Rieser, "Completions, coordination, and alignment in dialogue," *Dialogue and Discourse*, vol. 1, pp. 1–89, 2010.

[13] P. Weiß, O. Pustylnikov, A. Mehler, and S. M. Hellmann, "Patterns of alignment in dialogue: Conversatioanl partners do not always stay aligned on common object names," in *Oral presentation at the Conference on Embodied and Situated Language Processing (ESLP 2009), July 28-29, Rotterdam, The Netherlands*, 2009.

[14] H.-J. Eikmeyer and J. Salas, "Alignment und Ontologien," in *GI Jahrestagung (1)*, 2007, pp. 163–168.

[15] A. Clark, *Mindware: An Introduction to the Philosophy of Cognitive Science*. New York, NY: Oxford University Press, 2001.

[16] D. E. Rumelhart, *Introduction to Human Information Processing*. New York: John Wiley & Sons, 1977.

[17] J. H. Neely, "Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention," *Journal of Experimental Psychology*, vol. 106, no. 3, pp. 226–254, 1977.

[18] S. Grossberg, "Nonlinear neural networks: Principles, mechanisms, and architectures," *Neural Networks*, vol. 1, no. 1, pp. 17–61, 1987.

[19] A. Mehler, A. Lücking, and P. Weiß, "A network model of interpersonal alignment," *Entropy*, vol. 12, no. 6, pp. 1440–1483, 2010.

[20] A. Mehler, A. Lücking, and P. Menke, "Modelling lexical alignment in spontaneous direction dialogue data by means of a lexicon network model," in *Proceedings of 12th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), February 20–26, Tokyo*, ser. LNCS. Berlin/New York: Springer, 2011, accepted.

[21] K. Bock and Z. M. Griffin, "The persistence of structural priming: Transient activation or implicit learning?" *Journal of Experimental Psychology*, vol. 129, no. 2, pp. 177–192, 2000.

[22] V. S. Ferreira and K. Bock, "The functions of structural priming," *Language and Cognitive Processes*, vol. 21, no. 7, pp. 1011–1029, 2006.

[23] A. Kraskov and P. Grassberger, "MIC: Mutual information based hierarchical clustering," in *Information Theory and Statistical Learning*, F. Emmert-Streib and M. Dehmer, Eds. New York: Springer, 2008, pp. 101–123.

[24] W. D. Wallis, P. Shoubridge, M. Kraetz, and D. Ray, "Graph distances using graph union," *Pattern Recogn. Lett.*, vol. 22, no. 6-7, pp. 701–704, 2001.

[25] P. Weiß, T. Pfeiffer, G. Schaffranietz, and G. Rickheit, "Coordination in dialog: Alignment of object naming in the jigsaw map game," in *Proceedings of the 8th Annual Meeting of the Cognitive Science Society of Germany*, 4 2008, pp. 1–17.

[26] V. H. Yngve, "On getting a word in edgewise," in *Sixth regional meeting of the Chicago Linguistic Society*, vol. 6. Chicago: Chicago Linguistic Society, 1970, pp. 657–677.

[27] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[28] R. A. Botafogo, E. Rivlin, and B. Shneiderman, "Structural analysis of hypertexts: Identifying hierarchies and useful metrics," *ACM Transactions on Information Systems*, vol. 10, no. 2, pp. 142–180, 1992.

[29] A. Barrat, M. Barthélemy, and A. Vespignani, "The architecture of complex weighted networks," in *Large Scale Structure and Dynamics of Complex Networks*, G. Caldarelli and A. Vespignani, Eds. New Jersey: World Scientific, 2007, pp. 67–92.

[30] M. Dehmer, "Information processing in complex networks: Graph entropy and information functionals," *Applied Mathematics and Computation*, vol. 201, pp. 82–94, 2008.

[31] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, Feb 2004.

[32] A. Mehler, "Compactness in an axiomatic perspective: a formal account of compactness measures," In preparation, 2011.

[33] ——, "Structural similarities of complex networks: A computational model by example of wiki graphs," *Applied Artificial Intelligence*, vol. 22, no. 7&8, pp. 619–683, 2008.

[34] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.

[35] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, pp. 167–256, 2003.

[36] A. Mehler, P. Weiß, P. Menke, and A. Lücking, "Towards a simulation model of dialogical alignment," in *Proceedings of the 8th International Conference on the Evolution of Language (Evolang8), 14-17 April 2010, Utrecht*, 2010, pp. 238–245.

[37] A. Lücking, A. Mehler, and P. Menke, "Taking fingerprints of speech-and-gesture ensembles: Approaching empirical evidence of intrapersonal alignment in multimodal communication," in *LONDIAL 2008: Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*, King's College London, 2008, pp. 157–164.

[38] A. Mehler and A. Lücking, "A structural model of semiotic alignment: The classification of multimodal ensembles as a novel machine learning task," in *Proceedings of IEEE Africon 2009, September 23-25, Nairobi, Kenya*. IEEE, 2009.