FADLI 2017

Proceedings of the Workshop on Formal Approaches to the Dynamics of Linguistic Interaction

Christine Howes and Hannes Rieser (eds.)

Toulouse, 17–21 July 2017

ISSN 1613-0073

Proceedings are published online at CEUR-WS.org

http://ceur-ws.org/

FADLI

http://www.christinehowes.com/fadli

ESSLLI 2017

https://www.irit.fr/esslli2017/

Preface

The Workshop on Formal Approaches to the Dynamics of Linguistic Interaction is hosted by ESSLLI 2017 in Toulouse. The idea for bringing together researchers working on different formalisms, and interested in issues raised by linguistic interaction and how the dynamics of these interactions can be modelled came from a workshop on the Mechanisms of Linguistic Interaction, held in Ghent in 2015.

We received a total of 18 full paper submissions, 12 of which were accepted for talks after a peer-review process, during which each submission was reviewed by a panel of three experts. We are extremely grateful to the Programme Committee members for their very detailed and helpful reviews. The poster session hosts 3 of the remaining submissions. The papers for all accepted talks and posters are included in this volume.

The formalisms covered by the contributors include models of syntax, (Dynamic Syntax: Eshghi et al.; Gregoromichelaki; Kempson et al.), models of semantics (TTR: Breitholtz; Cooper; Dobnik and De Graaf; psi-calculus: Lawler et al.; Rieser), models of dialogue (SDRT: Schlöder; Information States: Cooper; Larsson and Myrendal; miscommunication: Mills and Redeker) models of evolution (Utterance selection model: Michaud and Schaden; Multi-level selection hypothesis: Kempson et al.) cognitive models (Kalociński) and models from pragmatics (Argumentation theory: Schaden; Social choice theory: Nishiguchi).

Notions of the dynamics of interaction range from the interaction between intonation and meaning within an utterance (Schlöder); the inter-sentential interaction between linguistic elements (Nishiguchi); the interaction of linguistic and non-linguistic inputs within turns in a dialogue (Gregoromichelaki; Lawler et al.; Rieser); human-agent interaction and learning (Dobnik and De Graaf, Eshghi et al.); the interactions between interlocutors (Cooper; Schaden) and how this effects language change through dialogue (Breitholtz; Kalociński; Larsson and Myrendal; Mills and Redeker) through contact with different languages (Michaud and Schaden) and at the evolutionary level (Kempson et al.).

Methodologically speaking the contributions range from empirical work (corpus and experimental: Lawler et al.; Rieser; Kalociński; Mills and Redeker; Schlöder; computer simulation and machine learning: Dobnik and De Graaf; Eshghi et al.; Michaud and Schaden) to more focussed formal analyses (Breitholtz; Cooper; Gregoromichelaki; Larsson and Myrendal; Schaden; Nishiguchi) to broad brush theoretical work (Kempson et al.).

We are proud to bring together researchers working on different formal approaches to the dynamics of interaction, and hope that the workshop fosters cross-disciplinary collaboration around these issues.

Finally, we would like to thank all our contributors and programme committee members, as well as the organisers of ESSLLI 2017 for hosting our workshop.

Christine Howes and Hannes Rieser

Gothenburg/Bielefeld

July 2017

Programme Committee

Ellen Breitholtz	University of Gothenburg
Stergios Chatzikyriakidis	University of Gothenburg
Robin Cooper	University of Gothenburg
David Devault	USC Institute for Creative Technologies
Simon Dobnik	University of Gothenburg
Arash Eshghi	Heriot-Watt University
Raquel Fernández	University of Amsterdam
Eleni Gregoromichelaki	King's College London
Pat Healey	Queen Mary University of London
Julian Hough	Bielefeld University
Christine Howes	University of Gothenburg
Julie Hunter	Universitat Pompeu Fabra, Barcelona and Université Paul Sabatier, Toulouse
Kristiina Jokinen	University of Helsinki
Ruth Kempson	Kings College London
Staffan Larsson	University of Gothenburg
Peter Ljunglöf	University of Gothenburg and Chalmers University of Technology
Gregory Mills	University of Groningen, Netherlands
Paul Piwek	The Open University
Matthew Purver	Queen Mary University of London
Hannes Rieser	Bielefeld University
David Schlangen	Bielefeld University
Matthew Stone	Rutgers

Table of Contents

Oral Presentations

Poster Presentations

Oral Presentations

Reasoning Incrementally with Underspecified Enthymemes

Ellen Breitholtz
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg
ellen@ling.gu.se

Abstract

In this paper we suggest a way of analysing mismatch in perceived common ground which is the result of dialogue participants adopting different *topoi*, or inference rules, based on which they interpret enthymematic (logically incomplete) arguments in dialogue. A contributing factor to this kind of mismatch is the use of *underspecified* enthymemes, that is enthymemes which are more general than the topoi that underpin them. We will account for an example of such reasoning using a game board style semantics cast in Type Theory with Records (TTR).

1 Introduction

In this paper we will show how one argument may be interpreted differently by two dialogue participants depending on the underpinning topoi they assume the argument to be based on. This is possible since arguments in dialogue are almost always *enthymematic* i.e. drawing on tacit premises and principles of reasoning. In the particular case we will look at the argument is not only enthymematic, it is also an example of an *underspecified* enthymeme. Generally, an underspecified enthymeme is one where the information given in the premise of the enthymeme is sparse, the consequence being that a wide range of topoi potentially could be used to underpin it. This kind of mismatch of topoi may go unnoticed in cases where consensus is reached. After all, if the interlocutors agree on the conclusion of the argument there is often no reason to argue about the rationale for agreeing. However, in the example below in (1) it is made explicit that the speaker and the listener interpret the enthymeme in (1a) drawing on different topoi.

(1)

a. P: Metal was actually the reason I started doing hip hop.

b. P: ...Because I hated metal

c. J: Oh, I thought you were going to say something completely different!

This snippet of dialogue is taken from a radio program where discussion alternates with music. The interviewee is Swedish hip hop artist Petter, and much of the dialogue relates to the songs being played in the music sections. Just before the dialogue a song by a metal band has been played. Petter is being asked for his opinion of the song. The sample suggests that incrementality in interaction concerns not only, as previously reported, phonetic and syntactic aspects of language, but also pragmatic inferences. We will suggest update rules needed to account for the reasoning performed by the speaker J, as well as other cases where an enthymematic argument used in dialogue is less specific than the topos it draws on. We will also suggest a formal definition of what it means for an enthymeme to be underspecified in relation to a topos.

2 Background

2.1 A Dialogue Semantics for Rhetorical Reasoning

In dialogue we frequently draw conclusions which are not, in a strict sense, logical. Following (Breitholtz and Cooper, 2011; Breitholtz, 2011; Breitholtz, 2014), we will use the Aristotelian term *enthymeme* in connection with such inferences. An enthymeme is an argument which appeals to what is in the listener's mind, i.e. an interlocutor must draw on background knowledge or contextual information to correctly interpret the argument. If

a dialogue participant presents the argument P therefore Q, an interlocutor must supply a warrant that P is a valid reason for Q in order for the argument to be successful. These warrants are often referred to as *topoi* (Aristotle, 2007), (Ducrot, 1988), (Ducrot, 1980). When we interact we expect topoi to be common ground, or to be accommodated (adopted by dialogue participants) during the course of the interaction.

The topoi in the resources of an agent may be drawn on to invent and interpret different kinds of enthymemes. Consider for example the dialogue excerpt in (2)

(2)
Anon 3: the monarchy are non political <pause >and therefore, when they choose to speak it's usually out of a genuine concern for that problem
(BNC FLE 233)

In situations such as the one where (2) occurs, the speaker typically assumes that the topos accessed by other conversational participants to interpret the argument, is similar to that which the speaker himself had in mind. However, sometimes our individual takes on the conversation do not match. It is possible that agents involved in dialogue accommodate different topoi which satisfy the criteria for underpinning a particular enthymeme, while not being the ones assumed by the speaker. To model the correspondence and differences between the topoi accessed by conversational participants we use a game board style semantics cast in TTR, similar to analyses found in (Ginzburg, 2012), (Breitholtz and Cooper, 2011), (Cooper and Ginzburg, 2015) (Schlöder et al., 2016). We model enthymematic arguments and the underpinning topoi in the dialogue participants' resources as functions which return types (dependent types). Subtyping is also essential in our account of how topoi may be employed in different enthymemes.

3 Analysis

Let us now return to the example in (1) where

P's first utterance in (1a) – "metal was the reason I started doing hip hop" – is in fact in itself an enthymeme – there is *something* about metal that made Petter start doing hip hop. Thus it may be described as a function from a situation of a type where the music genre "metal" occurs to a type of situation where P starts "doing hip hop", as seen in (3). We refer to this enthymeme as $\mathcal{E}_{reason}$.

(3)
$$\mathcal{E}_{reason} = \lambda r{:}\ \begin{bmatrix} \text{T=music:}\textit{Type} \\ \text{x=metal:T} \\ \text{z=Petter:}\textit{Ind} \\ \text{c}_1\text{:relevant(T)} \end{bmatrix} \cdot \begin{bmatrix} \text{y=hiphop:}r\text{.T} \\ \text{c}_2\text{:do(r.z,y)} \end{bmatrix}$$

There might be several topoi accessible to J which could be drawn on to underpin the enthymeme $\mathcal{E}_{reason}$. Judging from J's utterance she is surprised by P's assertion that he hated metal. We cannot say exactly in which way Petter hating metal is "completely different" from what J expected. However, it seems reasonable to assume that she expected metal being the reason for P starting to "do" hiphop to be due to some favourable relation between him and metal. Thus, a possible topos could be one saying that if two things are of the same type, and the speaker has a favourable attitude to one of them, that thing may cause someone to "do" the other thing. This principle does not follow classical logic, but still seems to be productive in everyday argumentation. Think of examples like "My grandma had poodles, that is what made me start breeding dalmatians", "Karate got me interested in Kung Fu", etc. We see a formalisation of this topos, $\mathcal{T}_{similar}$ in (4).

(4)
$$\mathcal{T}_{similar} = \lambda r{:}\ \begin{bmatrix} \text{T:}\textit{Type} \\ \text{x:T} \\ \text{z:}\textit{Ind} \\ \text{c}_1\text{:relevant(T)} \\ \text{c}_2\text{:like(z, x)} \end{bmatrix} \cdot \begin{bmatrix} \text{y:r.T} \\ \text{c}_3\text{:do(r.z, y)} \end{bmatrix}$$

(Breitholtz, 2014) suggests update rules for integrating topoi on the shared DGB, similar to the one in (5)

5 is a function from a situation of a type where a speaker has access to a topos (in the private field of the DGB) to another function from a type of situation where one such topos is a specification of the max eud, to a situation type where the topos in question is integrated on the shared DGB. This function thus only applies when the domain- or antecedent part of the enthymeme is a subtype (less specific or identical to) of the corresponding part of the topos. Secondly, the result of applying the enthymeme to a record r must be a subtype of the result of applying the topos to the same record.

In the case of $\mathcal{E}_{reason}$ the antecedent type is *not* a subtype of the antecedent type of $\mathcal{T}_{similar}$, since it lacks the constraint c_2: like(z, x). Both requirements for a standard update of shared topos

(5)

$$\mathcal{F}_{integrate\_shared\_topos} = \lambda r{:}\begin{bmatrix}\text{private:}\begin{bmatrix}\text{topoi:list(}\textit{Topos}\text{)}\end{bmatrix}\\ \text{shared:}\begin{bmatrix}\text{eud:list(}\textit{Enthymeme}\text{)}\\ \text{topoi:list(}\textit{Topos}\text{)}\end{bmatrix}\end{bmatrix} \cdot \lambda e{:}\begin{bmatrix}\text{t:}\textit{Topos}\\ \text{c}_1\text{:in(t, r.private.topoi)}\\ \text{c}_2\text{:specification(fst(r.shared.eud), t))}\end{bmatrix} \cdot \begin{bmatrix}\text{shared:}\begin{bmatrix}\text{topoi=[e.t | r.private.topoi]:list(}\textit{Topos}\text{)}\end{bmatrix}\end{bmatrix}$$

is thus not met. However, since dialogue participants sometimes do accommodate topoi based on underspecified enthymemes, we want to be able to model how topoi may be integrated based on less strict requirements. In order to do this we introduce an additional update rule – $\mathcal{F}_{integrate\_topos'}$ – for integrating topoi based on underspecified enthymemes, as seen in (6).

According to $\mathcal{F}_{integrate\_topos'}$ – which is to be applied if there is no topos that is a more general version of the max eud – we may integrate a topos which is more specified than the enthymeme evoking it. We say that an enthymeme $\mathcal{E} = T_3$ is underspecified in relation to a topos $\mathcal{T}$ if $\mathcal{T} = T_1 \cdot T_2$, $\mathcal{E} = T_3 \cdot T_4$, $T_1 \sqsubset T_3$ and, for any r, $\mathcal{E}(r) \sqsubseteq \mathcal{T}(r)$

After the application of $\mathcal{F}_{integrate\_shared\_topos'}$, J's information state is of the type in (7).

(7)

$$\begin{bmatrix}\text{shared:}\begin{bmatrix}\text{eud=[}\mathcal{E}_{metal\_reason}\text{]:list(}\textit{Enthymeme}\text{)}\\ \text{topoi=[}\mathcal{T}_{similar}\text{]:list(}\textit{Topos}\text{)}\\ \text{l-m:}\begin{bmatrix}\text{prev:}\textit{Rec}\\ \text{e:}\begin{bmatrix}\text{x:}\textit{Ind}\\ \text{y:}\textit{Ind}\\ \text{z=Petter:}\textit{Ind}\\ \text{s:}\textit{Ind}\\ \text{c}_1\text{:metal}\\ \text{c}_2\text{:hiphop}\\ \text{c}_3\text{:spec(x, s)}\\ \text{c}_4\text{:spec(y, s)}\\ \text{c}_5\text{:start\_doing}\\ \text{c}_6\text{:reason(z, c}_5\text{, c}_2\text{)}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

After P's second utterance in (1b) – "Because I hated metal" – a new enthymeme, $\mathcal{E}_{reason'}$, is integrated at the top of the list of enthymemes under discussion.

(8) $\mathcal{E}_{reason'} =$

$$\lambda r{:}\begin{bmatrix}\text{T:}\textit{Type}\\ \text{x=metal:T}\\ \text{c}_1\text{:relevant(T)}\\ \text{z=Petter:}\textit{Ind}\\ \text{c}_{hate}\text{:hate(z, x)}\end{bmatrix} \cdot \begin{bmatrix}\text{y=hiphop:}r\text{.T}\\ \text{c}_2\text{:do(}r\text{.z,y)}\end{bmatrix}$$

We need an update rule making sure that shared topoi is updated with a topos which supports the max eud. The rule $\mathcal{F}_{update\_topoi}$ in (9) says that if there is an information state where a topos on shared.topoi supports the max eud, we are licensed to update that information state so that the topos in question is moved to the max topoi position at the top of the list of topoi. If b is a list and $a \in b$, the function μ applied to b, $\mu(a, b)$, moves a to the top of list b regardless of what position a has had previously.

The update rule in (9) applies when a topos which is already integrated on the shared gameboard is being actualised by an enthymeme. However, in cases such as this the topos available seems to be incompatible with the enthymeme: $\mathcal{E}_{reason'}$ says that since Petter hated metal, he started doing hip hop, and the topos $\mathcal{T}_{similar}$ says that if someone likes something s/he might start doing something similar. The antecedents include concepts that we would probably want to model as mutually exclusive, namely *like* and *hate*. The formula in (10) is our version of a meaning postulate, and reads " T_1 precludes T_2", that is there is no situation in which both T_1 and T_2 apply (for a thorough discussion of preclusion in TTR, see (Cooper, in prep).)

(10) If $\begin{bmatrix}\text{x:}\textit{Ind}\\ \text{c:hate(x)}\end{bmatrix} = T_1$ and $\begin{bmatrix}\text{x:}\textit{Ind}\\ \text{c:like(x)}\end{bmatrix} = T_2$ then $T_1 \perp T_2$

When we engage in conversation we normally try to interpret underspecified or implicit content drawing on information already introduced on the dialogue gameboard. This is the case with for example resolution of anaphora. Thus it seems reasonable that an algorithm for applying update rules meant to pick out a topos to underpin the enthymeme currently under discussion first tries to apply the rule $\mathcal{F}_{update\_shared\_topoi}$ which looks for a suitable topos already on the DGB, and not until that fails, apply a rule which looks into the long term memory of the conversational participant (modelled here as private.topoi).

The only topos on the list of shared topoi at the point where J has just integrated $\mathcal{E}_{reason'}$ is such that the max eud cannot be a specification

(6)

$$\mathcal{F}_{integrate\_shared\_topos'}(r) = \lambda r{:}\ \begin{bmatrix}\text{private:}\begin{bmatrix}\text{topoi:list(}\textit{topos}\text{)}\end{bmatrix}\\ \text{shared:}\begin{bmatrix}\text{eud:list(}\textit{Enthymeme}\text{)}\\ \text{topoi:list(}\textit{Topos}\text{)}\end{bmatrix}\end{bmatrix} \cdot$$
$$\lambda e{:}\ \begin{bmatrix}\text{t:}\textit{Topos}\\ \text{c}_1\text{:in(t, r.private.topoi)}\\ \text{c}_2\text{:underspec.(fst(r.shared.eud), t))}\end{bmatrix} \cdot$$
$$\begin{bmatrix}\text{shared:}\begin{bmatrix}\text{topoi=[e.t | r.private.topoi]:list(}\textit{Topos}\text{)}\end{bmatrix}\end{bmatrix}$$

(9)

$$\mathcal{F}_{update\_topoi} = \lambda r{:}\ \begin{bmatrix}\text{shared:}\begin{bmatrix}\text{eud:list(}\textit{Enthymeme}\text{)}\\ \text{topoi:list(}\textit{Topos}\text{)}\end{bmatrix}\end{bmatrix} \cdot$$
$$\lambda e{:}\ \begin{bmatrix}\text{t:}\textit{Topos}\\ \text{c}_1\text{:in(}r\text{.shared.topoi, t)}\\ \text{c}_2\text{:specification(fst(}r\text{.shared.eud), t)}\end{bmatrix} \cdot \begin{bmatrix}\text{shared:}\begin{bmatrix}\text{topoi=[}\mu(e\text{.t, }r\text{.sh.topoi)]:list(}\textit{Topos}\text{)}\end{bmatrix}\end{bmatrix}$$

of it, nor can the topos be a specification of the max eud, since $\mathcal{E}_{reason'} \perp \mathcal{T}_{similar}$. Thus the conditions for applying $\mathcal{F}_{update\_topoi}$ are not fulfilled. So, we move on to once again applying rule $\mathcal{F}_{integrate\_topoi}$. A topos that would work here would be one capturing the notion of " the lesser of two evils", or any other topos saying that dislike of something may cause someone to do some activity of the same type. $\mathcal{T}_{l\_t\_e}$. The point is that in the first assumed topos, the focus is on the *similarity* between two things of the same type, in the second it is on the *dissimilarity*.

(11) $\mathcal{T}_{l\_t\_e} = \lambda r{:}\ \begin{bmatrix}\text{T:}\textit{Type}\\ \text{x:T}\\ \text{y:T}\\ \text{z:}\textit{Ind}\\ \text{c}_1\text{:relevant(T)}\\ \text{c}_2\text{:hate(z, y)}\end{bmatrix} \cdot \begin{bmatrix}\text{e:do(r.z, r.x)}\end{bmatrix}$

We assume thus, that J's information state, T_{IS_J}, when she has integrated $\mathcal{E}_{reason'}$ is of the type in (12).

(12)
$$\begin{bmatrix}\text{priv:}\begin{bmatrix}\text{topoi=[}\mathcal{T}_{l\_t\_e}\text{]:list(}\textit{Topos}\text{)}\end{bmatrix}\\ \text{shar:}\begin{bmatrix}\text{eud=[}\mathcal{E}'_{reason}, \mathcal{E}_{reason}\text{]:list(}\textit{Enthymeme}\text{)}\\ \text{topoi=[}\mathcal{T}_{similar}\text{]:list(}\textit{Topos}\text{)}\end{bmatrix}\\ \end{bmatrix}$$

Since the application of update rule $\mathcal{F}_{update\_shared\_topoi}$ fails in this situation, we move on to once more apply $\mathcal{F}_{integrate\_topoi}$. The resulting type has a max topos that is a specification of the max eud, which is what we would typically expect after integration of a topos on the shared game board.

(13) $\mathcal{F}_{integrat\_topoi}(T_{IS_J}) =$
$$\begin{bmatrix}\text{shared:}\begin{bmatrix}\text{eud=[}\mathcal{E}'_{reason}, \mathcal{E}_{reason}\text{]:list(}\textit{Enthymeme}\text{)}\\ \text{topoi=[}\mathcal{T}_{l\_t\_e}, \mathcal{T}_{similar}\text{]:list(}\textit{Topos}\text{)}\end{bmatrix}\\ \end{bmatrix}$$

4 Conclusion

When a dialogue participant sets about to interpret an enthymematic utterance, they try to access a topos that may serve as underpinning for the enthymeme. typically this means a topos which is a more general than the enthymeme. We looked at an example providing evidence that we may actually start reasoning before an argument is fully spelled out, in the sense that there is a topos that warrants the enthymeme by being a generalised version of it. This indicates that the way we process rhetorical structure is analogous to the way we process sentential and non-sentential utterances as described in e.g. (Eshghi et al., 2015) – by incrementally constraining the search space. We have suggested rules to account for information state updates based on fully specified as well as underspecified enthymemes. In further work we want to investigate to what degree underspecified enthymemes are actually used. Intuitively, in situations where dialogue participants know each other very well and/or the context allows it, they may well infer topoi based on underspecified enthymemes, which turn out to be exactly the ones intended by the speaker. Furthermore the possibility of asking follow up questions and other types of feed back may make it efficient to reason based on underspecified enthymemes in situations where the stakes are not too high.

References

George A. Kennedy Aristotle. 2007. *On Rhetoric, a theory of civic discourse*. Oxford University Press.

Ellen Breitholtz and Robin Cooper. 2011. Enthymemes as rhetorical resources. In Ron Artstein, Mark Core, David DeVault, Kallirroi Georgila, Elsi Kaiser, and Amanda Stent, editors, *Proceedings of*, volume SemDial 2011 (LosAngelogue) Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 149–157.

Ellen Breitholtz. 2011. Enthymemes under discussion: Towards a micro-rhetorical view of dialogue modelling. In Kepa Korta and Maria Ponte, editors, *Proceedings od SPR-11 ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric*, pages 65–69, Donostia, November 9 - 11.

Ellen Breitholtz. 2014. *Enthymemes in Dialogue: A micro-rhetorical approach*. Ph.D. thesis, University of Gothenburg.

Robin Cooper and Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. In Shalom Lappin and Chris Fox, editors, *The Handbook of Contemporary Semantic Theory*, pages 375–407. Wiley-Blackwell.

Robin Cooper. "in prep". *Type theory and language - From perception to linguistic communication*. June. Draft, June 17 2014.

Oswald Ducrot. 1980. *Les échelles argumentatives*.

Oswald Ducrot. 1988. Topoï et formes topique. *Bulletin d'études de la linguistique française*, 22:1–14.

A. Eshghi, C. Howes, E. Gregoromichelaki, J. Hough, and M. Purver. 2015. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics*, London UK.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Julian Schlöder, Ellen Breitholtz, and Raquel Fernández. 2016. Why? In Mandy Simons Julie Hunter and Matthew Stone, editors, *Proceedings of JerSem*, pages 5–14.

Turn-taking with a hidden agenda

Robin Cooper
Centre for Linguistic Theory and Studies in Probability (CLASP)
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg
cooper@ling.gu.se

Abstract

We propose a simple model of turn-taking in an information state based approach to dialogue using TTR (Type Theory with Records). The information state (dialogue gameboard) contains an agenda formulated as a list of speech event types that the dialogue participant plans to realize. A novel aspect of the proposal is that the agenda also includes types of events that intuitively should be carried out by an interlocutor. We argue that all dialogue events should be regarded as events jointly carried out by the dialogue participants and that this yields a simple formal method for representing turn-taking in a formal treatment of dialogue.

1 Perception and types

In the literature on TTR (Type Theory with Records), see Cooper and Ginzburg (2015) for a recent introduction, a connection is made between the notion of judgement in type theory (judging that an object or event is of a certain type) and perception, that is, perception involves classifying something as being of a certain type. We will describe this in this section. As we interact with our environment we not only perceive objects but also create new objects of certain types. Performing an action is creating an event of a particular type. A plan is a list of types which we hope to realize in this way. Thus we obtain a simple theory of action based on type theoretic ideas, which we will describe in Section 2. In Section 3 we will consider how coordinated action can be modelled in terms of games in this framework. We will see in Section 4 that this type theoretical view of action leads naturally to a notion of joint action and that this is important in order to obtain a theory of coordinated action. Finally, in Section 5, we will apply this view of action to turn taking in dialogue.

TTR is a type theory which takes many ideas from Martin-Löf type theory (Martin-Löf, 1984; Nordström et al., 1990). This kind of type theory differs from the version of the simple theory of types that Montague used (Montague, 1973; Montague, 1974) in that it allows for a rich collection of types including types like *Dog* and, following a suggestion by Ranta (1994), types of situations like *A_boy_hug_a_dog* in addition to the kind of types corresponding to basic ontological categories (for example, in Montague's case, types like *Entity* and *Truth_value*) and all types of functions based on the basic types which are introduced in simple type theory. Central to this kind of type theory is the notion of a judgement that an object a is of a type T, in symbols, $a : T$. We will sometimes refer to a as a *witness* for the type T. In the literature on TTR this notion of judgement is connected to a theory of perception. An act of perception involves making such a type judgement. When we perceive something we perceive it as being of a certain type. That is, perceiving an object a as a dog involves making the type judgement $a : \mathit{Dog}$. Similarly perceiving a situation, e, as one in which a boy hugs a dog involves making the type judgement $e : \mathit{A\_boy\_hug\_a\_dog}$. Agents are thought of as having a collection of types available as a resource which they can employ in, among other things, acts of perception. The types available to an agent are in part limited by their perceptual apparatus.

2 Action and types

A judgement can be thought of as an action which an agent carries out, for example, when an object is presented to its perceptual apparatus. Cooper (2014; Cooper (in prep) calls it a kind of *type act*

(meant as a parallel to speech act) and presents a simple theory of action based on types. Basically, there are three things that you can do with types: (i) judge an object to be of a type (ii) wonder whether an object is of a type (iii) create a new object of a type. The third of these is important for this paper. Since we have types of situations (including events) we can regard actions as involving the creation of a situation of a certain type. Consider a particular boy, b and a particular dog, d. The type of situation in which b hugs d can be represented in TTR as a *ptype* (a type constructed from a predicate and appropriate arguments), 'hug(b,d)'. It is no accident that the notation for this situation type is the same as that for a logical formula corresponding to a proposition. In this kind of type theory, types can serve as propositions (the "propositions as types" slogan[1]). When considered as propositions, they are true if there is something of the type and false if there is nothing of the type. Thus if b creates a situation of the type 'hug(b,d)', then b has guaranteed that the type is "true".

3 Coordinated action and games

Let us consider a slightly more complicated kind of situation. Suppose we have a human and a dog playing the game of fetch, where the human picks up a stick and throws it, the dog runs after the stick and brings it back to the human. This involves a string of events[2] which could be regarded as witnesses for ptypes in TTR. If $T_1, \ldots, T_n$ are types, then $T_1 ^\frown \ldots ^\frown T_n$ is a type whose witnesses are strings $a_1 \ldots a_n$ such that $a_i : T_i$ for $1 \leq i \leq n$. Thus a game of fetch between a human, a, and a dog, b, involving a stick, c could be characterized as having the type:

pick_up(a,c)$^\frown$attract_attention(a,b)$^\frown$throw(a,c)$^\frown$
run_after(b,c)$^\frown$pick_up(b,c)$^\frown$return(b,c,a)

For technical reasons that will become apparent below we will use record types containing these ptypes instead of the simple ptypes:

$\left[\text{e:pick\_up}(a,c)\right]^\frown\left[\text{e:attract\_attention}(a,b)\right]^\frown$
$\left[\text{e:throw}(a,c)\right]^\frown\left[\text{e:run\_after}(b,c)\right]^\frown$
$\left[\text{e:pick\_up}(b,c)\right]^\frown\left[\text{e:return}(b,c,a)\right]$

This gives us a label 'e' which we can use as a pointer to pick out the individual subevents. A record, $\left[\text{e=s}\right]$, would be of type $\left[\text{e:pick\_up}(a,c)\right]$ just in case s:pick_up(a,c). We can think of a record as modelling a situation with one or more facts which hold in it. Witnesses for record types may contain more facts than those required by the type. Thus $\begin{bmatrix}\text{e} = s\\ \text{e}' = s'\end{bmatrix}$ would also be a witness of this record type just in case the object, s, in the field labelled by 'e' is of the appropriate type. Fields with labels not mentioned in the type are ignored.

One aspect of coordination between the human and the dog is that they both realize that the game they are playing has this type. For each type in the string an event of that type has to be created and this has to be carried out in the appropriate order, of course. One way to do this is to think of the rules of the game as a collection of update functions addressing an agenda in the agents' information state. An agenda is a list of types (that is, it is of type '[*Type*]') which the agent plans to realize in order. An update function can come in one of two forms. The first form will map an information state of a given type to a new type which can then be used to compute a type for the new information state. The second form will map an information state of a given type and an event of a given type to a new type which can be used to compute a type for the new information state. The type of information state we are using here is $\left[\text{agenda:[}\textit{RecType}\text{]}\right]$, that is the type of records which have a field labelled 'agenda' which contains a list of record types. (*RecType* is the type of record types and [*RecType*] is the type of lists of record types.) We can restrict the type of information states to be one where the agenda is required to be some specific list, L, by using a manifest field: $\left[\text{agenda=}L\text{:[}\textit{RecType}\text{]}\right]$, the type of information states whose 'agenda'-field contains the list, L.

The two forms of update function are illustrated with respect to the fetch game below.

$\lambda r{:}\left[\text{agenda=[]:[}\textit{RecType}\text{]}\right]$.
$\left[\text{agenda=[}\left[\text{e:pick\_up}(a,c)\right]\text{]:[}\textit{RecType}\text{]}\right]$

This function maps a state with an empty agenda to the type of states where the agenda contains a sole member the type of situation where a picks up c.

$\lambda r{:}\left[\text{agenda=[}\left[\text{e:pick\_up}(a,c)\right]\text{]:[}\textit{RecType}\text{]}\right]$.

[1] See Wadler (2015) for an account of the origins of this slogan from the perspective of computer science and Ranta (1994) for a discussion of its relevance for linguistic semantics

[2] The idea of events as strings come from work by Tim Fernando, most recently presented in Fernando (2015).

λe: $\left[\text{e:pick\_up}(a,c)\right]$.

$\left[\text{agenda=}[\left[\text{e:attract\_attention}(a,b)\right]]:[\textit{RecType}]\right]$

This function maps a state with the event type "a picks up c" on the agenda and an event where a picks up c to the type of state which has the type "a attracts b's attention" on the agenda. Such functions can be used by an agent to predict what type of information state could be licensed on the basis of the agent's current information state and, in the case of the second function, also an external event of a given type. The idea is that, if f is such a function of type $T_i \rightarrow$ *RecType* (or $T_i \rightarrow T_e \rightarrow$ *RecType*) and $r : T_i'$ is the current information state where T_i' is a subtype of T_i (and also $e : T_e'$, where T_e' is a subtype of T_e), the type of the next information state is licensed to be $T_i' \boxed{\wedge\!\!\!\cdot} f(r)$ (or $T_i' \boxed{\wedge\!\!\!\cdot} f(r)(e)$). '$\boxed{\wedge\!\!\!\cdot}$' is the operation of *asymmetric merge* (Cooper and Ginzburg, 2015; Cooper, in prep). Basically if one of T_1, T_2 is not a record type then $T_1 \boxed{\wedge\!\!\!\cdot} T_2 = T_2$. If T_1, T_2 are both record types, then for labels they do not have in common, $T_1 \boxed{\wedge\!\!\!\cdot} T_2$ will contain both the fields from T_1 and T_2. For labels, ℓ, they do have in common, $T_1 \boxed{\wedge\!\!\!\cdot} T_2$ will contain a field labelled ℓ with the asymmetric merge of the two types in that field in T_1 and T_2. Asymmetric merge corresponds to the notion of priority unification in the feature-based grammar literature (Shieber, 1986). For example, the asymmetric merge of

$$\begin{bmatrix}\text{agenda=}[\left[\text{e:pick\_up}(a,c)\right]]:[\textit{RecType}]\\ \text{other-info:}T\end{bmatrix}$$

with

$\left[\text{agenda=}[\left[\text{e:attract\_attention}(a,b)\right]]:[\textit{RecType}]\right]$

is

$$\begin{bmatrix}\text{agenda=}[\left[\text{e:attract\_attention}(a,b)\right]]:[\textit{RecType}]\\ \text{other-info:}T\end{bmatrix}$$

An important word in the characterization of update above is *licensed*. Actions are licensed by previous events of the appropriate type as specified by the game. There is, of course, no necessary inference that such an action will occur or even that the type will appear on anybody's agenda. We can at any point decide to stop playing the game. What we can infer is that if we stop in the middle we will not have completed the game and that certain actions are necessary if we are to create an instance of the game type we have in mind. In this way the kind of inferencing that is involved here is enthymematic in the sense of Breitholtz (2014a), Breitholtz (2014b).

4 Joint action to achieve coordination

In Section 3 we have said something about what it might mean for agents to be coordinated on the type of the game they are playing and how they might update their agendas on the basis of previous events considered as events in a particular instance of the game. But we have said nothing about which event types go on which agent's agenda. At first blush it seems there is a clear division of duties between the human and the dog in the game of fetch. The human has to pick up the stick, attract the dog's attention and throw the stick. The dog has to run after it and bring it back to the human. Therefore it might appear that the first three types should, at the appropriate point in the game, appear on the human's agenda and the other two types, again at an appropriate point in the game, appear on the dog's agenda.

But let us think about this a little more carefully before we develop a formal treatment which involves the different types arriving on the appropriate agenda. Suppose the human picks up the stick and tries to realize the event type of attracting the dog's attention. But the dog is facing the other way gnawing on a bone. The human perhaps calls to the dog but gets no response. Perhaps the human walks around the dog so that she is in the dog's line of sight. The dog turns around taking the bone and faces away from the human. The game cannot continue. The dog has to make a contribution to the realization of the type 'attract_attention(a,b)', look at the stick, and look excited, bark or jump up and down or something. The agent who realizes the type is not just the intuitive "first argument" to the predicate. The dog has to give some kind of feed-back that it is up for the game.

Consider another scenario, a little further on in the game. The stick has been thrown and the dog has run after it and has it in its mouth but then discovers that the human has disappeared. How can the dog realize the type 'return(b,c,a)' if a has wandered off somewhere and is nowhere to be seen? No, the human has to contribute to the realization of this type by at least staying close enough to the dog and in the dog's line of sight when it turns round, quite possibly also by encouraging the dog and looking like she expects the stick to be brought back to her.

These actions are joint actions in the sense of Clark (1996), even if one of the agents is active and the other is fairly passive. Realizing the situ-

ation types in a game for two agents is not something that you can do on your own. The technical conclusion I would draw from this is that the types associated with the game are entered onto both agents' agendas as the appropriate juncture in the game as specified by the update functions. Then even if there are types where you don't have to make any kind of active contribution to realizing the type at least there will be a mechanism for causing you to wait until the type on the agenda has been realized before moving on and updating the agenda with a new type. This is known as *turn taking*.

5 Joint action and turn taking in dialogue

I have dwelt at length on the non-linguistic example of the game of fetch because I believe that the basic strategies of coordination, including turn-taking, in dialogue are really the same strategies needed by collaborating agents even without language. The event types involved are very different, involving types of speech events which on an evolutionary scale are extremely specialized and even arcane, but I would like to suggest that the basic turn-taking mechanism which enables coordination in speech is built on the kind of cognitive abilities and strategies necessary for coordinating agents independent of whether they have language or not. This is one reason that it seems important to embed a formal theory of language in a general formal theory of action.

In the literature involving gameboards of the kind Ginzburg has proposed (Ginzburg, 2012) there has not be a great deal of emphasis on getting turn taking to work out. For those of us working with agendas in this kind of framework following (Larsson, 2002), there has been the general assumption that what goes on the agenda are types of events in which the agent is the main actor. Thus, for example, if agent A asks a question of agent B, then the type of the question event first goes on A's agenda and this licenses A to realize an event of this type, that is, ask the question. B, on hearing A's utterance of the question, plans to answer the question, that is, puts a type on B's agenda which is the type of an answer to the question. (This assumes that A and B are playing a straightforward question-answer game rather than something more complicated like a clarification or rejection of the question.) At the point at which B is in this state and utters an answer, A's agenda is empty. There is nothing in such a formal account which represents the fact that when you ask a question you are supposed to wait an appropriate amount of time for an answer and be collaborative. That is, in a normal question-answer exchange you are not supposed to ask a question and then walk out of the room or sing at the top of your voice so that you cannot hear the answer. It seems that the kind of coordination that is required here is exactly like that required between the dog and the human when the dog is picking up the stick. Just like the "passive" role that the human has to play in the returning of the stick, a questioner has a role to play in realizing the type where the question is answered, namely by showing that they are ready to receive the answer. Both A and B should have a type of the answering event on the agenda and jointly play their respective roles in realizing the type. Note that it need not be the case that A and B have exactly the same type on the agenda. For example, A will have a type for a situation where B answers the question. It may be that, at least at some point, before actually answering the question, B has a subtype of A's type on the agenda, namely one that in addition specifies a content for the answer. That is, it is A's job to facilitate an answer, whatever it is. It is B's job to give some particular answer to the question. This shows that a notion of A and B being coordinated does not necessarily involve having the *same* types on their respective agendas. But perhaps what counts as coordination is that the respective types stand in the subtype relation and perhaps one could even claim that the type that the main actor of the event has must be a subtype of the type that the "supporting" actor has. If it is the other way around then perhaps the supporting actor was expecting something of the main actor that they didn't do in the end – a sign of miscoordination.

6 Conclusion

We have suggested a simple notion of turn taking as a kind of coordination between information states in agents both in linguistic and non-linguistic games and we have emphasized the importance of embedding a formal theory of language in a formal theory of action. It seems on this view that almost any speech act is a kind of joint action, albeit in many cases with a "leading" actor and a "supporting" actor.

References

Ellen Breitholtz. 2014a. *Enthymemes in Dialogue: A mico-rhetorical approach*. Ph.D. thesis, University of Gothenburg.

Ellen Breitholtz. 2014b. Reasoning with topoi – towards a rhetorical approach to non-monotonicity. In *Proceedings of AISB Symposium on "Questions, discourse and dialogue: 20 years after Making it Explicit"*.

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

Robin Cooper and Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. In Lappin and Fox (Lappin and Fox, 2015), pages 375–407.

Robin Cooper. 2014. How to do things with types. In Valeria de Paiva, Walther Neuper, Pedro Quaresma, Christian Retoré, Lawrence S. Moss, and Jordi Saludes, editors, *Joint Proceedings of the Second Workshop on Natural Language and Computer Science (NLCS 2014) & 1st International Workshop on Natural Language Services for Reasoners (NLSR 2014) July 17-18, 2014 Vienna, Austria*, pages 149–158. Center for Informatics and Systems of the University of Coimbra.

Robin Cooper. in prep. Type theory and language: from perception to linguistic communication. Draft of book chapters available from `https://sites.google.com/site/typetheorywithrecords/drafts`.

Tim Fernando. 2015. The Semantics of Tense and Aspect: A Finite-State Perspective. In Lappin and Fox (Lappin and Fox, 2015).

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford.

Shalom Lappin and Chris Fox, editors. 2015. *The Handbook of Contemporary Semantic Theory*. In Lappin and Fox (Lappin and Fox, 2015), second edition.

Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, University of Gothenburg.

Per Martin-Löf. 1984. *Intuitionistic Type Theory*. Bibliopolis, Naples.

Richard Montague. 1973. The Proper Treatment of Quantification in Ordinary English. In Jaakko Hintikka, Julius Moravcsik, and Patrick Suppes, editors, *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, pages 247–270. D. Reidel Publishing Company, Dordrecht.

Richard Montague. 1974. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven. ed. and with an introduction by Richmond H. Thomason.

Bengt Nordström, Kent Petersson, and Jan M. Smith. 1990. *Programming in Martin-Löf's Type Theory*, volume 7 of *International Series of Monographs on Computer Science*. Clarendon Press, Oxford.

Aarne Ranta. 1994. *Type-Theoretical Grammar*. Clarendon Press, Oxford.

Stuart Shieber. 1986. *An Introduction to Unification-Based Approaches to Grammar*. CSLI Publications, Stanford.

Philip Wadler. 2015. Propositions as Types. *Communications of the ACM*, 58(12):75–84.

KILLE: learning grounded language through interaction

Simon Dobnik and Erik Wouter de Graaf
Dept. of Philosophy, Linguistics & Theory of Science
University of Gothenburg, Sweden
simon.dobnik@gu.se and kille@masx.nl

Abstract

Testing and computational implementation of formal models of situated linguistic interaction imposes demands on computational infrastructure. We present our system called KILLE and provide a proof-of-concept evaluation of interactive situated learning of object categories and spatial relations.

1 Grounded meaning in interaction

Contemporary approaches to semantics of natural language (Cooper, 2016; Fernández et al., 2011) are based on two important premises: (i) meanings are not universal and static but are agent-relative and are continuously adapted in interaction with other agents and environment (Clark, 1996; Pickering and Garrod, 2004); and (ii) meanings (sense and reference) are multi-modal where different lexical items are sensitive to different modalities in different contexts to different degrees (Coventry and Garrod, 2005).

Both aspects have changed the focus in computational semantics from engineering formal rules that cover a domain or a fragment of linguistic data off-line to approaches that are data driven and involve continuous online fine-tuning of the model's parameters (Skočaj et al., 2011; Matuszek et al., 2012). In robotics a shift in the approach has happened much earlier as it quickly became apparent that robots with static models cannot deal with any changes in the environment or with the environment's uncertainty. Instead, modern robotics uses models which are learned from data and refined continuously as the robot's interaction with the environment develops (for example (Dissanayake et al., 2001) for map building). We argue that the same paradigm should also be adopted when dealing with computational models of language. In this view the focus of building a computational system is not on designing representations but investigating and modelling interactive strategies or dialogue games (Kowtko et al., 1992) that will allow construction of such representations or fine-tuning of their features, depending on how much of representations are pre-available to such a system.[1]

The interactive semantics of a computational system have also implications on the models of meaning used. The predominant semantic representations used in computational semantics today are vector-space representations that define meaning as semantic similarity between lexical items on the basis of their co-occurrence in contexts (Turney et al., 2010; Clark, 2015). Such models can be successfully extracted from large corpora of text and are very successful in representing meaning. However, they nonetheless represent meaning in an indirect way as they never consider a relation between an expression and situations in which that expression applies to or is true for. The reason why words in particular linguistic contexts are lexically similar is because words in linguistic strings as a whole refer to (more or less) the same situations which we do not have access to or ignore when we built vector space models. However, in an interactive scenario described above we can explore linking linguistic expressions and perceptual features directly, a process which is commonly known as grounding (Harnad, 1990; Roy, 2002). Such models are required for situated dialogue agents or conversational robots which have to link language and situations that they jointly attend to with human conversational partners.[2]

[1] This sounds similar to the Chomsky's innateness claim but here we are thinking of purely engineering a system and make no claims about human cognition.

[2] It is important to emphasise nonetheless that vector space models may provide an important source of back-

Grounded meanings of linguistic descriptions such as "close to the table" and "red" correspond to some function from physical or colour space to a degree of acceptability of that description (Logan and Sadler, 1996; Roy, 2002; Skočaj et al., 2011; Matuszek et al., 2012; Kennington and Schlangen, 2015; McMahan and Stone, 2015). Cognitive structures are hierarchically organised at several representation layers focusing on and combining different modalities (Kruijff et al., 2007). Since the functions predict distributions of degree of applicability several descriptions may be equally applicable for the same perceptual situation: the chair can be "close to the table" or "to the left of the table" which means *vagueness* is prevalent in grounding. This however, can be resolved through interaction by adopting appropriate interaction strategies (Kelleher et al., 2005; Skantze et al., 2014; Dobnik et al., 2015).

A formal model of perceptual semantics in interaction has been the focus of Type Theory with Records (TTR) (Cooper, 2016; Larsson, 2013; Dobnik et al., 2013). Implementing, validating and testing such models imposes complex demands on computational infrastructure in the sense that this involves connecting perceptual sensors with dialogue systems and machine learning algorithms. Processing language in interaction also presents challenges from the computational perspective as it is often not trivial to employ existing language technology tools and (machine learning) algorithms, which were developed for processing data offline, in an interactive tutoring scenario. To address both issues we have developed a framework for situated agents that learn grounded language incrementally and online with a help of human tutor called KILLE[3] (Kinect Is Learning LanguagE). This paper focuses on the construction of the Kille framework and its properties while it also provides a proof-of-concept evaluation of such learning of simple object and spatial relations representations. We hope that this framework will be a useful tool for future studying and computational modelling language in interaction.

ground knowledge in such scenario and hence a dialogue agent does not have to learn every meaning representation through grounding. The challenges of integration of both meaning representations are a focus of ongoing research.

[3]Swedish for "fellow", "chap" or "bloke".

2 The KILLE system

KILLE is a non-mobile table-top robot connecting Kinect sensors with image processing (*libfreenect*), classification (clustering of visual features and location classification) and a spoken dialogue system OpenDial[4] (Lison, 2013) connected through Robot Operating System (ROS) (Quigley et al., 2009). The latter is a popular robotic middle-ware which ensures communication between them. It runs on a variety of popular robotic hardware implementations which means that our system could be ported to them without too much modification (Figure 1). We prefer a robotic middle-ware rather systems centred around dialogue systems because it allows us to represent and exchange perceptual and linguistic information together and in the same way: there is one information state for both. In addition to the integration of these modules, our main contribution is implementation of ROSDial which provides and interface between OpenDial and ROS, implementation of Kille Core which provides perceptual and spatial classification, and implementation of dialogue games that interface between dialogue and perceptual classification and therefore enable incremental perceptual learning.

The system learns to recognise objects presented to it by a human tutor from scratch. It can direct learning by asking for more objects of a particular category if it is not able to classify them with sufficient reliability, thus filling in the missing knowledge. If more objects of a particular category are available in the scene and the system is able to recognise them, the system queries the user to describe spatial relations between them. Each of these kinds of descriptions focus on a different perceptual features (visual features vs location) and represent two fundamental linguistic semantic categories: entities and entity relations. Overall, KILLE combines both passive (being instructed) and active (requesting examples to learn from) learning which is incremental in terms of learning to recognise situations denoted by the lexical categories. A demo of the system can be found at `http://www.youtube.com/watch?v=AsEgaka6tH0`.

Below there are some examples of dialogue games that we implemented as dialogue rules in the dialogue manager which act upon both linguistic and perceptual knowledge.

[4]https://github.com/plison/opendial

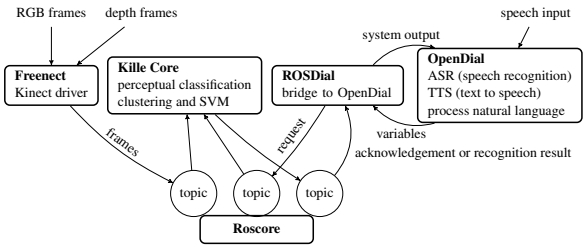


Figure 1: Kille modules

1. **Direct instruction: objects**

   User: This is a cup.
   System: [Object is focused on and observed.]
   S: OK, I learned a cup.

2. **Describing objects and relations with (un)certainty**

   U: What is this?
   S: [Object is focused on and observed.]
   U: The object is thought to be a book, but it might also be a mug.

3. **Reinforcement from feedback**

   U: What is this?
   S: [Object is focused on and observed.]
   S: This is a cup.
   U: That is correct.
   S: [Observation is added to the models.]

4. **Self initiative: spatial relations**

   S: [Objects are focused on and observed.]
   S: One of the objects is book, the other one is box. Please tell me where things are.
   U: The book is to the left of the box.
   S: OK, this relation is called "to the left of".

For visual representations we use Open Source Computer Vision (OpenCV)[5] (Bradski and Kaehler, 2008) which is a popular library for computer vision including real-time machine learning applications. Through ROS we receive real-time frames from Kinect which includes both data from the depth sensor and the visual RGB sensor. We use the depth information (which gives us a precise 3d location of points making up an object) to detect the object in focus and later take the pixels representing these points in focus to detect SIFT features (Scale-Invariant Feature Transform) (Lowe, 1999) over them which are used to represent objects in our model as shown in Figure 2.

Objects, including those that are very similar and belong to the same category, have different number of SIFT descriptors detected depending on

[5]http://opencv.org

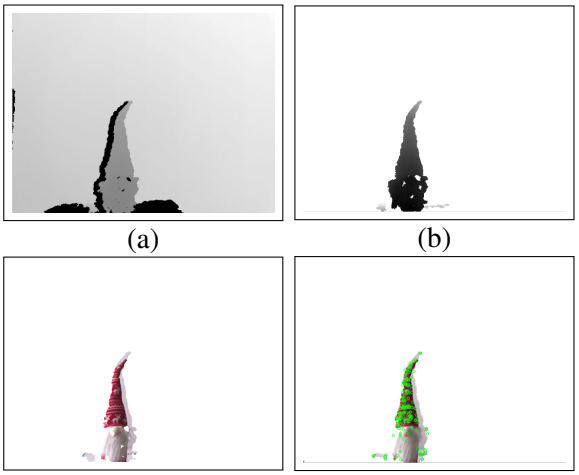


Figure 2: A perception of a plush gnome from the depth sensor (a) including the background, (b) with the background removed, (c) with the RGB image superimposed, and (d) with SIFT features detected in the image. The black border in (a) is a perceptual artefact arising from the interference of sensors.

their visual properties: some objects have more visual details than others. There is a bias that object with less features match objects with more (and similar looking) features. In our interactive scenario there is also no guarantee that the same features will be detected after the object is re-introduced (or even between two successive scans) as the captured frame will be slightly different from the previously captured one because of slight changes in location, lighting and camera noise.

3 Interactive perceptional learning

In the following subsections we present a proof-of concept implementation and evaluation of perceptual learning through interaction which demonstrates the usability of the Kille framework.

Learning to recognise objects To recognise objects we developed a nearest neighbour classification method based on the the FLANN library (Muja and Lowe, 2009) which works by comparing the SIFT descriptors of object to classify with the objects in the database and then returns the class of the closest matching object. In the evaluation, 10 consecutive scans are taken and their recognition scores are averaged to a single score. This improves the accuracy but increases the classification time (which is nonetheless still reasonable for the small domain of objects we are con-

sidering). The location of the recognised object is estimated by taking the locations of the twenty matched descriptors with the shortest distance.

To evaluate the system's performance in an interactive tutoring scenario we chose the following 10 objects: apple, banana, teddy bear, book, cap, car, cup, can of paint, shoe and shoe-box. A human tutor successively re-introduces the same 10 objects to the system in a pre-defined order over four rounds trying to keep the presentation identical as much as possible. In each round all objects are first learned and then queried. To avoid ASR errors both in learning and generation text input is used.

Taking the average SIFT feature matching scores over 4 rounds for each object and taking the class of the object with highest mean score, on average all but one object were recognised correctly. However, the cap was consistently confused with the banana. There were a couple of individual confusions that have been levelled out in the calculation of the average score. To test how distinct objects are from one another we calculated a difference of the matching scores of the highest-ranking object of the correct category and the other highest ranking candidate. If we arrange objects by this score, we get the following ranking (from more distinct to least distinct): book $>$ car $>$ shoe $>$ cup $>$ banana $>$ bear $>$ apple $>$ paint $>$ shoe-box $>$ cap. We also tested recognition of the same objects when rotated and recognition of new objects of the same category.

Learning to recognise spatial relations Before spatial relations can be learned the system must recognise the target and the landmark objects ("the gnome/TARGET is to the left of the book/LANDMARK") both in a linguistic string and in a perceptual scene. Twenty highest ranking SIFT features are taken for each object and their x (width), y (height) and z (depth) coordinates are averaged, thus giving us the centroid of the 20 most salient features of an object. The coordinate frame of the coordinates is transposed to the centre of the landmark object. The relativised location of the target to the landmark are fed to a Linear Support Vector Classifier (SVC) with descriptions as target classes.

A human tutor taught the system by presenting it the target object (a book) randomly 3 times at 16 different locations (2 distances/circles containing 8 points separated at 45°) in relation to the landmark (the car). The spatial descriptions that the human instructor used were *to the left of*, *to the right of*, *in front of*, *behind of*, *near* and *close to* (6). The performance of the system was evaluated by two human conversational partners, one of whom was also the tutor from the learning stage. The target object was randomly placed in one of the 16 locations and each location was used twice which gave us 32 generations. A particular location may be described with several spatial descriptions (but not all combinations of descriptions are possible) but some may be more appropriate than others. The evaluators first wrote down a description they would use to describe the scene and then the system would be queried about the location of the target to which it provided a response. The evaluators would then also record whether they agree with the generation. The observed blind agreement between the evaluators is 0.5313 with $\kappa = 0.4313$ which means that choosing a spatial description is quite a subjective task. The blind agreement between the evaluators and the system is 0.2344 with $\kappa = 0.0537$. The evaluators were happy with the system's generation in additional 37.5% of cases, which means that the system generated an appropriate description in 60.94% of cases which is encouraging and comparable to the similar task in the literature. Note also that the system tried to learn continuous functions from a very small number of examples, on an average only 46/6=8 instances.

4 Conclusion and future work

In this paper we argue that there is a need for a computational infrastructure that will allow us modelling dynamic grounded semantics in interaction for two reasons: (i) to verify semantic theories and (ii) to provide a platform for their computational implementations. We developed and framework called KILLE a simple interactive "robot" which we argue provides a good solution for modelling these aspects and at the same time can be ported to more sophisticated robotic hardware platforms. We demonstrated a proof-of-concept of learning object categories and spatial relations following the theoretical proposals in the literature. We hope that the platform will provide useful for testing further models of linguistic and perceptual interactions.

References

Gary Bradski and Adrian Kaehler. 2008. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.

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

Stephen Clark. 2015. Vector space models of lexical meaning. In Shalom Lappin and Chris Fox, editors, *Handbook of Contemporary Semantics — second edition*, chapter 16, pages 493–522. Wiley – Blackwell.

Robin Cooper. 2016. Type theory and language: From perception to linguistic communication. Draft of chapters 1-6, 30th November.

Kenny Coventry and Simon Garrod. 2005. Spatial prepositions and the functional geometric framework. towards a classification of extra-geometric influences. In Laura Anne Carlson and Emile van der Zee, editors, *Functional features in language and space: insights from perception, categorization, and development*, volume 2, pages 149–162. OUP.

M. W. M. G Dissanayake, P. M. Newman, H. F. Durrant-Whyte, S. Clark, and M. Csorba. 2001. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotic and Automation*, 17(3):229–241.

Simon Dobnik, Robin Cooper, and Staffan Larsson. 2013. Modelling language, action, and perception in Type Theory with Records. In Denys Duchier and Yannick Parmentier, editors, *Constraint Solving and Language Processing (CSLP 2012), Revised Selected Papers*, v8114 of *LNCS*, pages 70–91. Springer Berlin Heidelberg.

Simon Dobnik, Christine Howes, and John D. Kelleher. 2015. Changing perspective: Local alignment of reference frames in dialogue. In *Proceedings of goDIAL - Semdial 2015*, pages 24–32, Gothenburg, Sweden, 24–26th August.

Raquel Fernández, Staffan Larsson, Robin Cooper, Jonathan Ginzburg, and David Schlangen. 2011. Reciprocal learning via dialogue interaction: Challenges and prospects. In *Proceedings of the IJCAI 2011 ALIHT Workshop*, Barcelona, Catalonia, Spain.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D*, 42(1–3):335–346, June.

J.D. Kelleher, F. Costello, and J. van Genabith. 2005. Dynamically structuring updating and interrelating representations of visual and linguistic discourse. *Artificial Intelligence*, 167:62–102.

Casey Kennington and David Schlangen. 2015. Simple learning and compositional application of perceptually grounded word meanings for incremental reference resolution. In *ACL-IJCNLP 2015*, pages 292–301, Beijing, China, July. ACL.

Jacqueline C Kowtko, Stephen D Isard, and Gwyneth M Doherty. 1992. Conversational games within dialogue. HCRC research paper RP-31, University of Edinburgh.

Geert-Jan M. Kruijff, Hendrik Zender, Patric Jensfelt, and Henrik I. Christensen. 2007. Situated dialogue and spatial organization: what, where... and why? *International Journal of Advanced Robotic Systems*, 4(1):125–138. Special issue on human and robot interactive communication.

Staffan Larsson. 2013. Formal semantics for perceptual classification. *Journal of Logic and Computation*, online:1–35, December 18.

Pierre Lison. 2013. *Structured Probabilistic Modelling for Dialogue Management*. Ph.D. thesis, Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Oslo, 30th October.

Gordon D. Logan and Daniel D. Sadler. 1996. A computational analysis of the apprehension of spatial relations. In Paul Bloom, Mary A. Peterson, Lynn Nadel, and Merrill F. Garrett, editors, *Language and Space*, pages 493–530. MIT Press, Cambridge, MA.

David G Lowe. 1999. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. IEEE.

Cynthia Matuszek, Nicholas FitzGerald, Luke Zettlemoyer, Liefeng Bo, and Dieter Fox. 2012. A joint model of language and perception for grounded attribute learning. In *Proceedings of ICML 2012*, Edinburgh, Scotland, June 27th - July 3rd.

Brian McMahan and Matthew Stone. 2015. A Bayesian model of grounded color semantics. *Transactions of the ACL*, 3:103–115.

Marius Muja and David G Lowe. 2009. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (1)*, 2(331–340):2.

Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.

Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. 2009. ROS: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5.

Deb K. Roy. 2002. Learning visually-grounded words and syntax for a scene description task. *Computer speech and language*, 16(3):353–385.

Gabriel Skantze, Anna Hjalmarsson, and Catharine Oertel. 2014. Turn-taking, feedback and joint attention in situated human-robot interaction. *Speech Communication*, 65:50–66.

Danijel Skočaj, Matej Kristan, Alen Vrečko, Marko Mahnič, Miroslav Janíček, Geert-Jan M. Kruijff, Marc Hanheide, Nick Hawes, Thomas Keller, Michael Zillich, and Kai Zhou. 2011. A system for interactive learning in dialogue with a tutor. In *IROS 2011*, San Francisco, CA, USA, 25-30 September.

Peter D Turney, Patrick Pantel, et al. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1):141–188.

Interactional Dynamics and the Emergence of Language Games

Arash Eshghi
Interaction Lab
Heriot-Watt University
eshghi.a@gmail.com

Igor Shalyminov
Interaction Lab
Heriot-Watt University
o.lemon@hw.ac.uk

Oliver Lemon
Interaction Lab
Heriot-Watt University
o.lemon@hw.ac.uk

Abstract

Meaning is highly activity-specific, in that the action that a particular sequence of words is taken to perform is severely underdetermined in the absence of an overarching activity, or a 'language-game'. In this paper, we combine a formal, incremental model of interactional dynamics and contextual update - Dynamic Syntax and Type Theory with Records (DS-TTR) - with Reinforcement Learning for word selection. We show, using an implemented system, that trial and error generation with a DS-TTR lexicon – a process we have dubbed `babbling` – leads to particular domain-specific dialogue acts to be learned and routinised over time; and thus that higher level dialogue structures - or how actions fit together to form a coherent whole - can be learned in this fashion. This method therefore allows incremental dialogue systems to be automatically bootstrapped from small amounts of unannotated dialogue transcripts, yet capturing a combinatorially large number of interactional variations. Even when the system is trained from only a single dialogue, we show that it supports over 8000 new dialogues in the same domain. This generalisation property results from the structural knowledge and constraints present within the grammar, and highlights limitations of recent state-of-the-art systems that are built using machine learning techniques only.

1 Introduction

Meaning is highly activity-specific, in that the action that a particular sequence of words is taken to perform, together with any perlocutionary effect that action might give rise to, is severely underdetermined in the absence of a particular overarching activity, or a 'language-game'. Wittgenstein famously argued that the structure of a language-game, or how actions fit together to form a coherent whole, is irreducible. Arguably, this is the most unyielding obstacle facing not only theoretical approaches to pragmatics, but also dialogue system developers today. This suggests that particular dialogue structures are emergent, learned, and very frequently adjusted during interaction (Mills and Gregoromichelaki, 2010; Mills, 2011; Healey, 2008; Larsson and Cooper, 2008).

Despite this, recent and ongoing work in formal dialogue modelling suggests that not only language processing mechanisms, but also certain basic principles of contextual dynamics in dialogue do generalise across domains (Ginzburg, 2012; Kempson et al., 2016; Eshghi et al., 2015; Kempson et al., 2015; Purver et al., 2010). Even in a simple domain, there's a lot of *interactional variation* that does not ultimately affect the overall communicative goal of a dialogue. For example, the dialogues in Fig. 1 (specifically the top two rows, where the lexicon is held constant) all lead to a context in which the user wants to buy a phone by LG. These dialogues can be said to be *pragmatically synonymous* for this domain. Arguably, a good model of interactional dynamics should be able to capture this synonymy.

In this paper, we show, using an implemented system (Kalatzis et al., 2016), that given Dynamic Syntax and Type Theory with Records (DS-TTR) (Kempson et al., 2001; Eshghi et al., 2012; Eshghi et al., 2015) as a low-level, incremental model of interactional and contextual dynamics, one can see dialogue acts, together with their associated local dialogue structures and procedural conventions as emergent and learned from interaction; and thus

Interactional	USR: I would like an LG laperr, phone SYS: okay.	USR: I would like a phone by LG. SYS: sorry a what? USR: a phone by LG. SYS: okay.	SYS: what would you like? USR: an LG phone SYS: okay.
	SYS: what would you like? USR: a phone SYS: by which brand? USR: LG SYS: okay	SYS: you'd like a ...? USR: a phone SYS: by what brand? USR: LG. SYS: okay	SYS: so would you like a computer? USR: no, a phone. SYS: okay. by which brand? USR: LG. SYS: okay.
Lexical	USR: I want an LG phone. SYS: okay.	USR: Can I have an LG phone? SYS: Sure.	SYS: What do you want to buy? USR: a phone SYS: by which make? USR: LG SYS: Okay.

Figure 1: Some Interactional and Lexical Variations in a Shopping Domain

that fully incremental dialogue systems can be bootstrapped from raw, unannotated example successful dialogues within a particular domain.

The model we present below combines DS-TTR with Reinforcement Learning for incremental word selection, where dialogue management and language generation are treated as one and the same decision/optimisation problem, and where *the corresponding Markov Decision Process is automatically constructed*. Using our implemented system, we demonstrate that using this system one can generalise from very small amounts of raw dialogue data, to a combinatorially large space of interactional variations, including phenomena such as question-answer pairs, over-answering, self- and other-corrections, split-utterances, and clarification interaction, when most of these are not even observed in the original data (see section 4.1).

1.1 Dimensions of Pragmatic Synonymy

There are two important dimensions along which dialogues can vary, but nevertheless, lead to very similar final contexts: interactional, and lexical. Interactional synonymy is analogous to syntactic synonymy - when two distinct sentences are parsed to identical logical forms - except that it occurs not only at the level of a single sentence, but at the dialogue or discourse level - Fig. 1 shows examples. Importantly as we shall show, this type of synonymy can be captured by grammars/models of dialogue context.

Lexical synonymy relations, on the other hand, hold among utterances, or dialogues, when different words (or sequences of words) express meanings that are sufficiently similar in a particular domain or activity - see Fig 1. Unlike syntactic/interactional synonymy relations, lexical ones can often break down when one moves to another domain: lexical synonymy relations are domain specific. Here we do not focus on these, but merely note that lexical synonymy relations can be captured using Distributional Methods (see e.g. Lewis & Steedman (2013)), or methods akin to Eshghi & Lemon (2014) by grounding domain-general semantics into the non-linguistic actions within a domain.

2 Dynamic Syntax (DS) and Type Theory with Records (TTR)

Dynamic Syntax (DS) a is a word-by-word incremental semantic parser/generator, based around the Dynamic Syntax (DS) grammar framework (Cann et al., 2005) especially suited to the fragmentary and highly contextual nature of dialogue. In DS, words are conditional actions - semantic updates; and dialogue is modelled as the interactive and incremental construction of contextual and semantic representations (Eshghi et al., 2015) - see Fig. 2. The contextual representations afforded by DS are of the fine-grained semantic content that is jointly negotiated/agreed upon by the interlocutors, as a result of processing questions and answers, clarification requests, acceptances, self-/other-corrections etc. The upshot of this is that using DS, we can not only track the semantic content of some current turn as it is being constructed (parsed or generated) word by word, but also the context of the conversation as whole, with the latter also encoding the grounded/agreed content of the conversation (see e.g. Fig. 2, and see Eshghi et al. (2015); Purver et al. (2010) for details of the model). Crucially for our model below, the inherent incrementality of DS together with the word-level, as well as cross-turn, parsing constraints it provides, enables the word-by-word exploration of the space of grammatical dialogues,

and the semantic and contextual representations that result from them.

These representations are Record Types (RT, see Fig. 2) of Type Theory with Records (TTR, (Cooper, 2005)), useful for incremental specification of utterance content, underspecification, as well as richer representations of the dialogue context (Purver et al., 2010; Purver et al., 2011; Eshghi et al., 2012). For reasons of lack of space, we only note that the TTR calculus provides, in addition to other operations, the *subtype checking operation*, $\sqsubseteq$, among Record Types (RT), and that of the Maximally specific Common Supertype (MCS) of two RTs, which both turn out to be crucial for the automatic construction of our MDP model, and feature checking (for more detail on the DS-TTR Hybrid, see (Eshghi et al., 2012; Hough and Purver, 2014)).

3 The overall BABBLE method

We start with two resources: a) a DS-TTR parser *DS* (either learned from data (Eshghi et al., 2013), or constructed by hand), for incremental language processing, but also, more generally, for tracking the context of the dialogue using Eshghi et al.'s model of feedback (Eshghi et al., 2015; Eshghi, 2015); b) a set D of transcribed successful dialogues in the target domain.

Overall, we will demonstrate the following steps (see (Kalatzis et al., 2016) for more details):

1. Automatically induce the Markov Decision Process (MDP) state space, S, and the dialogue goal, G_D, from D;
2. Automatically define the state encoding function $F : C \rightarrow S$; where $s \in S$ is a (binary) state vector, designed to extract from the current context of the dialogue, the semantic features observed in the example dialogues D; and $c \in C$ is a DS context, viz. a pair of TTR Record Types: $\langle c_p, c_g \rangle$, where c_p is the content of the current, *PENDING* clause as it is being constructed, but not necessarily fully grounded yet; and c_g is the content already jointly built and *GROUNDED* by the interlocutors (loosely following the DGB model of (Ginzburg, 2012)).
3. Define the MDP action set as the *DS* lexicon L (i.e. actions are words);
4. Define the reward function R as reaching G_D, while minimising dialogue length.

We then solve the generated MDP using Reinforcement Learning, with a standard Q-learning method, implemented using BURLAP (MacGlashan, 2015): train a policy $\pi : S \rightarrow L$, where L is the DS Lexicon, and S the state space induced using F. The system is trained in interaction with a (semantic) simulated user, also automatically built from the dialogue data (see (Kalatzis et al., 2016) for details).

The state encoding function F , as shown in Figure 2 the MDP state is a binary vector of size $2 \times |\Phi|$, i.e. twice the number of the RT features. The first half of the state vector contains the grounded features (i.e. agreed by the participants) ϕ_i, while the second half contains the current semantics being incrementally built in the current dialogue utterance. Formally:
$s = \langle F_1(c_p), \ldots, F_m(c_p), F_1(c_g), \ldots, F_m(c_g) \rangle$;
where $F_i(c) = 1$ if $c \sqsubseteq \phi_i$, and 0 otherwise. (Recall that $\sqsubseteq$ is the RT subtype relation).

4 Discussion

We have so far induced two prototype dialogue systems, one in an 'electronic shopping' domain (as exemplified by the dialogues in Fig. 1) and another in a 'restaurant-search' domain showing that incremental dialogue systems can be automatically created from small amounts of dialogue transcripts - in this case both systems were induced from a single successful example dialogue.

From a theoretical point of view, this shows that DS-TTR as an incremental model of interactional dynamics, with a domain-specific reward signal/goal is sufficient for certain word sequences becoming routinised and learned as ways of performing specific kinds of speech act within the domain, without any prior, procedural specifications of such actions. Thus, a dialogue system learns not only *what* it needs to do, but also *how* and *when* to do it (e.g. in a 'restaurant-booking' task, it learns to ask "What kind of cuisine would you like?", in a situation where the user says she wants to book a table, but does not provide information about restaurant type): higher-, discourse-level dialogue structure is emergent from interaction in such a setting.

From the practical point of view of dialogue system development, the major benefits of this approach are in (1) more naturally interactive dialogue systems as the resulting systems are incremental and are thus able to handle inherently in-

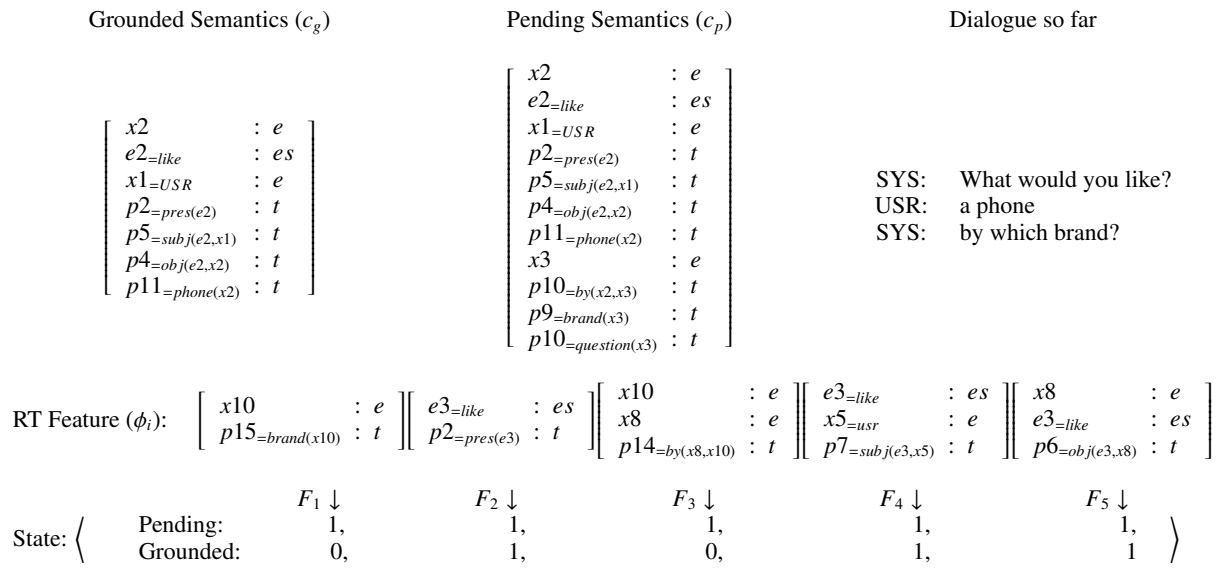


Figure 2: Semantics to MDP state encoding with RT features

cremental dialogue phenomena such as continuations, interruptions, and self-repair (see (Hough, 2015) for the DS-TTR model of self-repair); and (2) reduced development time and cost. To evaluate (2), below we consider the number of different dialogues that can be processed based on only 1 example training dialogue.

4.1 Number of interactional variations captured

Here we establish, as an example of the power of the method implemented, a lower-bound on the number of dialogue variants that can be processed based on training from *only 1 example dialogue*. Consider the training dialogue (which has only 2 'slots' and 4 turns) below:

SYS: What would you like?
USR: a phone
SYS: by which brand?
USR: by Apple

Parsing this dialogue establishes (as described above) a dialogue context that is required for success. The DS grammar is able to parse and generate many variants of the above turns, which lead to the same dialogue contexts being created, and thus also result in successful dialogues. To quantify this, we count the number of interactional variants on the above dialogue which can be parsed/generated by DS, and are thus automatically supported after training the system on this dialogue. Note that we do not take into account possible syntactic and lexical variations here, which would again lead to a large number of variants that the system can handle.

The *DS* grammar can parse several variants of the first turn, including overanswering ("I want an Apple laptop"), self-repair ("I want an Apple laptop, err, no, an LG laptop"), and ellipsis ("a laptop"), whose combinatorics give rise to 16 different ways the user can respond (not counting lexical and syntactic variations). These variations can also happen in the second user turn. If we consider the user turns alone, there are at least 256 variants on the above dialogue which we demonstrate that the trained system can handle. If we also consider similar variations in the two system turns (ellipsis, questions vs. statement, utterance completions, continuation, etc), then we arrive at a lower bound for the number of variations on the training dialogue of 8,192.

This remarkable generative power is due to the generalisation power of the DS grammar, combined with the system's DM/NLG policy which is created by searching through the space of possible (successful) dialogue variants.

5 Conclusion and ongoing work

We show how incremental dialogue systems can be automatically learned from example successful dialogues in a domain, with Dialogue Acts and discourse structure emergent rather specified in advance. This method allows rapid domain transfer – simply collect some example (successful) dialogues in a 'slot-filling' domain, and retrain. At present this is fully automated, and only requires checking that the DS lexicon covers the input data. We are currently applying this method to the problem of learning (visual) word meanings (groundings) from interaction.

References

Ronnie Cann, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language*. Elsevier, Oxford.

Robin Cooper. 2005. Records and record types in semantic theory. *Journal of Logic and Computation*, 15(2):99–112.

Arash Eshghi and Oliver Lemon. 2014. How domain-general can we be? Learning incremental dialogue systems without dialogue acts. In *Proceedings of Semdial 2014 (DialWatt)*.

Arash Eshghi, Julian Hough, Matthew Purver, Ruth Kempson, and Eleni Gregoromichelaki. 2012. Conversational interactions: Capturing dialogue dynamics. In S. Larsson and L. Borin, editors, *From Quantification to Conversation: Festschrift for Robin Cooper on the occasion of his 65th birthday*, volume 19 of *Tributes*, pages 325–349. College Publications, London.

Arash Eshghi, Julian Hough, and Matthew Purver. 2013. Incremental grammar induction from child-directed dialogue utterances. In *Proceedings of the 4th Annual Workshop on Cognitive Modeling and Computational Linguistics (CMCL)*, pages 94–103, Sofia, Bulgaria, August. Association for Computational Linguistics.

A. Eshghi, C. Howes, E. Gregoromichelaki, J. Hough, and M. Purver. 2015. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics (IWCS 2015)*, London, UK. Association for Computational Linguisitics.

Arash Eshghi. 2015. DS-TTR: An incremental, semantic, contextual parser for dialogue. In *Proceedings of Semdial 2015 (goDial), the 19th workshop on the semantics and pragmatics of dialogue*.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Patrick G. T. Healey. 2008. Interactive misalignment: The role of repair in the development of group sublanguages. In R. Cooper and R. Kempson, editors, *Language in Flux*. College Publications.

Julian Hough and Matthew Purver. 2014. Probabilistic type theory for incremental dialogue processing. In *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, pages 80–88, Gothenburg, Sweden, April. Association for Computational Linguistics.

Julian Hough. 2015. *Modelling Incremental Self-Repair Processing in Dialogue*. Ph.D. thesis, Queen Mary University of London.

Dimitrios Kalatzis, Arash Eshghi, and Oliver Lemon. 2016. Bootstrapping incremental dialogue systems: using linguistic knowledge to learn from minimal data. In *Proceedings of the NIPS 2016 workshop on Learning Methods for Dialogue*, Barcelona.

Ruth Kempson, Wilfried Meyer-Viol, and Dov Gabbay. 2001. *Dynamic Syntax: The Flow of Language Understanding*. Blackwell.

Ruth Kempson, Ronnie Cann, Arash Eshghi, Eleni Gregoromichelaki, and Matthew Purver. 2015. Ellipsis. In Shalom Lappin and Chris Fox, editors, *The Handbook of Contemporary Semantics*. Wiley-Blackwell.

Ruth Kempson, Ronnie Cann, Eleni Gregoromichelaki, and Stergios Chatzikiriakidis. 2016. Language as mechanisms for interaction. *Theoretical Linguistics*, 42(3-4):203–275.

Staffan Larsson and Robin Cooper. 2008. Corrective feedback and concept updates. In *Proceedings of The 2nd Swedish Language Technology Conference (SLTC-08)*.

Mike Lewis and Mark Steedman. 2013. Combined distributional and logical semantics. *Transactions of the Association for Computational Linguistics*, 1:179–192.

James MacGlashan. 2015. Burlap: Brown-umbc reinforcement learning and planning. In *http://burlap.cs.brown.edu/*.

G. Mills and E. Gregoromichelaki. 2010. Establishing coherence in dialogue: sequentiality, intentions and negotiation. In *Proceedings of SemDial (PozDial)*.

Gregory Mills. 2011. The emergence of procedural conventions in dialogue. In *IProceedings of the 33rd Annual Conference of the Cognitive Science Society*.

Matthew Purver, Eleni Gregoromichelaki, Wilfried Meyer-Viol, and Ronnie Cann. 2010. Splitting the 'I's and crossing the 'You's: Context, speech acts and grammar. In P. Łupkowski and M. Purver, editors, *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, pages 43–50, Poznań, June. Polish Society for Cognitive Science.

Matthew Purver, Arash Eshghi, and Julian Hough. 2011. Incremental semantic construction in a dialogue system. In J. Bos and S. Pulman, editors, *Proceedings of the 9th International Conference on Computational Semantics*, pages 365–369, Oxford, UK, January.

Procedural Syntax and Interactions

Eleni Gregoromichelaki
King's College London
University of Osnabrueck
elenigregor@gmail.com

Abstract

The view of NLs as codes mediating a mapping between "expressions" and the world is abandoned to give way to a view where utterances are seen as actions aimed to locally and incrementally alter the affordances of the context. Such actions employ perceptual stimuli composed not only of "words" and "syntax" but also elements like visual marks, gestures, sounds, etc. Any such stimuli can participate in the domain-general processes that constitute the "grammar". The function of the grammar is dynamic categorisation of various perceptual inputs and their integration in the process of generating the next action steps. Given these assumptions, a challenge that arises is how to account for the reification of such processes as exemplified in apparent metarepresentational practices like quotation, reporting, citation etc. It is argued that even such phenomena can receive adequate and natural explanations through a grammar that allows for the ad hoc creation of occasion-specific content through reflexive mechanisms.

1 Language as action and grammar

Standard models that describe natural languages (NLs) as representational systems belong to the 'language-as-product' paradigm (Clark, 1992), concerned with the definition of linguistic representations, the "product" of linguistic processing. In this tradition, it has been a standard assumption that NL properties should be explained by reifying NLs as abstract codes, mapping forms (strings of symbols) to propositional intentions. However, a substantial amount of evidence indicates that NL use substantially affects NL structuring indicating an alternative characterisation: within a 'language as action' paradigm, NL properties can be explicated as coinciding with those of human action; an agent's linguistic actions are structured sequentially, directed by predictions of upcoming inputs, interleaved and interacting with other activities and agents. Accordingly, in everyday conversation, utterances are not expected to display evidence of necessary hierarchical constituency, e.g. sentential structuring: non-sentential utterances are adequate to underpin interlocutor coordination and all linguistic dependencies are resolvable across more than one turn:

(1) Angus: But Domenica Cyril is an intelligent and entirely well-behaved dog who
Domenica: happens to smell
[radio play, 44 Scotland Street]

In such cases, postulating a notion of well-formedness based on a code licensing units ranging over strings of words, as an independent level of structuring, impedes a natural account of such phenomena. This is because joining overt forms together often results in illformedness or misleading interpretations:

(2) A: I heard a bang. Did you hurt
B: myself? No, but Mary is in a state

Moreover, at the level of semantics/pragmatics of dialogue, the issue of recoverability of propositional intentions is also problematic, e.g., in cases such as (5) where various speech acts are accomplished within the unfolding of a shared single proposition (see Gregoromichelaki et al. (2011)):

(3) Jack: I just returned
Kathy: from . . .
Jack: Finland. [Lerner (2004)]

(4) Eleni: A: Are you left or
Yo: Right-handed. [natural data]

(5) Hester Collyer: It's for me.
Mrs Elton the landlady: And Mr. Page?
Hester Collyer: is not my husband. But I'd rather you think of me as Mrs. Page. [The Deep Blue Sea (film)]

This endemic context-sensitivity and situatedness of NL use is indicative of the fact that both content and structure are emergent products of the

processes and practices underpinning human interaction. For these reasons, the more general approach to NL analysis argued for here revolves around the idea that structures, objects, concepts, concrete reality (and even the individual self) can all be taken as metaphysically emergent categories with processes, mechanisms, and change as ontologically primary.[1]

2 DS-TTR

A grammar architecture adopting this perspective can be articulated within DS-TTR (Cann et al. (2005); Purver et al. (2010); Gregoromichelaki (in press)). Here NLs are conceived as comprising sets of processes modelled formally as *procedures*. Both NLs' temporal structuring (syntax) and lexical specifications are analysed as involving stored sequences (*macros*) of elementary (epistemic) actions, defined in an IF-THEN-ELSE format. Such actions incrementally and predictively build or linearise conceptual categories expressed in TTR-representations (Cooper, 2012). The model assumes tight interlinking of NL perception and action: production uses simulation and testing of parse states in order to license the generation of strings; comprehension predictively builds structures to accommodate upcoming inputs in order to constrain efficiently the usual overwhelming ambiguity of NL stimuli. By imposing top-down predictive and goal-directed processing at all comprehension and production stages, interlocutor feedback is incrementally anticipated and integrated. The model includes subsentential tracking of the shifting contextual parameters of each word-utterance event (Eshghi et al. (2015); Gregoromichelaki (in press)). *Context* constitutes an integral part of the grammar, not only as a record of the shifting parameters that provide for the interpretation of various indexical elements (e.g. *myself* in (2)), but also storing (a) the emergent (partial) structures constructed from the contributions of all participants; (b) the phonological/graphical elements that have been employed; (c) the actions used, recorded as traversals of paths in a graph display; (d) processing paths that have been considered as probabilistically live options but not eventually pursued (Sato, 2011; Hough, 2015). Storing the action paths is necessary for the resolution of anaphora and ellipsis, especially "sloppy" or "paycheck" readings, whose resolution relies on re-executing ('rerunning') previous action sequences in an updated processing environment. Maintaining abandoned options is required for the modelling of backtracking in subsententially occurring conversational phenomena like clarification, self-/other-corrections, etc. but also humour effects and puns (Gregoromichelaki, in press). Consequently, coordination among interlocutors is seen not as inferential metarepresentational activity but as the outcome of the fact that the grammar consists of a set of licensed complementary actions that both speakers and hearers have to perform in synchrony (Gregoromichelaki et al., 2013).

[1]This view has its roots in an ancient philosophical programme starting in the Western world with Heraclitus, situated within a tradition following, among others, Martin Heidegger, Ilya Prigogine, Gilles Deleuze, and even encompassing current notions like the concept of the extended mind (Clark and Chalmers, 1998; Clark, 2008).

2.1 Quotation in DS-TTR

Given these assumptions, a challenge that arises is how to account for the reification of grammatical processes as exemplified in apparent metarepresentational practices like quotation, reporting, citation etc. As we saw earlier in (1)-(5), perfectly intelligible moves in dialogue can be achieved simply by initiating a grammatical dependency which prompts either interlocutor to fulfill it without specific determination or identifiability of a given speech-act. In various other cases though, the interlocutor completing somebody else's utterance might be seen as offering the completion along with a query as to whether such a (meta)representation is what the other interlocutor would have said (e.g. (2)). There are further such phenomena in cases of citation, quotation, reports, echoing uses, and code-switching:

(6) "Cities," he said, "are a very high priority."

(7) Wright won't disclose how much the Nike deal is worth, saying only that "they treat **me** well". [De Brabanter (2010)]

(8) A doctor tells him [Gustave Flaubert] he is like a "vieille femme hysterique"; he agrees. [De Brabanter (2010)]

(9) Alice said that life is "difficult to understand". [Cappelen and Lepore (1997)]

(10) Mary felt relieved. If Peter came tomorrow, she would be saved. [Recanati (2010)]

Despite recent attempts to integrate such phenomena within standard grammars (e.g., (Ginzburg and Cooper, 2014; Maier, 2014; Potts, 2007)), certain data are not amenable to appropriate treatment due to the lack of modelling incrementality within these

formalisms. For example, as can be seen in (6)-(9), quotation can appear subsententially, and discontinuously, at any point, which means that contextual parameters regarding the utterance event and semantic evaluation need to be able to shift incrementally at each word-by-word processing stage. Additionally, quotation is one of the environments where the phenomenon of split-utterances is observed frequently as an opportunity arises for co-constructing a vivid unified perspective of some (actual or imaginary) speech/thought event (Gregoromichelaki, in press):

(11) Clinician: So I watch this person being killed and then I go to bed and I'm you know lying there going, "well"
Patient: "did I hear something?" [Duff et al. (2007)]

The contextual parameters relevant to the resolution of indexicals (e.g. *I*) in such cases, even though needing to shift mid-sentence, do not necessarily track the current speaker/hearer roles. Moreover, such role-switches include cases where the same structure can be employed both as expressing a speaker's own voice and as a subsequent quotation:

(12) A: SOMEONE is keen [BBC]
B: says the man who slept here all night

In all such cases, issues of "footing" (Goffman, 1979), namely changes in perspectives and roles assumed by interlocutors, intersect with syntactic/semantic issues of direct/indirect speech constructions and speech-act responsibility and echoing. For these reasons, an adequate account of the function of such NL devices can be given straighforwardly in DS-TTR due to its incremental modelling of context shifting, the potential for sharing of syntactic/semantic dependencies, and the fact that there is no requirement to derive a global propositional speech act (Gregoromichelaki (in press); Gregoromichelaki & Kempson (2016)).

On the other hand, modelling the potential of partially assuming another speaker's role, being perceived as "demonstrating" what somebody else was going to say, and the "metalinguistic" appearance of various such phenomena might seem especially problematic aspects for the DS-TTR stance:

(13) "Life is difficult" is grammatical.
(14) James says that "Quine" wants to speak to us. [James thinks that McPherson is Quine]
(15) "I talk better English than the both of youse!" shouted Charles, thereby convincing me that he didn't.

A DS-TTR grammar takes words (and the operation of "syntax" in general) as offering affordances exploited by the interlocutors to facilitate interaction. This means that words and linguistic constructions are NOT conceptualised as abstract code elements, expression types, that are associated with referential/semantic values (cf Cooper (2014) where string structure is still presumed). With no privileged semantic entities corresponding to (types of) expressions, only domain-general mechanisms for processing stimuli, quotation thus offers a crucial test for the legitimacy of these DS-TTR claims: when processing a quoted/cited string, what happens within the quotation marks (or any other indications) according to these assumptions?

In fact, it turns out that such cases are also unproblematic for the DS-TTR model, and can be explicated in a natural manner that conforms with intuitions and parallels the modelling of anaphora/ellipsis. First, in order to model cases like (6)-(10), (14), (15), as well as mid-sentence general code-switching, it has to be assumed that the context keeps track incrementally, through a designated metavariable (**g** in (16)), of which and whose grammar is being employed at each particular subsentential stage (cf Ginzburg and Cooper (2014)). Next, consider the most challenging cases, namely, metalinguistic uses, for example (13), so-called *pure quotation*, where an NL-string appears in a regular NP position. Under DS-TTR assumptions, this will be a pointer position where the grammar has already generated a prediction for the processing of a singular term ($?Ty(e)$, other cases might involve $?Ty(cn)$, etc.). The explanation of what happens here is based on the fact that actions are first-class citizens in DS-TTR. This means that previous actions can be invoked by the grammar to be re-executed ('rerun') in order to provide parallel but distinct contents as needed in cases of sloppy-ellipsis or paycheck-pronoun readings. From this perspective, metalinguistic, echoic, and similar uses are cases where the actions specified by some grammar *g* for processing a particular string, e.g. the embedded sentential string in (13), come to be executed on the spot to provide an ad hoc conceptualisation of a demonstrated action. The formalisation of the basic mechanism is shown in (16) below. Different variants of this macro and combinations with other independently needed components of the grammar account for all such phenomena:

(16)

IF	$?Ty(\mathbf{x}_{\in\{e,cn,\ldots\}})$
THEN	put $Ty(\mathbf{x})$
	put $(u_{q=\mathtt{run}_{\mathbf{g}}(\langle \mathbf{a}_i,\ldots,\mathbf{a}_{i+n}\rangle)}: e_s)$
ELSE	abort

(a) demonstration action

In (16), the higher-order action `run`, also employed in cases of sloppy anaphora, is triggered. `run` is parameterised to some grammar *g* (replacing the metavariable **g**), which can be distinct from the grammar used for parsing/producing the rest of the string (see (8), (15)). At the same time, the executed sequence of actions $\langle \alpha_i, ..., \alpha_n \rangle$, bound to the rule-level variables $\langle \mathbf{a}_i,...,\mathbf{a}_n \rangle$, confers the ad hoc conceptual type of the quoting utterance event u_q which therefore functions as a demonstration. The performance of this demonstration event is then categorised as belonging to the already predicted semantic type, here, in (13), a referential term ($Ty(e)$) (feasible due to TTR's subtyping definitions). The rest of the string then delivers a content that combines with the reification of this ad hoc execution. In (13), this delivers the interpretive result that this demonstration of the execution of the grammatical actions is characterised as having the property derived from processing *is grammatical*. For echoic cases, where the interpretation of the indexicals shifts following parameter values supplied by the invoked context, e.g. (7), (15), a similarly triggered action execution is accomplished, this time, with parallel introduction of the quoting context as a mentioned utterance event *u*, replacing the metavariable **u** of type e_s, i.e. eventuality:

(17)

IF	$?Ty(\mathbf{x}_{\in\{e,cn,\ldots\}}),\ [\ \text{CONTEXT}\ :\ [\ldots[\ \mathbf{u}\ :\ e_s\ ]]\ ]$
THEN	put $Ty(\mathbf{x})$
	put $(u_{q=\mathtt{run}_{[\text{CONTEXT}\,:\,[\mathbf{u}\,:\,e_s]]}(\langle \mathbf{a}_i,\ldots,\mathbf{a}_{i+n}\rangle)}\ :\ e_s)$
ELSE	abort

(b) demonstration-and-echoing action

Cases of direct quotation (e.g. (11), (12), (15)) are those where such a freely-available contextual switch has been grammaticalised in English.

Notably, given that the DS-TTR grammar does not provide form-meaning correspondences but only provides for the parsing/generation of stimuli in context, the same mechanism can be applied to non-linguistic signals/demonstrations: reifying the processing of some upcoming element to provide ad hoc content of another already predicted type explains how non-linguistic signals can compose subsententially with linguistic ones as the conceptualisation of some experience being demonstrated:

(18) John saw the spider and was like "ahh!"
(19) John was eating like [gobbling gesture]
(20) She went "Mm Mmmrn Mphh"

The existence of such compositions, along with all the previous data, might be challenging, under one construal, for the account of NL-gesture coordination in Rieser (2014; 2015). Rieser presents a framework (the λ-π calculus) where NL and gesture are modelled as independent but communicating processes. Even though the process metaphysics incidentally mentioned there is a welcome development, the assumption of independence might be questioned. First, this assumption seems to be an artifact of presupposing that NLs are structured codes mediating arbitrary mappings from standard syntactic forms (trees inhabited by words) to propositional meanings (e.g. λ-calculus formulae). Since the co-speech gestures examined are related to imagery (aural, visual, etc.) in an iconic manner, modelling their contribution in the standard way needs to abstract representations from the kinematics that cannot be unified with NL syntactic representations. In contrast, the view taken here is that NL actions do not require an independent syntax relying on the hierarchical structuring of stimuli sequences. Hence production/comprehension of stimuli in various modalities need not be segregated. Second, the major argument in Rieser's analysis comes from SaGA data (Lücking et al., 2013) where NL segments and gesture-strokes seem not to synchronise perfectly. However, this is not an argument for considering such stimuli qualitatively distinct. Perfect synchronisation is not necessary within a single modality either, e.g. dialogue participants do not perfectly synchronise their turns. In a predictive framework like DS-TTR, such asynchrony might reveal a purpose, for example, co-speech gestures can be modelled as elaborating or narrowing down predictions that precede the processing of NL input. But then, under this view, there is a viable and useful application of the λ-π calculus in the DS-TTR framework too. Given that DS-TTR processing is strictly incremental pursuing only one path at a time, it is possible that various sources of information might compete for sequential positions. Introducing ad-hoc channel interfaces, modelled with resources from the λ-π calculus, can provide for the implementation of a sequentiality-repair mechanism, ordering inputs/outputs, even within the same modality, so that they can be processed strictly incrementally.

References

Ronnie Cann, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language*. Elsevier, Oxford.

Herman Cappelen and Ernie Lepore. 1997. Varieties of quotation. *Mind*, 106(423):429–450.

Andy Clark and David Chalmers. 1998. The extended mind. *Analysis*, pages 7–19.

Herbert H. Clark. 1992. *Arenas of Language Use*. University of Chicago Press.

Andy Clark. 2008. *Supersizing the mind: Embodiment, action, and cognitive extension*. OUP USA.

Robin Cooper. 2012. Type theory and semantics in flux. In R. Kempson, T. Fernando, and N. Asher, editors, *Handbook of the Philosophy of Linguistics*, volume 14: Philosophy of Linguistics, pages 271–323. Elsevier.

Robin Cooper. 2014. Phrase structure rules as dialogue update rules. In V. Rieser and P. Muller, editors, *Proceedings of DialWatt - Semdial 2014: The 18th Workshop on the Semantics and Pragmatics of Dialogue, pages 4352, Edinburgh, 13 September 2014.*

Philippe De Brabanter. 2010. The semantics and pragmatics of hybrid quotations. *Language and Linguistics Compass*, 4(2):107–120.

Melissa C Duff, Julie A Hengst, Daniel Tranel, and Neal J Cohen. 2007. Talking across time: Using reported speech as a communicative resource in amnesia. *Aphasiology*, 21(6-8):702–716.

Arash Eshghi, Christine Howes, Eleni Gregoromichelaki, Julian Hough, and Matthew Purver. 2015. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 261–271.

Jonathan Ginzburg and Robin Cooper. 2014. Quotation via dialogical interaction. *Journal of Logic, Language and Information*, 23(3):287–311.

Erving Goffman. 1979. Footing. *Semiotica*, 25(1-2):1–30.

Eleni Gregoromichelaki and Ruth Kempson. 2016. Reporting, dialogue, and the role of grammar. In Alessandro Capone, Ferenc Kiefer, and Franco Lo Piparo, editors, *Indirect reports and pragmatics*, pages 115–150. Springer.

Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, Gregory J. Mills, Ronnie Cann, Wilfried Meyer-Viol, and Patrick G. T. Healey. 2011. Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, 2(1):199–233.

E. Gregoromichelaki, R. Kempson, and C. Eshghi Howes. 2013. On making syntax dynamic: the challenge of compound utterances and the architecture of the grammar. In Ipke Wachsmuth, Jan de Ruiter, Petra Jaecks, and Stefan Kopp, editors, *Alignment in Communication: Towards a New Theory of Communication*. John Benjamins, "Advances in Interaction Studies".

Eleni Gregoromichelaki. in press. Quotation in dialogue. In & Johnson M. Saka, P., editor, *The Semantics and Pragmatics of Quotation.* Dordrecht: Springer.

Julian Hough. 2015. *Modelling Incremental Self-Repair Processing in Dialogue*. Ph.D. thesis, Queen Mary University of London.

Gene H. Lerner. 2004. Collaborative turn sequences. In *Conversation analysis: Studies from the first generation*, pages 225–256. Gene H. Lerner, John Benjamins.

Andy Lücking, Kirsten Bergman, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2013. Data-based analysis of speech and gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications. *Journal on Multimodal User Interfaces*, 7(1-2):5–18.

Emar Maier. 2014. Mixed quotation: The grammar of apparently transparent opacity. *Semantics & pragmatics*, 7(7):1–67.

Christopher Potts. 2007. The dimensions of quotation. In Chris Barker and Polly Jacobson, editors, *Proceedings from the workshop on direct compositionality*, pages 405–431. Oxford: Oxford University Press.

Matthew Purver, Eleni Gregoromichelaki, Wilfried Meyer-Viol, and Ronnie Cann. 2010. Splitting the i's and crossing the you's: Context, speech acts and grammar. In P. Łupkowski and M. Purver, editors, *Proceedings of SemDial 2010 (PozDial)*, pages 43–50, Poznan, Poland, June 2010. Polish Society for Cognitive Science.

François Recanati. 2010. *Truth-conditional pragmatics*. Clarendon Press Oxford.

Hannes Rieser. 2014. Gesture and speech as autonomous communicating processes. MS, University of Bielefeld.

Hannes Rieser. 2015. When hands talk to mouth. gesture and speech as autonomous communicating processes. In *Proceedings of SEMDIAL 2015-GoDIAL*, page 122.

Yo Sato. 2011. Local ambiguity, search strategies and parsing in Dynamic Syntax. In R. Kempson, E. Gregoromichelaki, and C. Howes, editors, *The Dynamics of Lexical Interfaces*. CSLI Publications.

Communicative and Cognitive Pressures in Semantic Alignment

Dariusz Kalociński
Institute of Philosophy
University of Warsaw
d.kalocinski@uw.edu.pl

Abstract

Descriptions used by participants in conversation tend to be progressively systematized. A paradigmatic example of this phenomenon is the global shift from concrete to abstract descriptions observed in Maze Task dialogues. We propose to explain this trend by the appeal to communicative and cognitive pressures exerted on participants during conversation. We conclude that models of meaning coordination in dialogue should incorporate communicative and cognitive biases towards expressiveness and ease of processing.

1 Introduction

One of the most robust findings in experimental psychology of dialogue is that participants tend to spontaneously systematize their means of referring in task-oriented conversation. Since the seminal maze task experiment by Garrod and Anderson (1987), the evidence for this has been ubiquitous (Garrod and Doherty, 1994; Healey, 1997; Healey and Mills, 2006; Mills and Healey, 2006; Healey, 2008). Despite several empirically motivated approaches to modelling meaning coordination in dialogue (Garrod and Anderson, 1987; Garrod and Doherty, 1994; Pickering and Garrod, 2004b; Healey, 1996; Healey, 2008), the acclaimed global trend of conceptual and semantic change has remained largely unexplained.

The maze task involves two participants, connected by a two-way audio link and seated in separate rooms in front of a computer displaying a two-dimensional maze. Each player is supposed to reach a target node by moving his position marker through the maze. None of the players can see the position nor the target of the other participant. Crucial paths are blocked by gates which can be opened by stepping onto switch nodes but this can only happen by guiding one's partner and making him step onto the switch he cannot see. Thus, participants are faced with the recurrent coordination problem of developing and sustaining a system of descriptions to refer to maze locations.

Figural: refers to salient features of the maze

"the l-shape sticking out at the top"

"the uppermost box"

Path: refers to a route from one node to another

"Go 2 up, 1 down, 2 along, 5 up"

"up, right, down, up"

Line: refers to nodes treated as intersects of horizontal and vertical vectors

"3rd row, 5th box", "4th column, 2nd square"

"The third row, fifth to the left"

Matrix: coordinate-system

"4,2", "A,1"

Figure 1: Description types used in Maze Task experiments.

Garrod and Anderson (1987) classify descriptions used by participants in maze task experiments into four types (see Figure 1). It has been repeatedly observed that description types used most frequently initially tend to be abandoned later on in favour of new, previously less frequent forms (Garrod and Anderson, 1987; Garrod and Doherty, 1994). Crucially, descriptions used in Maze Task experiments tend to migrate across trials from more "concrete" (Figural and Path) to more "abstract" (Line and Matrix). As reported by Mills and Healey (2008), a typical shift is exemplified by the excerpt of dialogue presented

in Table 1. Still though, participants occasion-

0 mins:	The piece of the maze sticking out
2 mins:	The left hand corner of the maze
5 mins:	The northenmost box
10 mins:	Leftmost square of the row on top
15 mins:	3rd column middle square
20 mins:	3rd column 1st square
25 mins:	6th row longest column
30 mins:	6th row 1st column
40 mins:	6 r, 1 c
45 mins:	6,1

Table 1: Semantic shift from Figural/Path to Line/Matrix descriptions in Maze Task dialogues.

ally change descriptions to more "concrete", especially when they encounter problematic dialogue (Healey, 1996; Healey and Mills, 2006; Mills and Healey, 2006).

The question is why the migration pattern looks as in Figure 1? Crucially, the pattern cannot be seen as a simple contraction of form as different description schemes seem to rely on incompatible situation models (Garrod and Anderson, 1987). The drift of description types is thus better seen as a directional conceptual and semantic change.

The migration pattern is also difficult to reconcile with existing models of semantic alignment in dialogue. For example, the input-output coordination model by Garrod and Anderson (1987) and the interactive alignment model by Pickering and Garrod (2004b) are based on a tacit priming mechanism and as such are claimed too conservative to account for innovative changes in description schemes (Garrod, 1999; Healey, 2004; Pickering and Garrod, 2004a; Mills and Healey, 2006). The repair-driven account by Healey (1997; 2006; 2008) sketches how alignment might proceed through local resolution of problematic understanding but does not explain why meanings tend to migrate the way they actually do.

What we propose is to account for the directional drift of description types by the appeal to communicative and cognitive pressures acting on interlocutors during alignment in dialogue.

2 Expressiveness and Ease of Processing

The idea that certain features of natural language stem from the pressures imposed on subjects during language learning and use has been explored in linguistics successfully on many levels. Perhaps one of the earliest such theories explains the inverse relationship between frequency and length of words by the appeal to competing motivations of speaker and hearer (Zipf, 1949). According to a more recent theory, language structure is, to a large extent, an adaptation of language itself to multiple constraints imposed during learning and use (Christiansen and Chater, 2008). For example, it has been argued that compositionality arises from the trade-off between pressures for compressibility and expressivity (Kirby et al., 2015).

If we want to explain the migration pattern in terms of pressures acting on discussants, the putative pressures should fit the timescale of a conversation. In our explanation we refer to two generic pressures which are equally applicable to dialogue situations: expressivity and ease of processing.

The pressure for expressiveness plays an important role in the maze task. Due to the novelty of the task, participants start with a little common ground and possibly few semantic precedents. To accomplish the game, they need to develop linguistic means to refer to relevant maze locations. In principle, a salient maze location could be any location in the maze whatsoever. Thus, the nature of the task imposes pressure for expressiveness on language being used and developed by participants in dialogue. We envisage a fully functional language as allowing for information exchange about arbitrary locations.

Ease of processing is another important factor which is likely to affect descriptions developed by participants. There are at least two levels at which this pressure applies. First, speaker may tend to use shorter descriptions in order to reduce his effort (Zipf, 1949). This tendency partially explains shortening of descriptions (see Table 1). Second, ease of processing is tightly coupled with deeper levels of production and comprehension. On the cognitive side, descriptions are associated with procedures which are intermediaries between formal and semantic levels of representation (Tichý, 1969; Suppes, 1980). For example "xth row, yth box" may be coupled with a particular procedure which, given a relevant situational model of the maze, and the location intended by the speaker, computes n (say, by counting rows from the bottom) and m (say, by counting boxes from the right) which are then plugged into the description form. If situational model and se-

mantic representations are sufficiently aligned between participants (Pickering and Garrod, 2004b), the hearer's interpretation boils down to almost the same procedure: counting n rows from the bottom, m boxes from the right and thus getting the intended location right.

When thinking about semantic representations in terms of procedures, it is natural to ask about complexity of corresponding problems (functions from inputs to outputs) and linking relevant complexity measures with cognitive reality (see, e.g., Szymanik (2016)). It is also natural to expect that greater complexity of a procedure may provide a pressure for finding more efficient solutions. For example, Schlotterbeck and Bott (2013) have shown that intractable meanings tend to be avoided by human participants in verification of sentences having both tractable and intractable interpretations. It seems, however, that the pressure for ease of processing may be equally important in selecting between feasible interpretations, which are nevertheless distinguished by different complexity characteristics. We return to this in Section 5.

3 Amount of Ambiguity vs Alignment

Participants in maze task dialogues are often misaligned at the level of semantic representation and situation model. Let us define the concept of semantic misalignment in terms of procedures which participants associate with descriptions. We say that the meaning of a given description form (say, "xth row, yth column") is misaligned between participants if the procedures they associate with the description form are not extensionally equivalent. What it means is that for some instances of the description, participants' procedures fail to give the same output.

Consider a Matrix description "4,3" as an example. There are several natural algorithms matching this type of input. The input itself does not specify which coordinates correspond to horizontal and vertical vectors. Moreover, the description does not hint about counting procedure—should one start from the top or from the bottom? From left or from right? Taking into account only this sort of underspecification, we get eight extensionally non-equivalent procedures.

As for Line descriptions like "5th row, 3rd column", underspecification is less severe. Provided that "row" designates horizontal vectors, the association between coordinates and horizontal/vertical vectors is fixed and thus one degree of freedom disappears which reduces ambiguity twice (procedures not conforming to the coordinate-dimension convention are discarded). Moreover, some description forms which are classified as Line descriptions ("The third row, fifth to the left") are even less ambiguous.

Path descriptions can still manifest some amount of ambiguity. Perhaps the most precise way of tracing the route along connected nodes is by means of descriptions like "up, right, down" etc. This way we are able to trace the path to the destination node unambiguously. However, using "2 along" or even "2 up" is potentially ambiguous as it is not specified whether one should start counting from the current position (Pickering and Garrod, 2004b). Hence, certain Path descriptions seem to manifest similar amount of ambiguity as Line descriptions.

Figural descriptions pick out easily identifiable features of the maze and seem least ambiguous ("the northenmost box"). Obviously, figural descriptions sometimes fail to denote precisely one box like in "the l-shape sticking out at the top". However, they allow participants to focus on particular, easily identifiable portions of the maze without the risk of misunderstanding.

An important link between ambiguity and semantic coordination is that greater ambiguity hinders alignment. Based on the foregoing considerations, the order of migration pattern (Figural/Path $\rightarrow$ Line/Matrix) respects the increasing order of ambiguity and, hence, of alignment complexity. This view is strengthened by the fact that meanings usually associated with each type of description are equally expressive and complex (see Sections 4, 5) which makes them roughly equally likely to be selected during alignment.

4 Expressiveness

Figural descriptions are least expressive. Certain boxes are easily describable ("the leftmost box of the row on top") while others are not identifiable by any simple figural description, especially if the maze does not contain easily distinguishable parts. On the other hand, there are maze configurations which are particularly likely to invoke Figural descriptions (Garrod and Anderson, 1987).

Path descriptions are more expressive than Figural—in principle, one can trace a route along

interconnected nodes to any location reachable from a given starting point. Thus, alignment on Path description is sufficient to solve the entire maze and is strictly favoured by the pressure for expressiveness. Moreover, even if interlocutors are not aligned on Path descriptions, using them seems to be a safer strategy as it gives participants more control over the location of their partner.

Line and Matrix descriptions are most expressive. They allow to identify any node in the maze whatsoever. Hence, alignment on Line or Matrix description is also sufficient for solving the maze. However, acting according to misaligned Line or Matrix descriptions can lead to serious troubles as non-equivalent procedures of this sort fail to produce the same outputs for most inputs and—moreover—output boxes generated by such procedures may be distant from each other in the maze.

5 Complexity

By inspecting Table 1, we see that the longer the description, the earlier its place in the migration ordering. Hence, descriptions which come out as earlier in this ordering are associated with greater effort on the part of the speaker. Note, however, that Path descriptions make this picture somewhat more complicated as their lengths may vary considerably depending on the length of the denoted path. Indeed, Path descriptions of short routes can be more concise than Line and Matrix descriptions ("up, right") whereas Path descriptions of long routes can easily surpass the length of long Figural descriptions. Consequently, a Path description may be preferred or dispreferred, depending on its actual length, accordingly.

We now turn to the complexity measure associated with procedures. First, observe that Path descriptions correspond to quite a different task than Line and Matrix descriptions. In abstracto, the underlying problem is that of finding a route connecting two nodes of the graph. Obviously, participants cannot bypass this sort of problem as this is actually what they are required to do: solve the maze by going from their positions to other dedicated positions. However, this sort of task is more difficult than simply computing the position of a given node which always requires at most linear time with respect to n, where, conceptually, the maze is arranged on n horizontal/vertical lines of length n or $n \times n$ matrix. Finding a path between two nodes may require non-linear time; for example, inspecting half of the nodes of the maze, which amounts to roughly $n^2/2$ steps. Note, however, that the actual influence of this factor depends on the size of the maze and, presumably, on its structure as well.

6 Explaining the Migration Pattern

Abandoning Figural descriptions seems to be explained by the pressure for expressiveness. As already noted, crucial parts of the maze may be difficult to pinpoint using mere Figural descriptions.

Migration from Path to Line (or Matrix) seems to be driven by cognitive pressures exerted on the speaker. Line or Matrix descriptions are shorter and the associated procedures are less complex. By abandoning Path descriptions, the effort of production is greatly reduced and the cost of computing the path is delegated to the hearer. Moreover, using Path descriptions may take longer on average. Hence, steering away from them may reduce the joint effort of participants (Clark and Wilkes-Gibbs, 1986).

Finally, the advantage of Matrix over Line forms seems to be associated solely with their lengths as computational complexity of Line and Matrix procedures is the same.

As we have observed, Matrix descriptions seem to be highly ambiguous. Line descriptions are less ambiguous but can still be quite problematic. This ambiguity and its potential for causing misalignment is perhaps the main reason for not using Line or Matrix descriptions consistently right from the start. This means that language users may resolve to less ambiguous (Figural) or less ambiguous/more safe (Path) strategies. Nonetheless, due to the presence of cognitive and communicative pressures we should expect that participants will tend to align on short forms associated with computationally efficient procedures.

7 Conclusions

We have proposed to explain the migration pattern observed in dialogues from Maze Task experiments by the appeal to communicative and cognitive pressures exerted on participants during conversation. Considering the effort associated with production of descriptions and computation of referential information by means of procedures seems to be an important aspect that should be taken into account when developing models of alignment in dialogue.

Acknowledgments

This work is supported by the Polish National Science Centre grant 2015/19/B/HS1/03292.

References

Morten H. Christiansen and Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(05):489–509.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1 – 39.

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

Simon Garrod and Gwyneth Doherty. 1994. Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3):181–215.

Simon Garrod. 1999. The challenge of dialogue for theories of language processing. *Language processing*, pages 389–415.

Patrick G.T. Healey and Gregory J. Mills. 2006. Participation, Precedence and Co-ordination in Dialogue. In R. Sun and N. Miyake, editors, *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pages 1470–1475, Vancouver, Canada.

Patrick G.T. Healey. 1996. Communication as a Special Case of Misunderstanding: Semantic Coordination in Dialogue.

Patrick G.T. Healey. 1997. Expertise or Expert-ese? The Emergence of Task-Oriented Sub-Languages. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, pages 301–306.

Patrick G.T. Healey. 2004. Dialogue in the degenerate case? *Behavioral and Brain Sciences*, 27(02):201–201. Peer Commentary on Pickering and Garrod: 'The Interactive Alignment Model'.

Patrick G.T. Healey. 2008. Interactive misalignment: The role of repair in the development of group sublanguages. In R. Cooper and R. Kempson, editors, *Language in Flux*, volume 212, pages 13–39. Palgrave-McMillan.

Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.

Gregory J. Mills and Patrick G.T. Healey. 2006. Clarifying spatial descriptions: Local and global effects on semantic co-ordination. In *Proceedings of Brandial06 The 10th Workshop on the Semantics and Pragmatics of Dialogue. University of Potsdam, Germany*, pages 122–129.

Gregory J. Mills and Patrick G.T. Healey. 2008. Semantic negotiation in dialogue: the mechanisms of alignment. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 46–53. Association for Computational Linguistics.

Martin J. Pickering and Simon Garrod. 2004a. The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, 27(02):212–225.

Martin J. Pickering and Simon Garrod. 2004b. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02):169–190.

Fabian Schlotterbeck and Oliver Bott. 2013. Easy Solutions for a Hard Problem? The Computational Complexity of Reciprocals with Quantificational Antecedents. *Journal of Logic, Language and Information*, 22(4):363–390.

Patrick Suppes. 1980. Procedural Semantics. In R. Haller and W. Grassl, editors, *Language, Logic, and Philosophy: Proceedings of the 4th International Wittgenstein Symposium*, pages 27–35. Hölder-Pichler-Tempsy, Vienna.

Jakub Szymanik. 2016. *Quantifiers and Cognition: Logical and Computational Perspectives*. Number 96 in Studies in Linguistics and Philosophy. Springer International Publishing, 1 edition.

Pavel Tichý. 1969. Intension in terms of Turing machines. *Studia Logica*, 24(1):7–21.

George K. Zipf. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley Press, Cambridge, MA, USA.

Cognitive science, language as a tool for interaction, and a new look at language evolution

Ruth Kempson
Department of Philosophy
King's College London
ruth.kempson@kcl.ac.uk

Stergios Chatzikyriakidis & Christine Howes
Philosophy, Linguistics and Theory of Science
University of Gothenburg
{stergios.chatzikyriakidis, christine.howes}@gu.se

Abstract

We explore prerequisites necessary for embedding Dynamic Syntax within an account of language evolution. We show how the dynamics of processing as modelled in Dynamic Syntax display remarkable parallelism with Clark's (2016) Predictive Processing Model and that the interactive stance of a combined DS/PPM model of language/cognition reflects the Multi-Level Selection Hypothesis – with groups as units for evolutionary purposes, not just individuals. With these assumptions, language emerges without necessary invocation of rich innate encapsulated structures or mind-reading capacities, paralleling first language acquisition.

1 Introduction

This paper sets out a new direction for language evolution research that brings together three disciplines in a novel cross-disciplinary perspective.

2 Dynamic Syntax

The starting point is from within Linguistics, and the Dynamic Syntax stance (DS: Cann et al., 2005) in which the grammar framework itself sets out the dynamics of how the information growth process is achieved in both parsing and production (Kempson et al., 2016). This tight, system-internal integration is strikingly made evident by informal dialogue exchanges in which speakers and hearers jointly induce structure, fluently and effortlessly switching roles at arbitrary points in an exchange (1)-(9). Interlocutors can each contribute a fragment (1) with the overall content and inferential effect only emerging across parties in a composite group activity. Although these fragments are structurally collaborative, this does not necessitate the recovery of some previously or even subsequently agreed intended propositional content or speech-act (1,8): notably, even very young children are able to join in with appropriate increments in the joint activity of co-creating dialogue (5,6), well before evidence of productive mind-reading capability (Breheny, 2006).

This universal phenomenon, i.e. *split utterances*, is highly problematic for all major grammatical frameworks, with their exclusive emphasis on licensing sentence-string/interpretation pairings as output of the grammar system. As these are taken to model the ideal speaker/hearer's capacity in language, all split utterance data are beyond their remit. They are largely ignored and/or explained as performance disfluencies. This judgement, however, flies in the face of the seamless fluency of the phenomenon in informal conversation. DS apart (Gregoromichelaki et al., 2011, a.o.), the only studies addressing this challenge (Poesio and Rieser, 2010; Pickering and Garrod, 2013) are at best incomplete in modelling only the subset of deliberately 'helpful' completions (2,3); and such accounts involve complex externally imposed operations: e.g. high-level inference (Gibson et al., 2013), abduction (Friston and Frith, 2015) or the creation of efference-copies as in standard models of action control (e.g. Pickering and Garrod, 2013; though see Clark, 2016). Since very young children freely join in on such utterance exchanges, the full complexity of any such mechanisms has to be assumed to be in place prior to the acquisition process if these are to be captured, a view notably embraced in its strongest form by Tomasello and colleagues (2005; 2008) in the form of innate specification of Gricean inference capacities.

In DS, no such commitment is necessary, as the split utterance effects follow directly from the system itself. Production and parsing both involve the top-down anticipation-driven construction of al-

(1) **A:** We're going to ... **B:** Burbage to see Ann, Auntie Ann **C:** with the dogs? **B:** if you look after them.	(2) **Homeowner:** I shall need the ... **Gardener:** mattock. For breaking up clods of earth. [BNC]	(3) **A:** Have all the students handed in **B:** their term papers? **A:** or even any assignments
(4) **Sue:** I'm afraid I burned the kitchen ceiling. **Michael:** Did you burn **Sue:** myself? No, fortunately not.	(5) **Carer:** Old McDonald had a farm... E-I-E-I-O. And on that farm he had a... **Child:** cow.	(6) **Teacher:** And your name is ... **Child:** Mary
(7) **A:** And they ignored the conspirators who were **B:** Geoff Hoon and Patricia Hewitt [BBC Radio 4 06/09/10]	(8) (A and B arguing:) **A:** It's clear from what you've just said that **B:** that I am completely vindicated	(9) (T(herapist) and C(lient)) **T:** Your sponsor before... **C:** was a woman. **T:** Yeah. **C:** But I only called her every three months. **T:** And so your sobriety now, in AA [is] **C:** [is] at a year [Ferrara 1992]

Figure 1: Examples

ternative interpretations directly establishing step by step coordination, with competing emergent interpretations involving probabilistic weightings with consistency checking filtering out errors (Eshghi et al., 2013; Hough and Purver, 2014; Kempson et al., 2015), and positive and negative feedback constraining the searchspace (Eshghi et al., 2015). Online processing is thus modelled as system-internal structural growth (Chatzikyriakidis and Kempson, 2011; Kempson et al., 2016) and not via the grammar plus externally defined parsing/production modules. This captures standard sentence or subsentential-level morphosyntactic phenomena, all the way up to discourse effects such as ellipsis and dialogue data (1)-(9). This is an advantage over other frameworks, in which some (sometimes non-conservative) extensions need to be made.

On the DS view, to the contrary, the interaction emerges from the fact that all parties are using the same structure building strategies. The coordinative effect is a direct consequence of incorporating the dynamics of online processing within the grammar formalism. There is no invocation or presumption of grammar-external inference to achieve the interactivity intrinsic to dialogic exchanges. However, although this ability is freely made use of once it has become available to the language user, it is not a sine qua non for language development. Indeed, even on the assumption that a mind-reading capacity plays a large part in adult cooperativity, it is notable that the assumptions of shared IDENTITY are never guaranteeable. Despite this, the effectiveness of interaction is almost never jeopardised, and even in cases where corrections/clarifications are warranted, successful interactive exchange is buttressed through such overt clarifications.

Recent computational work confirms both the viability of DS as a grammar formalism, its fit to include multimodal data to parse or generate complex dialogue data (e.g. corrections, elliptical fragments, split utterances), and the provision it makes available to enable a large amount of dialogue data to be acquired from very small amounts of unannotated data (actually, just one sentence), using a combination of DS and Reinforcement Learning (Yu et al., 2016; Kalatzis et al., 2016). Earlier work (Eshghi et al., 2013) has shown that an incremental semantic grammar can be acquired using limited data. Both the training and test data are taken from utterances in the CHILDES corpus paired with their logical forms expressed using Type Theory with Records (TTR: Cooper, 2005). The system input is comprised of: (i) a fixed set of computational DS actions (language general structure building mechanisms); (ii) a training set of the form $<S_i,RT_i>$, where $<S_i>$ are sentences of the language and $<RT_i>$ their targeted semantic representations. The output induces lexical actions for the individual words, probabilistically decomposing the possible sequences of actions that lead to the complete target semantic representations.[1] The results suggest that grammar induction of a probabilistic grammar in an incremental, semantic model like DS can be done effectively without prior assumptions of syntactic structure. On a more general level, this might point to the possibility of a set of language independent (and potentially domain general) computational actions that given appropriate data can induce domain specific systems (e.g. lexical actions for individual words).

3 The Predictive Processing Model

This perspective fits directly into a larger cognitive science perspective along two dimensions, that of

[1] The interested reader is directed to Eshghi et al. (2013) for more details.

an integrated nonmodular cognitive system, and that of how language might have evolved as a departure within such a cognitive system. First, we find notable parallels between the Dynamic Syntax perspective and the Clark (2013; 2016) view of cognition modelled as a generative Predictive Process (PPM). PPM equally argues that action and perception act in tandem and invokes neither higher order intentions nor an efference copy construction to yield interactional effects. In PPM, as in DS, context is everything: brains are predictive engines (not passive modular input systems) using their own immediate and encyclopaedic context at every step to guess the structure/shape of the incoming sensory array, which forms 80% of the burden of processing. DS can thus be viewed as a specialisation of the dynamics of the PPM framework, viewing language as a set of action transitions from one context state to another. Context itself is also characterised in process-terms of growth rather than a static store. The addition of DS extends the PPM by adding the dimension of manifest interaction for which language is the central tool. Within this extended model, there are two variants; and on either variant, this nesting of the DS model of language within PPM yields a yet larger perspective – that of language evolution.

Current work on language evolution is split between two extremes. At one end are those who see language as innate and encapsulated. Language is just one amongst a (large) number of modules, each with their specialised niche, requiring some form of glue-language to relate one vocabulary with another; and no adaptationist account is possible (Fitch et al., 2005). At the other extreme are those who see acquisition of language as emerging out of the dynamics of communication, with the panoply of Gricean-style axioms and mind-reading capacities taken to be a necessary prerequisite for acquisition of language, hence innate (Tomasello, 2008; Jaeger, 2007; Christiansen and Chater, 2016). Under all these views, the emergence of language is parasitic on the not unproblematic assumption that language itself is a specialised module, not reducible to more general cognitive architecture. Either type of account requires a significant shift away from a general inferentialist system by some form of switch mechanism to an encapsulated language faculty not expressible in the same terms as the general cognitive system. There are variants between these extremes, amongst which Kirby et al. (2008) argue that compositionality in language can be learned without predispositions, offering a counter-argument to the innateness view of language and its anti-adaptationist stance.

4 The Multilevel Selection Hypothesis

In addition, a combined DS/PPM perspective suggests a view which reflects recent work in evolutionary biology urging a re-evaluation of groups as a unit for evolutionary purposes: the Multilevel Selection Hypothesis (MLSH: Sober and Wilson 1998; Wilson, 2002). On the MLSH view, evolution is seen as driven by two separate dimensions, individual- and group- level adaptivity. The potential of a group to form an adaptive unit turns on the successful balancing of these two conflicting dimensions requiring intra-group pressure to moderate rampant individualism. This dual-level perspective has not so far been taken up within the language evolution narrative, which remains based on individualistic competing selfish considerations. We now explore this in two steps.

Taking first an individual-centred basis DS offers a view of language broadly following (Kirby et al., 2008) in not having to stipulate either rich innate attributes of structure, or externally imposed innate higher-order inference capabilities as pre-requisite to language development, while opening up the potential for an MLSH form of explanation. With DS assumptions, the interactivity displayed by split utterances is seen as emerging from a background of rich interaction between co-participants without any necessity of shared agreed content, as vividly displayed in first language acquisition (Hilbrink et al., 2015). The in tandem co-construction by speaker and hearer of some sound-interpretation pairing is grounded in the already robustly established pattern of situated interactional behaviour between carer and child. The infant's non-language-based verbalising behaviour is interpreted by the carer as contributing to some verbal frame which she herself may have as the basis for engaging with the child in order to create the bonding achievable – even without any signalled content being conveyed (e.g. the peekaboo games which pre-linguistic children so enjoy; Clark and Casillas, 2016). Fragments such as one word utterances initiating the child's emerging language capability are also interpreted against the rich contextualisation of the carer, ei-

ther in interpreting the child's minimal utterance, or in providing a frame relative to which the utterance provides an entirely successful completion (as in (5,6)), building on the pleasure in interaction which the infant and carer already share.

It is then a small second step to see this established interactivity as the basis for a new group-oriented perspective on language evolution. Successful utterance exchanges, even one word utterances, can be seen as achieving the same context-dependent interactional effect displayed by other primates but with the addition of manifest signalling of that interactional effect – from which the step of ascribing content to a signal could have developed (Kirby et al., 2008; Scott-Phillips et al., 2009). The inexorable interactive duplication by all parties in jointly building up the substructure to meaningfully support such utterances yields cumulative interactive effects, multiplied recursively with each additional language token. And with such interactions, repeated reiterations combine with internal cognitive pressure for simplification and cognitive economy and inexorably lead to routinisation effects, with macro sequences of actions becoming stored for ease of recoverability. This leads to recursive buttressing of the group ethos, without ever needing the identity of word tokens or their interpretation to be manifestly confirmed. Hence the uncontentiously effective group-forming trait of language which creates sharp barriers against those who cannot control the stored routinised string-interpretation pairings necessary to achieve the interactiveness that the language makes manifest. Moreover, though the role of mind-reading and explicit seeking of common goals in later stages of language and cognitive development remains an undoubted buttressing force for group consolidation, it no longer plays a role in triggering language emergence: rather, the merely approximate cross-speaker correspondence of string-interpretations set by each participant contributes to gradual language change in the face of cognitive and social pressures.

This contrasts with the Tomasello (2008) account in particular, which claims that the full apparatus of Gricean reasoning has to be innate: "communicative intentions of the cooperative (Gricean) kind [are] clearly a prerequisite for understanding symbols"; and "the idea of language without shared intentionality, even in one-unit expressions, is simply incoherent."(Tomasello et al., 2005, 724). It is notable that the considerable empirical data supposedly confirming this innate cooperativity and desire to be helpful to others, claimed to need dual representation of both speaker and hearer perspective for each individual participant, can all be explained relative to the weaker stance that it is the potential for interaction which is innate, and not a necessity of "shared contents" or "shared goals", with their problematic concept of identity of content. Tomasello et al. (2005) note the potential functionality of shared intentionality at the level of group selection, but do not develop it. On the DS view, the group dynamic IS the story, irreducibly so, in virtue of the characterisation of language as manifest mechanisms for securing online interactive exchange.

Finally, we can now see how the evolutionary advantage of language lies in its adaptivity both at the individual and group level. In contradistinction to both biological and cultural evolution (Sober and Wilson, 1998; Wilson, 2002), in which selfish behaviour is seen as having to be kept within bounds if optimal adaptivity is to be ensured, capacity for language is advantageous both for individual- and group- level adaptivity. It is adaptive for the individual because it enhances potential for interactive and cooperative exchange with others, with individual benefits in such cooperation not achievable without language. It is adaptive for the group because it buttresses group potential for survival in accentuating and distinguishing other competing groups. The claimed relative adaptivity of individual languages in their progressive shift to meeting the brain desiderata of providing input able to be processed fast against ever-evolving contexts (Christiansen and Chater, 2016), can now be replaced with the more appropriate view of languages as differing in the various culturally evolved sets of actions they license, all of them being subject to cognitive constraints associated with the pressures for rapid real-time processing.

References

Richard Breheny. 2006. Communication and folk psychology. *Mind & Language*, 21(1):74–107.

Ronnie Cann, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language*. Elsevier, Oxford.

Stergios Chatzikyriakidis and Ruth Kempson. 2011. Standard modern and pontic greek person restric-

tions: A feature-free dynamic account. *Journal of Greek Lingusitics*, pages 127–166.

Morton H. Christiansen and Nick Chater. 2016. *Creating Language: Integrating Evolution, Acquisition, and Processing*. MIT Press.

Eve V. Clark and M. Casillas. 2016. First language acquisition. In Keith Allan, editor, *The Routledge Handbook of Linguistics*, pages 311–329. Routledge.

Andy Clark. 2013. Are we predictive engines? perils, prospects, and the puzzle of the porous perceiver. *Behavioral and Brain Sciences*, 36(03):233–253.

Andy Clark. 2016. *Surfing uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Robin Cooper. 2005. Records and record types in semantic theory. *Journal of Logic and Computation*, 15(2):99–112.

Arash Eshghi, Julian Hough, and Matthew Purver. 2013. Incremental grammar induction from child-directed dialogue utterances. In *Proceedings of the 4th Annual Workshop on Cognitive Modeling and Computational Linguistics*, pages 94–103. ACL.

Arash Eshghi, Christine Howes, Eleni Gregoromichelaki, Julian Hough, and Matthew Purver. 2015. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Linguistics IWCS*.

Tecumseh Fitch, Mark Hauser, and Noam Chomsky. 2005. The evolution of the language faculty: clarifications and implication. *Cognition*, 97:197–201.

Karl Friston and Christopher Frith. 2015. A duet for one. *Consciousness and cognition*, 36:390–405.

Edward Gibson, Steven T Piantadosi, Kimberly Brink, Leon Bergen, Eunice Lim, and Rebecca Saxe. 2013. A noisy-channel account of crosslinguistic word-order variation. *Psychological Science*.

Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, Greg J. Mills, Ronnie Cann, Wilfried Meyer-Viol, and Patrick G. T. Healey. 2011. Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, 2(1):199–233.

Elma E. Hilbrink, Merideth Gattis, and Stephen C. Levinson. 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of motherinfant interaction. *Frontiers in Psychology*, 6:1492.

Julian Hough and Matthew Purver. 2014. Probabilistic type theory for incremental dialogue processing. In *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, pages 80–88, Gothenburg, Sweden. ACL.

Gerhard Jaeger. 2007. Evolutionary game theory and typology: A case study. *Language*, (3(1)):74–109.

Dimitrios Kalatzis, Arash Eshghi, and Oliver Lemon. 2016. Bootstrapping incremental dialogue systems: using linguistic knowledge to learn from minimal data. In *Proceedings of the NIPS 2016 workshop on Learning Methods for Dialogue*, Barcelona.

Ruth Kempson, Eleni Gregoromichelaki, Arash Eshghi, and Julian Hough. 2015. Ellipsis in dynamic syntax. In Jeroen van Craenenbroeck and Tanja Temmerman, editors, *The Oxford Handbook of Ellipsis*. Oxford University Press.

Ruth Kempson, Ronnie Cann, Eleni Gregoromichelaki, and Stergios Chatzikyriakidis. 2016. Language as Mechanisms for Interaction. *Theoretical Linguistics*, 42(3-4):203–276.

Simon Kirby, Kenny Smith, and Hannah Cornish. 2008. Language, learning and cultural evolution. In Robin Cooper & Ruth Kempson, editor, *Language in Flux: dialogue coordination, language variation, change and evolution*, pages 81–108. College Pubs.

Martin J. Pickering and Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36:329–347, 8.

Massimo Poesio and Hannes Rieser. 2010. Completions, coordination, and alignment in dialogue. *Dialogue and Discourse*, 1:1–89.

Thomas C. Scott-Phillips, Simon Kirby, and Graham R. S. Ritchie. 2009. Signalling signalhood and the emergence of communication. *Cognition*, 113(2):226–233.

Elliott Sober and David Sloan Wilson. 1998. *Unto others: The evolution and psychology of unselfish behavior*. Number 218. Harvard University Press.

Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28:675–691.

Michael Tomasello. 2008. *Origins of Human Communication*. MIT press.

David Sloan Wilson. 2002. *Darwin's cathedral: Evolution, religion, and the nature of society*. University of Chicago Press.

Yanchao Yu, Arash Eshghi, and Oliver Lemon. 2016. Incremental generation of visually grounded language in situated dialogue. In *Proceedings of INLG 2016*, Los Angeles.

Towards Dialogue Acts and Updates for Semantic Coordination

Staffan Larsson
Department of Philosophy, Linguistics and Theory of Science
Gothenburg University, Sweden
sl@ling.gu.se

Jenny Myrendal
Department of Education, Communication and Learning
Gothenburg University, Sweden
jenny.myrendal@gu.se

Abstract

This paper sketches a formal account of semantic coordination, combining parts of two dialogue act taxonomies related to semantic coordination and relating these to meaning updates on an abstract level.

1 Introduction

Semantic coordination is the process of interactively agreeing on the meanings of words and expressions, and (simultaneously) agreeing on which words are appropriate in a given context. Shared meanings are achieved by agents interactively coordinating their respective takes on those meanings (Larsson, 2008).

In this paper, we will sketch a general account of dialogue acts for semantic coordination in dialogue by (1) sketching a synthesis of two existing taxonomies of dialogue acts relating to semantic coordination and (2) relating these dialogue acts to different kinds of updates to (agents takes on) meanings.

2 Dialogue acts for Semantic Coordination

In this section, we will begin to synthesize two taxonomies for dialogue acts related to semantic coordination. While these taxonomies are designed for different settings (first language acquisition and online discussion forums), they nevertheless overlap in interesting ways. By combining and relating them, we hope to eventually provide a more comprehensive overview of the dialogue acts used in semantic coordination independently of setting and domain.

2.1 Dialogue acts for word meaning negotiation

In Myrendal (2015) and Myrendal (submitted), a taxonomy for dialogue acts involved in Word Meaning Negotiations (WMNs) in online discussion forum communication is presented. We here show only parts of the taxonomy. All examples are taken from Myrendal (2015).

Frequently, the question under discussion (QUD) in a WMN concerns whether a certain *trigger expression* T correctly describes a situation S under discussion (what may be called a *SUD* in analogy with QUD). However, in some cases there is no particular SUD, but meanings are negotiated more abstractly.

Explicification[1]**:** Provides an explicit (partial or complete) definition of T. Myrendal (2015) distinguishes between two types of explicifications. **Generic explicifications** foreground the meaning potential of T; a complete or partial definition D of T is provided, but D is not clearly derived from S. For example, Myrendal (2015) shows an example where a DP (Dialogue Participant) is asked to clarify the meaning of *sexism* and in response offers a definition: "That people are treated differently because of their gender."

By contrast, **specific explicifications** foreground conversational context; particular aspects of the SUD S are made explicit and presented as a (typically partial) definition of T. One example is taken from a discussion about whether or not piercing the ears of young children is morally acceptable, or if it constitutes *(child) abuse*: "Clearly ABUSE to pierce the ears of young children! [...] - you inflict pain upon the child and a physical change which the child herself has not chosen and which cannot be made undone."

[1]The term explicification is borrowed from Ludlow (2014), but is adapted and elaborated in Myrendal (2015).

Specific explicifications can also be negative. In one discussion the trigger word *boozing* (Sw. *super*). This discussion is about a woman who is denied alcohol in a restaurant. The bartender refuses to serve the woman a second glass of wine when he notices that she is breastfeeding her baby at the table. The thread starter in this discussion describes the womans behaviour as "boozing" which hich receives the following response: "2 glasses of wine is not boozing and it is not dangerous to drink while breastfeeding."

Exemplification: Providing examples of what the trigger word can mean, or usually means.In a discussion about dietary habits, many DPs state that they prefer to include full fat products in their diet. One DP requests clarification about the meaning of the trigger word ("What counts as full fat?"). Another DP then exemplifies the meaning of the trigger word: "When it comes to dairy products ordinary full cream milk, the fattest cheese and regular double cream (...)".

Similar to (specific) explicifications, exemplification can be negative. In a discussion about fast food, a DP protests against another DP's claim that (all) food from McDonald's is *unhealthy* (T): "Hamburgers with lettuce and water is not especially unhealthy." (Note that in this case the discussion does not revolve around a particular SUD, but rather around a general claim.)

Contrast: A third way of contributing to a WMN sequence is to contrast T against another word C, thus indicating a difference in meaning as well as updating the meanings of both T and C with respect to some example situation or entity.

In a discussion about whether or not it is acceptable to flirt with a married person, after a while it becomes clear that the participant asking this question has a specific situation in mind. The person doing the alleged flirting has expressed strong feelings towards the married person, sending her many text messages and e-mails per week and also sending flowers to her workplace. At this point, one participant objects to the trigger word being used to describe the SUD, and contrasts the trigger word with other words taken to be more suitable descriptions of the situation: "This is pure and utter courtship/picking someone up/declaration of infatuation! This is not how you flirt... at least not how I flirt. This is clearly way way beyond flirting in my world." Here, the behavior is claimed to go beyond "flirting" and to be more accurately described as "courtship", "picking someone up" or "declaration of infatuation".

2.2 Dialogue acts for first language acquisition

Clark and Wong (2002) provide a taxonomy of dialogue acts involved in first language acquisition. We will here describe a subset of this taxonomy. (Note that we will be using some terminology from Myrendal (2015) when describing these acts, even if this is not exactly how they are described in Clark and Wong (2002).)

Direct offers are utterances where speakers offer conventional terms or expressions, and nothing else; the primary function of the utterance is as an offer. Direct offers tend to be made using only a limited set of frames for presenting the term being offered. For example, "That's a pen", "That's called a dentist", "What is this? Chair.", "What's that called? Dancing".

There are also *indirect offers*, where speakers (adults) use their next utterance, whatever it is, to include the term that is simultaneously being offered as a correct form of a term in the addressees' (child's) utterance. We will here concern ourselves with one kind of indirect offer, namely *explicit* ones. In cases of **explicit replace**, a term or expression C is proposed as a replacement for T. An example from Clark and Wong (2002) is the following:

> Naomi: Birdie birdie.
> Mother: Not a birdie, a seal.

Here, "seal" (C) is offered as a replacement for "birdie" (T).

2.3 Towards a synthesis

A basic difference between WMN in online discussion forums (henceforth ODF) as described in (Myrendal, 2015) and first language acquisition (1LA) is that the latter setting typically requires a shared perceptually available situation, whereas ODF pretty much exclude this possibility. Deictic phrases (e.g. "that") in 1LA typically refer to aspects of the shared perceptual situation, whereas in ODF they typically refer to aspects of the situation under discussion, which is only available to DPs through verbal descriptions.

Also, in ODF speakers are assumed to be competent, so attempts at unprovoked teaching of

words (which is frequent in 1LA) are not motivated. Furthermore, ODF interaction is written whereas adult-child dialogues are spoken and arguably more interactive. Despite these differences, we believe it may be interesting to also briefly note some similarities between the respective dialogue act taxonomies for ODF and 1LA.

Firstly, Clark and Wong's **explicit replace** ("that's not an X, that's a Y") is very similar to Myrendal's **contrast**, but where the example is provided by the jointly perceived situation rather than by a verbal description. Secondly, Clark and Wong's **direct offer** is similar to Myrendal's (positive) **specific explicification**, where again the the jointly perceived situation provides the SUD.

For our current purposes, we will simply assume that direct offers can be treated as exemplifications and that explicit replace can be treated (more or less) as contrast. Importantly, doing so requires allowing for jointly observable situations (potentially including subsymbolic information derived from the sensory apparatuses of agents) to serve as the basis for the updates involved in both exemplification and contrast.

3 Meaning representations and updates

A full account of semantic updates involved in WMNs would require capturing the sequential updates at various stages of the negotiation process. Our goals here are more modest, in that we will not consider sequential updates or rejected proposals, but only try to capture isolated updates for *accepted* dialogue acts.

The exact way in which meaning updates are formalised will depend on how meanings are represented. Marconi (1997) distinguishes between inferential meanings of words, which enables to draw inferences from uses of the word, and referential meaning, allowing speakers to identify the objects and situations referred to by the word. We will regard inferential meaning as high-level (symbolic) rules governing inference, e.g. meaning postulates in modal logic or record types (and associated functions) in TTR (Larsson and Cooper, 2009). Secondly, referential meaning may be represented at least in part as low-level (subsymbolic) statistical or neural classifiers of perceptual data (Harnad, 1990; Steels and Belpaeme, 2005; Larsson, 2013; Kennington and Schlangen, 2015). A key insight here is that the step from perception to language can be conceptualised and implemented as the application of a classifier to perceptual data, yielding linguistically relevant classification results as output.

Correspondingly, we may distinguish kinds of meaning updates. High-level structures can be modified e.g. by adding and retracting meaning postulates or "possible languages" (Barker, 2002), or by adding and removing fields in record types representing inferential meanings (Larsson and Cooper, 2009). Low-level aspects of meanings, modeled as classifiers, can be modified by retraining the classifier with new (positive or negative) data.

However, there are also intermediate cases. For example, as shown in the account of vagueness involving comparison classes (Fernández and Larsson, 2014), meanings may involve both high-level (e.g. comparison class for vague terms) and low-level information (e.g. perceived height). Similarly, meaning updates may concern both high-level and low-level information (e.g. perceived height).

We will adopt a fairly abstract formalism for conceptual updates, where we assume that either a full or partial (verbal and hence symbolic/high-level) definition D of the trigger word T has been provided, or alternatively an example situation or entity[2] E (represented using high or low level information, or a combination thereof). D or E is then used for updating the meaning in question.

- δ^+(T, D): T updated with D as a partial definition of T
- δ^-(T, D): T updated with D as a negative partial definition of T
- ϵ^+(T, E): T updated with E as a positive example of a situation described by T
- ϵ^-(T, E): T updated with E as a negative example of a situation described by T

These abstract update operations can then be further specified depending on the semantic formalism used. The abstract meaning update functions thus serve as a sort of API between dialogue acts and their consequent meaning updates. The existence and usefulness of this level of representation remains to be demonstrated in future work; here, we are simply aiming to formulate our account as clearly as possible.

Although it is not explicit in the formalism used here, semantic updates always concern a particular

[2]Insofar as entities can be reified as situations involving them, we need only to talk about example situations.

agent's take on the meaning of the word in question. Meanings become shared by being interactively coordinated. Also, the viability of a semantic update may be limited to a specific dialogue, or it may eventually spread over a community and become part of "the language" (Larsson, 2008).

4 Meaning updates for dialogue acts

In this section, we present an initial characterisation of explicification, exemplification (including direct offers) and contrast (including explicit replace) in terms of the meaning updates described in the previous section.

Note that we are here formalising the update effect of successful (i.e. accepted) meaning updates. In general, proposed updates may not be accepted immediately but can lead to negotiation that may end up with coordinating on proposed update, no update or modified update. Formalising such exchanges is left for future work.

We will sidestep the problem of interpreting verbal definitions by simply using [square brackets] to indicate meanings of linguistic expressions. Updated meanings are indicated by a prime ($'$).

Explicification: By definition, explicifications provide a (full or partial) definition D of T, and the update is thus symbolic (linguistic) in nature which means that only the δ function is needed here.

As mentioned above, in the case of specific explicifications, the definition D is derived by abstraction over the (verbally described) SUD S.

- Generic explicification
  - Update: $T' = D$ (full) or $T' = \delta^+(T, D)$ (partial)
  - Example: ⟦sexism⟧′=⟦that people are treated differently because of their gender⟧
- Specific explicification ($S \sqsubseteq D$)
  - Positive update: $T' = \delta^+(T, D)$
  - Example: ⟦child abuse⟧′ $= \delta^+$(⟦child abuse⟧, ⟦to inflict pain upon the child and a physical change which the child herself has not chosen and which cannot be made undone⟧)
  - Negative update: $T' = \delta^-(T, D)$
  - Example: ⟦boozing⟧′ $= \delta^-$(⟦boozing⟧, ⟦(drinking) 2 glasses of wine (or less)⟧)

Exemplifying Proposes an example E of a situation or entity appropriately (or not, in the case of negative exemplification) described by T. The example can either be given verbally or it can be relevant aspects of the jointly perceived situation (often indicated by a deictic reference ("that")).

- Update: $T' = \epsilon^+(T, E)$ or $T' = \epsilon^-(T, E)$
- Example: ⟦full fat⟧′ $= \epsilon^+$(⟦full fat⟧, ⟦full cream milk⟧)
- Example: ⟦pen⟧′ $= \epsilon^+$(⟦pen⟧, S) where S is a jointly perceivable situation.
- Example: ⟦unhealthy⟧′ $= \epsilon^-$(⟦unhealthy⟧, ⟦hamburgers with lettuce and water⟧)

The last example above shows that the meanings negotiated may sometimes be specific to a domain (here, fast food).

Contrast: Proposes contrasting word C as an appropriate description of an example entity or situation E (as in positive exemplification), and trigger word T as inappropriate (as in negative exemplification).

- Updates: $T' = \epsilon^-(T, E)$, $C' = \epsilon^+(C, E)$
- Example:
  ⟦flirting⟧′ $= \epsilon^-$(⟦flirting⟧, E)
  ⟦courtship⟧′ $= \epsilon^+$(⟦courtship⟧, E),
  where E= ⟦involves expressing strong feelings, sending many texts and emails, and sending flowers to the workplace⟧.
- Example:
  ⟦birdie⟧′ $= \epsilon^-$(⟦birdie⟧, E),
  ⟦seal⟧′ $= \epsilon^+$(⟦seal⟧, E),
  where E is the jointly perceived (by Naomi and Mother) SUD in the example in Section 2.2.

5 Conclusion

We have sketched a formal account of semantic coordination, combining parts of two dialogue act taxonomies and relating these to meaning updates on an abstract level. In future work, we will increase the coverage of the taxonomy, verify and if necessary extend the range of meaning update functions, and show how the meaning update functions can be specified in TTR.

References

C. Barker. 2002. The Dynamics of Vagueness. *Linguistics and Philosophy*, 25(1):1–36.

Eve V. Clark and Andrew D. W. Wong. 2002. Pragmatic directions about language use: Offers of words and relations. *Language in Society*, 31:181–212.

Raquel Fernández and Staffan Larsson. 2014. Vagueness and learning: A type-theoretic approach. In *Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics (∗SEM 2014)*.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1990):335–346.

Casey Kennington and David Schlangen. 2015. Simple learning and compositional application of perceptually grounded word meanings for incremental reference resolution. In *Proceedings of the Conference for the Association for Computational Linguistics (ACL)*, pages 292–301.

Staffan Larsson and Robin Cooper. 2009. Towards a formal view of corrective feedback. In A Alishahi, T Poibeau, and A Villavicencio, editors, *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, EACL*, pages 1–9.

Staffan Larsson. 2008. Formalizing the dynamics of semantic systems in dialogue. In Robin Cooper and Ruth Kempson, editors, *Language in flux - dialogue coordination, language variation, change and evolution*. College Publications, London.

Staffan Larsson. 2013. Formal semantics for perceptual classification. *Journal of Logic and Computation*, 25(2):335–369. Published online 2013-12-18.

Peter Ludlow. 2014. *Living Words: Meaning Underdetermination and the Dynamic Lexicon*. Oxford University Press.

Diego Marconi. 1997. *Lexical competence*. MIT press.

Jenny Myrendal. 2015. *Word Meaning Negotiation in Online Discussion Forum Communication*. Ph.D. thesis, University of Gothenburg.

Jenny Myrendal. submitted. Negotiating meanings online: disagreements about word meaning in discussion forum communication.

Luc Steels and Tony Belpaeme. 2005. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28(4):469–89, August. Target Paper, discussion 489-529.

Gesture meaning needs speech meaning to denote – A case of speech-gesture meaning interaction

Insa Lawler
Department of Philosophy
University of Duisburg-Essen
insa.lawler@uni-due.de

Florian Hahn and Hannes Rieser
Faculty for Linguistics and Literary Studies
Bielefeld University
{fhahn2,hannes.rieser}
@uni-bielefeld.de

Abstract

We deal with a yet untreated issue in debates about linguistic interaction, namely a particular multi-modal dimension of meaning-dependence. We argue that the shape interpretation of speech-accompanying iconic gestures is dependent on its co-occurrent speech. Since there is no prototypical solution for modeling such a dependence, we offer an approach to compute a gesture's meaning as a function of its speech context.

1 Introduction

Speakers often convey multi-modal content by pointing at things or shaping their contours while talking. The semantics of the verbal part is intertwined not only with the communicative situation and the agent's informational situation, but also with the semantics of the non-verbal part. So, one information providing system (gesture) depends on another one (language) for its interpretation. In gesture research, there are at least three claims about how a gesture's interpretation depends on its accompanying speech context: (i) The *classification* of gestures is speech-dependent (see, e.g., (McNeill, 1992; Kendon, 2004; Müller, 2010; Fricke, 2014)). Whether a movement by the index finger is interpreted as drawing a line or as indexing an area in gesture space depends on the respective utterances. Such a movement is likely to be interpreted as *indexing* when the speaker says 'There is my ball,' but it is likely to be interpreted as a *drawing* if the speaker utters 'The path continues for ten miles.' (ii) The *individuation* of gestures is speech-dependent. For instance, it depends on the context whether one interprets an iterative movement as one gesture or as several directly subsequent ones (an example by Lascarides and Stone (2009): 403). (iii) Lascarides and Stone argue that an interpretation of a gesture's meaning does not only depend on its shape, but also on its *rhetorical connection* to its speech context (e.g., (Lascarides and Stone, 2009)). We set these three types of dependencies aside here. Instead, we argue that there is another type of dependence: The meaning of gestures with respect to their *shape* interpretation depends on their accompanying speech. In this paper, we present an approach how to model this particular meaning-dependence of iconic gestures.

2 The meaning-dependence of iconic gestures on their co-occurring speech

The iconic gestures we are concerned with are spontaneous movements of hands or fingers that do not have a lexical meaning. Here, we employ McNeill's conception of a stroke and its semantic synchrony with the accompanying speech (McNeill, 1992), but we acknowledge the idealizations involved in these matters (for treatments of asynchronous strokes, see, e.g., (Hahn and Rieser, 2012)). We take for granted that modeling the meaning of gestures *qua* linguistic signs requires a well-founded concept of meaning and benefits from a formal semantics approach.

Humans do not gesticulate geometrical shapes. If one takes a closer look at roundish-looking gestures, one quickly notices that such gestures are mostly if not always spiral. If a speaker iterates such a sloppy gesture, it looks helix-like. Moreoever, gestures that are intended to be angular are often roundish. This sloppiness is presumably due to the physiological features of humans, time limits, etc. Despite this fact it is common to interpret gestures as conveying meanings like round′ or square′. It seems natural to interpret, say, a roundish gesture as an imperfect sign for the

meaning round′. Roundish gestures can be interpreted as *approximating* geometrical shapes like circles. If so, the gesture's speech-*independent* morphological features alone, such as its hand shape, movements, could provide the core of the gesture's meaning. This view has been (implicitly or explicitly) suggested by authors of formal theories of gesture meaning (which range from employing HPSG (e.g., (Johnston, 1998; Lücking, 2013; Alahverdzhieva and Lascarides, 2010)), to LTAG (e.g., (Kopp et al., 2004)), to λ-calculus (e.g., (Rieser, 2004)), to Montague grammar (e.g., (Giorgolo, 2010)), to SDRT (e.g., (Lascarides and Stone, 2009))[1], and to TTR (e.g., (Lücking, forthcoming)). One might argue for such an approach by suggesting that humans abstract away from the sloppiness while interpreting gestures, since most if not all gestures are sloppy. Sloppiness itself need not pose a problem (apart from the problem of exact depiction). Nonetheless, we found that the sloppiness is the reason for a specific speech-dependence of gesture meaning. In what follows, we argue that the interpretation of a gesture's *shape* is dependent on the meaning of its accompanying speech. Only interpreted in particular contexts are roundish gestures interpreted as meaning round′ rather than angular′.

First, gestures that share all relevant morphological features (i.e., that are of the same type) can be interpreted differently given different speech contexts. If a helix-gesture accompanies an utterance like 'The window is round' it is likely to be taken as meaning circular′ or round′. If it accompanies 'The townhall features a staircase' it is likely to be interpreted as meaning spiral′. Depending on the standard of precision at stake, a roundish gesture might be interpreted as conveying round′ when accompanied by 'ball', but as conveying angular′ when accompanied by 'box'. Such an ambiguity is also found when the sloppiness of the gesture is extreme. Take a look at the examples given in Fig. 1. In Fig. 1a the speaker is uttering 'But not round spiral staircases, but so eh. If the house is rectangular, can the stairs outside be [truncation].' (English translation, gesture stroke underlined) The emphasis on 'rectangular' and the overlapping stroke together with other parts of the dialogue suggests that the speaker employs the gesture to illustrate the shape of the house. Of course, it is also plausible to interpret her gesture as modeling the house, but that seems dispreferred because of the stroke overlap and the content of the overlapping speech. Interestingly, the same speaker uses a similar gesture also in the following speech context: 'And it stands on such a round base?' (see Fig. 1b) Here, it is again plausible that the gesture illustrates a shape. But this time it seems to illustrate roundness. So, we encounter very similar gestures with quite different meanings due to different speech contexts. Our corpus provides more of these examples. The general observation is that one type of gesture (individuated via a similar gesture annotation) can have different gesture meanings when accompanying different utterance segments:

(a) 'rectangular house' (b) 'round base'

Figure 1: Similar gesture morphology, but different meaning

(I) One type of gesture accompanying different utterance segments has different meanings as value.

Second, gestures with a significantly different gesture morphology can represent the same meaning. For instance, different gestures can convey the meaning rectangular′ if they relate to the same utterance segment, etc. Take as examples the ones shown in Fig. 2. In Fig. 2a the speaker utters 'It is just a rectangular building.' Compare this displaying of rectangular′ with Fig. 2b which is identical to Fig. 1a. Although the gestures display some similarity, they are clearly different. Nonetheless, they both seem to mean rectangular′ or angular′. Here, the general observation is that different types of gesture accompanying the same or semantically similar utterance segments can select the same gesture meaning as value:

(II) Different types of gestures accompanying the same utterance segment have one and the same meaning.

(I) and (II) support the idea that the meaning of an iconic gesture is determined to a significant extent by the meaning of its accompanying speech.[2]

[1] Lascarides and Stone employ annotations featuring geometrical shapes, such as circles and cylinders, for their underspecified gesture meanings (e.g., (Lascarides and Stone, 2009): 402, 407, 430, 436).

[2] Our examples feature single words, but our account is

(a) 'rectangular building'

(b) 'rectangular house'

Figure 2: Different gesture morphology, but same meaning

3 Modeling the dependence

This dependence of gesture meaning on speech meaning has not been modeled. The gesture theories mentioned above could only cope with it by substantially underspecifying the gesture's meaning. This would allow the meaning of, say, a spiral gesture to be compatible with utterance segments with conflicting meanings, such as 'round' and 'rectangular'. But this would render the gesture's meaning too weak. It would not allow for recognizing the gesture's contribution to the communicated content and it would not fit the intuition that iconic gestures have a rich meaning on their own. There is also no prototypical solution to be found in other formal semantics: Formal semantics travels the inverse route, so to speak, modeling the context dependence of speech, whereas we model a dependence on speech as context.

A new model of the meaning of iconic gestures should meet at least the following desiderata: (a) The meaning of a gesture is determined to a significant extent by the meaning of the accompanying speech. A similar gesture morphology is *not* sufficient for a similar/identical meaning and a different gesture morphology is *not* sufficient for a different meaning. (b) Nonetheless, its morphology is not irrelevant for determining a gesture's meaning. Not just any gesture can have the meaning round′, for instance, a clearly articulated angular gesture cannot. So, a gesture's meaning is not completely determined by speech. Moreover, gesture content can *contradict* speech meaning. Our corpus has one remarkable instance in which a 'cup-upwards-word' is accompanied by a 'cup-down-wards' gesture.

From a formal point of view, (II) does not present new obstacles over and above those encountered in the context of observation (I). A roundish gesture accompanying, say, 'clock' or 'window' could either be drawn with one index-finger or shaped or modeled with both hands. According to our annotation practices, these would be different gestures, in part due to the different handshapes used. In addition, more subtle differences in terms of gesture morphology could arise. According to the account presented here, the different gestures might all yield ⟦round⟧ if combined with ⟦clock⟧ or ⟦window⟧.[3] Arguments supporting that would have to be given for (I), too.

For (I) our account has to specify the speech-dependent meaning of the gesture. Here is an outline of our approach: The gesture meaning is a *function* of the gesture's initial (topological) meaning based on its morphology and the speech context. The gesture's morphology is described by attribute-value pairs (AVMs) concerning hand shape, movements, etc. One computes the *initial meaning* of the gesture mapping the AVMs onto a logical formula. The *final gesture meaning* is a function of the initial meaning and the speech context. Then, speech meaning and final gesture meaning can be combined to gain a multi-modal proposition (see Fig. 3).

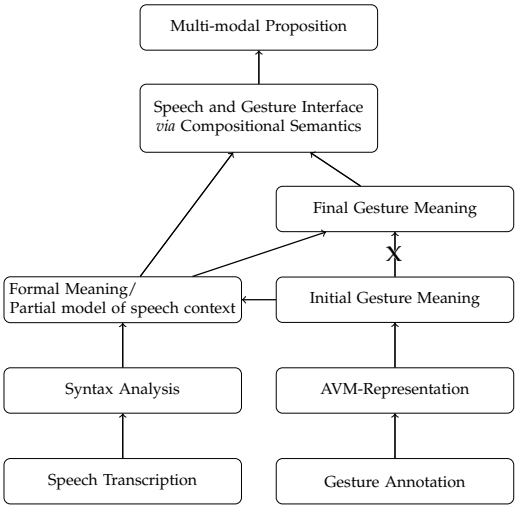


Figure 3: Methodology of our approach

Such an approach is best pursued using a dynamic semantics, because we need a device that is able to model the evolution of the interpretation of gesture processes and speech processes as well as their interaction. The interaction handles compositionality of non-speech and speech meanings. No known static semantics can fulfill such desiderata. We use the ψ-calculus, a recent extension of Milner's π-calculus (see, (Milner, 1999; Johansson, 2010; Rieser, 2015)). ψ has concurrent chan-

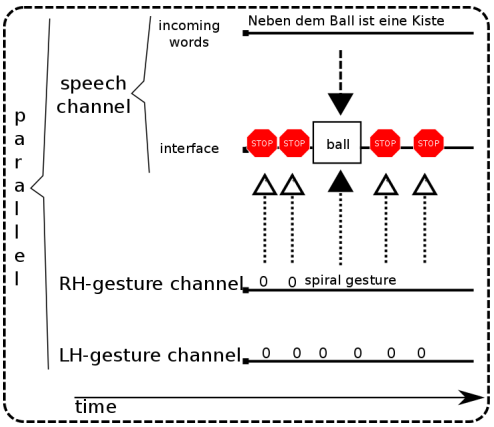
not, in principle, restricted to gesture-word relations.

[3] '⟦A⟧' denotes A's extension; 'A'' the whole meaning.

nels to transmit and process information specified as data structures. Channels are the input-output-devices known from concurrent programming. We represent channels as ψ-operators. They can transport any logic information, such as expressions of a typed λ-calculus and their partial models.

Implementing our approach roughly works as follows: The initial semantics of the gesture formulated in λ-terms is passed onto a channel containing the gesture's speech context. The speech context may modify the gesture meaning in various ways (see, e.g., (I)). Assume that the gesture's initial meaning is spiral$'$, its speech context ball$'$. Roughly, ball$'$ is sent to spiral$'$ which changes it to round$'$ and finally uses it as a modifying information. So, transported and modified meanings are treated in the end as fixed points. In Fig. 4 you can see the basic idea illustrated. The example utterance is 'Neben dem Ball ist eine Kiste.' (Engl.: 'Next to the ball there is a box.') As shown in Fig. 4a, the idea is that a spiral gesture in the context of objects like ⟦ball⟧ and other roundish things designates round$'$ (observe the use of meta-language and object language expressions here which is vital) and $\perp$ (undefined) else. So, the multi-modal meaning of 'ball' + spiral gesture is ball$'(x) \wedge$ round$'(x)$. More specifically, if the partial model input '⟦...⟧' to (2), instantiating bae, yields $z \in \{$⟦circle⟧, ⟦clock-face⟧, ⟦mirror⟧, ⟦sign⟧, ⟦ball⟧, ⟦cup-bottom⟧, ...$\}$ and the projection of spiral$'$, f(spiral$'$), approximates circle$'$ in context c to degree $r \geq$ the threshold in c then round$'$ is substituted for ro, [round$'/ro$], and output on ch$_2$; else $\perp$ is substituted for ro, $[\perp/ro]$, and is output on ch$_2$. The $z \in$ clause and the threshold shall guarantee that not just any gesture can mean round$'$. Gestures accompanying a phrase whose extension is not an element of the set (say, 'square'), as well as gestures that do not approximate a circle to the context-sensitive threshold cannot mean round$'$. The threshold can be determined algorithmically through a simulation device as shown in Pfeiffer et al. (2013) for two-dimensional cases. For three-dimensional cases we still rely on intuition.

This account is not an underspecification account of gesture meaning. We suggest a change of the initial meaning gained from the described morphology. It is triggered by the meaning of the accompanying speech, given that restrictions like the satisfaction of an approximation function hold.



(a) Basic intuition: contact point of spiral gesture and word 'ball'. The spiral gesture + 'ball' yields round$'$, according to (b).

$\underline{ch}_1\ bae\ \overline{ch}_2\ ro\ <\ \lambda z \exists f \exists c \exists r \exists \text{thr}_c\ (\text{spiral}' \wedge \text{approximates}(f(\text{spiral}'), c, x) = r \wedge r \geq \text{thr}_c \wedge \text{circle}'(x) \wedge \text{context}(c) \wedge z \in \{$⟦circle⟧, ⟦clock-face⟧, ⟦mirror⟧, ⟦sign⟧, ⟦ball⟧, ⟦cup-bottom⟧, ...$\}) \rightarrow$ [round$'/ro$][$else$][$\perp/ro$] $> (bae)$

(b) If-else rule for interpreting a spiral gesture in the context of, say, ⟦ball⟧ as round$'$

Figure 4: Modeling with the λ-ψ-calculus.

4 Conclusion and further research

We argued that gestures have a speech-dependent meaning and proposed to model their meanings as a function of the gesture's initial meaning and the speech context employing the ψ-calculus. On account of this, gestures with the same morphology can have even conflicting meanings if they appear in different speech contexts, e.g., we can assign meanings like rectangular$'$ vs. circular$'$ to similar gestures. For future research we aim at integrating the speech context's influence on the gesture classification and individuation as well as the role of rhetorical relations, and at expanding our model for analyzing more complex gestures.

5 Acknowledgements

We are grateful to three reviewers for their critical comments. We tried to accommodate most, but some suggestions, such as the generalization of the model to discourse data (see, however, (Rieser, 2017)) or to the speaker's perspective during the production of co-speech gestures or discussing whether speech disambiguates gesture meaning rather than changing its initial meaning, have to be tackled on another occasion.

References

Katya Alahverdzhieva and Alex Lascarides. 2010. Analysing language and co-verbal gesture in constraint-based grammars. In Stefan Müller, editor, *Proceedings of the 17th International Conference on Head-Driven Phase Structure Grammar (HPSG)*, pages 5–25, Paris.

Ellen Fricke. 2014. Between reference and meaning: Object-related and interpretant-related gestures in face-to-face interaction. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem, editors, *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, pages 1788–1802. De Gruyter Mouton.

Gianluca Giorgolo. 2010. *Space and Time in Our Hands*. UIL-OTS, Universiteit Utrecht.

Florian Hahn and Hannes Rieser. 2012. Non-compositional Gestures. In *International Workshop on Formal and Computational Approaches to Multimodal Communication held under the auspices of ESSLLI 2012*, Opole.

Magnus Johansson. 2010. Psi-calculi: a framework for mobile process calculi.

Michael Johnston. 1998. Unification-based multimodal parsing. In *Proceedings of the 36th Annual Meeting on Association for Computational Linguistics*, volume I, pages 624–630, Montreal, Quebec. ACL.

Adam Kendon. 2004. *Gesture – Visible Action as Utterance*. Cambridge University Press, Cambridge, NY, fourth edition.

Stefan Kopp, Paul Tepper, and Justine Cassell. 2004. Towards integrated microplanning of language and iconic gesture for multimodal output. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 97–104, New York, NY. ACM.

Alex Lascarides and Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics*, 26(4):393–449.

Andy Lücking. 2013. *Ikonische Gesten. Grundzüge einer linguistischen Theorie*. De Gruyter Mouton, Germany.

Andy Lücking. forthcoming. Modeling co-verbal gesture perception in type theory with records. In *Proceedings of the 1st International Workshop on AI Aspects of Reasoning, Information, and Memory*, AIRIM'16. Federated Conference on Computer Science and Information Systems.

David McNeill. 1992. *Hand and Mind. What Gestures Reveal About Thought*. The University of Chicago Press.

Robin Milner. 1999. *Communicating and mobile systems: the π-calculus*. Cambridge University Press.

Cornelia Müller. 2010. Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. *Sprache und Gestik. Sonderheft der Zeitschrift Sprache und Literatur*, 41(1):37–68.

Thies Pfeiffer, Florian Hofmann, Florian Hahn, Hannes Rieser, and Insa Röpke. 2013. Gesture semantics reconstruction based on motion capturing and complex event processing: a circular shape example. In *Proceedings of the SIGDIAL 2013 Conference*, pages 270–297, Metz, France. Association for Computational Linguistics.

Hannes Rieser. 2004. Pointing in dialogue. In *Proceedings of Catalog 04. The 8th Workshop on the Semantics and Pragmatics of Dialogue*, pages 93–101, Barcelona, July.

Hannes Rieser. 2015. When hands talk to mouth. gesture and speech as autonomous communicating processes. In *Proceedings of the 19th workshop on the semantics and pragmatics of dialogue*, pages 122–130.

Hannes Rieser. 2017. A Process Algebra Account of Speech-gesture Interaction. Revised and Extended Version. In *International Workshop on Formal Approaches to the Dynamics of Linguistic Interaction held under the auspices of ESSLLI 2017*, Toulouse.

Language Contact: Peaceful Coexistence or Emergence of a Contact Language

Jérôme Michaud
University of Edinburgh, SOPA
Peter Guthrie Tait Road
EH9 3FD Edinburgh, UK
jerome.michaud84@gmail.com

Gerhard Schaden
Université Lille SHS
CNRS UMR 8163 STL
59000 Lille
gerhard.schaden@univ-lille3.fr

Abstract

This paper presents a simple model of linguistic priming between languages in contact, based on the *utterance selection model* (USM) for language change of Baxter et al. (2006). It will be shown that the emergence or the non-emergence of a new contact language depends on the way potentially bilingual agents choose a language to communicate.

1 Introduction

One major factor driving language evolution is the interaction of its speakers. In our paper, we consider a situation where speakers of two different communities are in contact, and where (at least some of) the speakers of the two groups need to communicate with one another. There are basically two ways of resolving the communicative problem in such cases: speakers can either use (some variant of) their community languages or a contact language can emerge — which corresponds to neither of the two community languages. This new language, which can take the form of a pidgin, is not random and highly correlates with the two languages it originates from. The fine-grained processes controlling this process are poorly understood. In this paper, we provide a simple computational simulation of the stochastic dynamics of a contact situation. We show that the way agents choose a language when they interact partly controls the emergence of a contact language.

In order to capture the stochastic aspects of linguistic interactions, Baxter et al. (2006) designed the *utterance selection model* (USM) for language change (see also Croft, 2000). This is a stochastic agent-based model that accounts for the evolution of a single (socio-)linguistic variable (Tagliamonte, 2011), which can be instantiated in a finite number of equivalent variants. USMs can be seen as formal models of what Calvet (1999) calls an *ecolinguistic system*. The USM is well-adapted to capture the dynamics of a single linguistic variable and its stochastic evolution. It can also be used to predict the evolution of a linguistic variable in a larger population using coarse-graining techniques as shown in Michaud (2017). Other modelling methods, such as the model of Tria et al. (2015), can accurately reproduce the conditions under which a creole emerges based on census data. However, their model is highly idealized and makes assumptions such as "*if the hearer does not already possess the language of the utterance in her repertoire and therefore cannot make sense of it, she learns it by adding it to her repertoire*" (Tria et al., 2015, p. 6), which do not seem very realistic. Our aim is to provide a (still very simple) model, but whose agents correspond more closely to 'real' humans' capacities.

We study a simple extension of the USM that models potentially bilingual agents and explicitly takes into account a priming effect between the two languages to model a situation of language contact. In particular, we study how the choice of a specific language in the interaction can lead either to the coexistence of the two group languages or to the emergence of a new contact language.

2 Methodology

Our model is an extension of the USM for language change (Baxter et al., 2006) that takes into account potentially bilingual agents and models a priming effect between a group language and a non-group language. Below, we recall the definition of the USM and then explain the modifications made to model potentially bilingual agents. We conclude this section by explaining how we measure this stochastic system and explain how we decide when a contact language emerges.

2.1 The standard USM

The USM models the evolution of the usage frequency of a linguistic variable with V equivalent variants. The probability distribution over the different variants of an agent i is represented by the vector $\boldsymbol{x}^{(i)}$, where a component $x_v^{(i)}$ represents the probability/frequency with which agent i uses variant v.

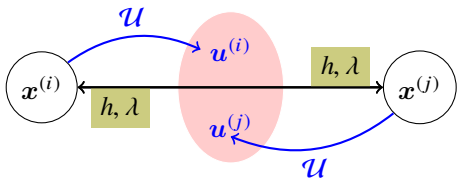


Figure 1: Structure of the USM interaction. On the chosen edge, the agents use their usage frequency vector $\boldsymbol{x}$ to produce an utterance $\boldsymbol{u}$ through the process $\mathcal{U}$, which depend on the matrix M. The utterances are then used to update the internal beliefs depending on the parameters h and λ (see below).

In order to communicate, an agent i produces an utterance $\boldsymbol{u}^{(i)}$ of length L from a production process $\mathcal{U}$ ($\boldsymbol{u} := \mathcal{U}\boldsymbol{x}$). The process $\mathcal{U}$ is defined by

$$\mathcal{U}\boldsymbol{x} = \frac{1}{L} M \text{Multi}(L, \boldsymbol{x}), \tag{1}$$

where M is a matrix representing production errors or innovations and $\text{Multi}(L, \boldsymbol{x})$ is a vector counting the outcome of L multinomially sampled variables.

During an interaction, two connected agents are randomly selected. Then, they both produce an utterance $\boldsymbol{u}$ and update their usage frequency distribution $\boldsymbol{x}$ by

$$\boldsymbol{x}^{(i),\text{new}} = \boldsymbol{x}^{(i),\text{old}} + \delta\boldsymbol{x}^{(i)}, \tag{2}$$

where the increment $\delta\boldsymbol{x}^{(i)}$ is defined by

$$\delta\boldsymbol{x}^{(i)} = \lambda[(1-h)\boldsymbol{u}^{(i)} + h\boldsymbol{u}^{(j)} - \boldsymbol{x}^{(i),\text{old}}], \tag{3}$$

where λ is a usually small learning parameter and the attention parameter h controls the relative importance of the utterance $\boldsymbol{u}^{(j)}$ of the other agents with respect to her own utterance $\boldsymbol{u}^{(i)}$. The presence of $\boldsymbol{u}^{(i)}$ accounts for a *self-monitoring* process and the presence of $\boldsymbol{u}^{(j)}$ accounts for an *accommodation* process. This learning rule assumes communicative success.[1]

[1]One way of interpreting this is to assume that the context and non-verbal communication provide sufficient clues for the interpretation.

The USM has been used to study the conditions under which a consensus can be achieved in a population (Baxter et al., 2006; Michaud, 2017). It has been applied to test the hypothesis of Trudgill about the emergence of New Zealand English (Baxter et al., 2009) and to test under which conditions the time series of usage frequency of an innovative variant takes the form of an S-shaped curve (Blythe and Croft, 2012).

2.2 Bilingual agents, social structure and priming

In order to model a language contact situation, the USM needs to take into account the possibility that agents become bilingual. We assume that each agent belongs to a group labelled by capital letters ($A, B, \ldots$) and every agent knows the group membership of every other agents. An agent belonging to some group Y is able to represent two languages and we denote the corresponding frequency vectors $\boldsymbol{x}_Y$ for the group language Y and $\boldsymbol{x}_{\bar{Y}}$ for the non-group language $\bar{Y}$. With this modification, the utterance production and learning rules have to be adapted.

During an interaction, if two agents belong to the same group, they interact as usual using the standard USM production and learning rules. If the two agents belong to different groups, we consider two scenarios:

Scenario 1: Symmetric adaptation When two agents of different groups interact, they both adapt to the other agent. For example if agent i of group A and agent j of group B interact, they both use their non-group language, i.e. $\bar{A}$ and $\bar{B}$, respectively.

Scenario 2: Unilateral adaptation When two agents of different groups interact, *for each interaction* they randomly choose a group language to use, either A (with probability p) or B (with probability $1-p$), and the agent who doesn't know the language uses his non-group language. For example, if agent i belongs to group A and agent j belongs to group B, then one language is chosen randomly, say language of group A, then agent i uses her group language and j her non-group language $\bar{B}$.

When an agent uses her group language, her knowledge of the language is assumed to be perfect and she uses the corresponding frequency vector. However,

when an agent needs to use a non-group language, her knowledge is only partial and the non-group language is primed by the group language. This priming is implemented by the rule that whenever a non-group language has to be used, instead of using the frequency vector $\boldsymbol{x}_{\bar{A}}$ purely, the group language frequencies modifies the distribution through

$$\boldsymbol{x}_{\bar{A},\text{eff}} = (1-\rho)\boldsymbol{x}_{\bar{A}} + \rho\boldsymbol{x}_{A}. \tag{4}$$

The priming parameter ρ models the degree of mixing between languages A and $\bar{A}$. If $\rho = 0$, then there is no priming and the effective frequency vector boils down to $\boldsymbol{x}_{\bar{A}}$ and if $\rho = 1$, then priming is total and the effective frequency distribution $\boldsymbol{x}_{\bar{A},\text{eff}} = \boldsymbol{x}_{A}$. In the production rule (1), it is the effective frequency vector $\boldsymbol{x}_{\bar{A},\text{eff}}$ that is sampled. The learning rule (2) is the same but is only applied to the languages associated with the interaction.

The social structure used in our model is made of two random regular graphs of degree 3, containing 20 agents each, connected with each other by 5 connexions, see Fig. 2. The agents situated at an end of an intergroup connexion are the potentially bilingual agents, the other agents are monolingual, since they never use their non-group language.

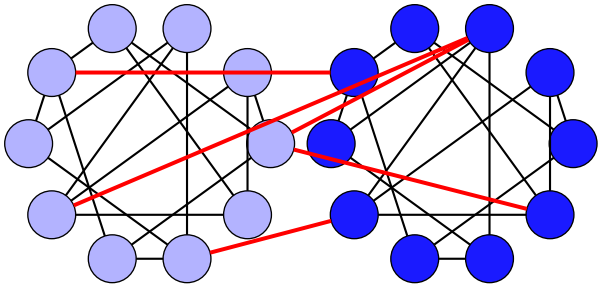


Figure 2: Illustration of the social structure. In the figure the groups have 10 agents instead of 20; red links are intergroup links.

2.3 Measuring the outcome of simulations

We measure the outcome of the simulation by computing Pearson's correlation coefficient between the time series of the averaged use of a language by each group. Note that the non-group languages are only used by agents with intergroup connexions and only these agents are updating their non-group language and can, therefore, become bilinguals.

We introduce the following notation: r_{XY} correlation between language X of group A and language Y of group B, illustrated in Fig. 3. If r_{XY} is close to 1, then the two languages can be considered as being the same. If r_{XY} is close to 0, the two languages are independent. For medium values of r_{XY} the languages are different but correlated.

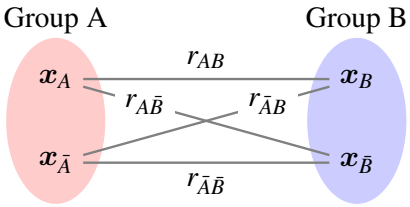


Figure 3: Illustration of the correlation coefficients between the different languages of the two groups.

3 Results

For the simulation of the two scenarios, we used the network topology discussed in Sec. 2.2 and illustrated in Fig. 2. The parameters are $N = 40$ agents with 5 intergroup connexions, the number of variants is $V = 3$, and the utterance length is $L = 2$. The learning parameter $\lambda = 0.1$ and the attention parameter $h = 0.5$. The matrix M used to simulate errors and innovations is of the form

$$M = \begin{bmatrix} 1-q & 0 & q \\ q & 1-q & 0 \\ 0 & q & 1-q \end{bmatrix}, \tag{5}$$

where $q = 3 \times 10^{-4}$. The structure of this matrix is such that the innovations are ordered and variant 1 can only be transformed into variant 2, but not into variant 3, and similarly for the other variants. The pattern of mutation/innovation should be read columnwise. The simulations have been performed for $T = 5000$ full network updates and the priming parameter ρ is varied.

In Scenario 1, two interacting agents of different groups used their non-group language. Results are displayed in Fig. 4 and we observe that the correlation between $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ (r_{AB}) is close to zero for all values of the priming parameter ρ; the correlation between $\boldsymbol{x}_{\bar{A}}$ and $\boldsymbol{x}_{\bar{B}}$ ($r_{\bar{A}\bar{B}}$) is close to one for all values of the priming parameter ρ; the other correlation coefficients grow from 0 to about 0.7 when ρ is increased. From these results, one can conclude that there are three languages in these settings, the language of group A, the language of group B, and a new contact language $\bar{A} = \bar{B}$ partly correlated with both languages.

In Scenario 2, when two agents of different groups interact, at each interaction, they choose language A with probability p and language B

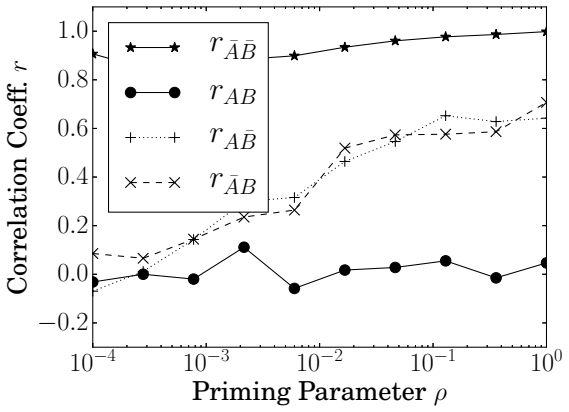


Figure 4: Scenario 1: Emergence of a contact language. The 3 groups of curves represent the different languages.

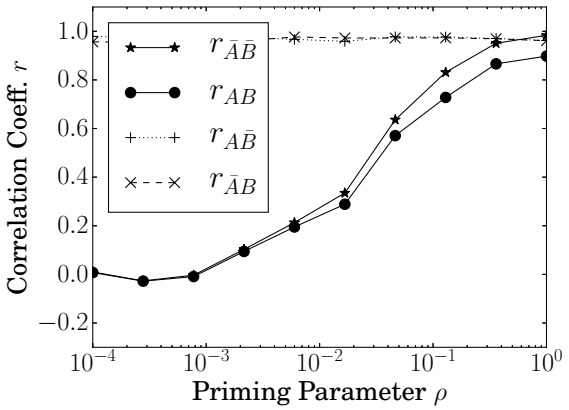


Figure 5: Scenario 2: No emergence of a contact language. The 2 groups of curves represent the different languages merging for high ρ.

with probability $1 - p$. Here $p = 0.5$ and the two languages are equivalent. Results are displayed in Fig. 5 and we observe that $r_{A\bar{B}}$ and $r_{\bar{A}B}$ are close to one for all values of ρ and the other correlation coefficients increase from zero to one as ρ increases. In this situation, there are only two languages present, the two group languages. When ρ is large enough, the two languages converge to the same language and there is a single language remaining.

4 Discussion

We have shown that the decision of which language to use has an important impact on the outcome of the simulation, and can lead either to the emergence of a contact language, or to the stable cohabitation of the two group languages. Compared to the naming game model of Tria et al. (2015), the agents of our model do not instantaneously learn or forget a language but gradually adapt their behaviour. As a result, the emergence of a contact language, or absence thereof, is more gradual and better accounts for the influence of the stance that agents take during intergroup communication.

Our model makes a number of idealising assumptions. First of all, we assume that there is no reason to choose one language rather than the other for intergroup communication — which implies the absence of any hierarchy between the languages (or groups). This is probably a rather rare setting in the wild. There are different degrees of divergence from this configuration: instead of a perfectly symmetric situation with a probability $p = 0.5$ for using each language, there may be a different p tilted towards one group language. In extreme cases, if $p = 1$ or 0, or when the priming parameter $\rho = 1$, the agents of one group do not adapt to the language of the other group at all. Therefore, their group language will always be used, forcing the agents of the other group to adapt. Furthermore, in our model, the preferences and attitudes of the agents as well as the network structures do not evolve over time (bilinguals cannot switch group allegiance, etc.).

That being said, in which circumstances of real-life language contact would we expect the two scenarios we have considered to arise? Notice first that the asymmetric scenario should have a lower cognitive cost than the symmetric one, since only one agent in an intergroup interaction needs to adapt his behaviour, whereas scenario 1 requires both agents to do so. Using this argument, scenario 2 should be preferred overall and no contact language should emerge. One can also argue that an asymmetric scenario will take longer to reach a consensus through the population. As a consequence, if the pressure for communication is strong enough, the more costly, but more rapidely converging scenario 1 would be preferred and a contact language is likely to emerge. The additional cognitive cost of a symmetric adaptation should be partly compensated by the fact that contact languages are usually simpler than fully-fledged languages.

To conclude, scenario 1 is expected if communication pressure is strong and the group languages are unrelated. Otherwise, we expect scenario 2. This is consistent with the conclusions of Tria et al. (2015) concerning the influence of population structure on communicative needs and creole-formation.

Acknowledgments

We would like to thank the three anonymous reviewers for their comments on a previous version of the paper. We would also like to thank the members of the project "Parallel Evolutions", on whom we inflicted a first version of this paper. All remaining errors and omissions are ours alone.

References

Gareth J. Baxter, Richard A. Blythe, William Croft, and Alan J. McKane. 2006. Utterance selection model of language change. *Physical Review E*, 73(4):046118.

Gareth J. Baxter, Richard A. Blythe, William Croft, and Alan J. McKane. 2009. Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, 21(02):257–296.

Richard A. Blythe and William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language*, 88(2):269–304, June.

Louis-Jean Calvet. 1999. *Pour une écologie des langues du monde*. Plon, Paris.

William Croft. 2000. *Explaining language change: An evolutionary approach*. Pearson Education.

Jérôme Michaud. 2017. Continuous time limits of the utterance selection model. *Phys. Rev. E*, 95:022308, Feb.

Sali A. Tagliamonte. 2011. *Variationist sociolinguistics: Change, observation, interpretation*, volume 39. John Wiley & Sons.

Francesca Tria, Vito D.P. Servedio, Salikoko S. Mufwene, and Vittorio Loreto. 2015. Modeling the emergence of contact languages. *PloS one*, 10(4):e0120771.

Amplifying signals of misunderstanding improves coordination in dialogue

Gregory Mills
Centre for Language & Cognition Groningen (CLCG)
University of Groningen
Netherlands
g.j.mills@rug.nl

Gisela Redeker
Centre for Language & Cognition Groningen (CLCG)
University of Groningen
Netherlands
g.redeker@rug.nl

Abstract

We report a dialogue task which investigates how the mechanisms of miscommunication contribute toward referential coordination. Participants communicate via a text-based instant messaging tool which is used to identify turns that were edited prior to sending. These turns are transformed by the server into artificial self-corrections, and sent to the participants. The patterns observed in the dialogues show that these interventions have a beneficial effect on referential coordination.

1 Introduction

A central finding in research on dialogue is that interlocutors rapidly converge on referring expressions (Krauss and Weinheimer, 1966; Clark, 1996), which become progressively, contracted, systematized and abstract. This occurs for a wide range of referents, e.g. when describing spatial locations (Garrod and Doherty, 1994), music (Healey et al., 2007), conceptual structures (Schwartz, 1995; Voiklis, 2012), confidence (Fusaroli et al., 2012), temporal sequences (Mills, 2011; Verhoef et al, 2016), and also when describing how to manipulate physical objects (Shirozou, 2002). Systematization of referring expressions also occurs across modalities - in spoken interaction (Pickering and Garrod, 2004), text-based interaction (Healey and Mills, 2006) and in graphical, mediated interaction (Healey, 2001).

The development of systematicity is not simply due to the coordination problem of creating a novel referring expression: once referring expressions have been used successfully, they continue to develop (Garrod, 1999; Healey, 2004). This pattern is observed both when interlocutors are faced with the task of describing unfamiliar referents using novel referring expressions (Galantucci, 2005), as well as in situations where interlocutors already possess referring expressions and concepts that are sufficient for uniquely individuating the referents (Pickering and Garrod, 2004). Even when the names of the referring expressions are given experimentally, as in the map task (Anderson et al., 1991), interlocutors coordinate on the semantics of their referring schemas (Larsson, 2007).

Cumulatively, these findings suggest that processing that occurs in dialogue places important constraints on the semantics of referring expressions. However, there is currently no consensus about how best to account for how convergence develops. The iterated learning model of Kirby et al (2002) explains convergence as arising out of *individual* speakers' cognitive biases - simply being exposed to another's linguistic output should yield more abstract descriptions. The interactive alignment model (Pickering and Garrod, 2004) proposes that convergence arises as a consequence of mutual priming and alignment, while the collaborative model of Clark (1996) emphasizes the role of positive feedback. One central problem with these accounts is that the basic mechanisms they propose are inherently conservative (Healey, 2004). Once a particular form is the most successfully and widely used by members of a group, there is no mechanism to explain how it might be supplanted by another. Yet interlocutors continue to develop more systematized descriptions throughout the interaction.

Further, a series of experiments (Healey and Mills, 2006; Mills and Healey, 2008) suggest that the development of abstraction can be driven by participants encountering and resolving problematic understanding. In these experiments, participants played an online version of the maze game (Pickering and Garrod, 2004) and

communicated via an experimental chat-tool which inserts artificial clarification requests into the interaction. The clarification requests appear, to participants, to originate from each other. For example in the following conversation between two participants A and B , the second turn "row?" is an artificial turn produced by the server, but appears to originate from participant B.

A: Go to the 3rd row 1st box
B: row? *(produced by the server)*
A: yeah from the top

When participants received these interventions, they produced less abstract descriptions. However, once the interventions stopped, participants subsequently used more abstract descriptions than participants who had received no interventions (Mills, 2015).

In a subsequent experiment (Healey, Mills, Eshghi, 2013) , this methodology was used to automatically detect naturally occurring clarification requests and transform them into more severe signals of miscommunication. For example in the following conversation between two participants A and B, B's clarification request "5th?" is intercepted and transformed into "what?" and sent to A.

A: go to the 5th row 2nd square
B: 5th? *(intercepted by server)*
B: what? *(transformed turn sent to B)*
A: yeah from the top

Notice that this transformation reduces the diagnostic specificity of the clarification request; A has less evidence of B's level of (mis)understanding. Since there is an expectation that a conversational partner should provide diagnostic information that is sufficient to resolve misunderstanding (Clark, 1996), this manipulation makes it appear to A that B is experiencing more difficulty than is actually the case. Participants who received these artificially amplified clarification requests also converged on more abstract descriptions than participants in a baseline condition.

Taken together, these results suggest that (1) When interlocutors encounter problematic understanding, they initially decrease the level of abstraction of their referring expressions, allowing them to identify and diagnose the nature of the misunderstanding, and (2) Once the problem has been resolved, this subsequently allows the participants to coordinate on even more abstract and systematized referring expressions.

However, these experiments have focused solely on "trouble" that is signalled in clarification requests about the content of another's turns, i.e. in "other-initiated" repair (Schegloff, 2007). It is currently unclear whether negative evidence in self-repair might also have an effect on the development of abstract referring conventions.

2 Method

To investigate in closer detail how negative evidence might contribute toward convergence, we report a variant of the maze-task. Here too, participants communicate with each other via an experimental chat tool which automatically transforms participants' private turn-revisions into public self-repairs that are made visible to the other participant. For example, if a participant, A types:

A: Now go to the square on the left, next to the big block on top

and then before sending, A revises the turn to:

A: Now go to the square on the left, next to the third column

The chat server automatically detects the left-most boundary of the edited portion of the turn and inserts a hesitation marker (e.g. "umm" or "uhhh" immediately preceding the revision), followed by the text that was deleted. This would yield the following turn, sent to B: :

A: Now go to the square on the left, next to the big block on top umm..I meant next to the third column

Two self-repair formats were used:

- A: original turn + hesitation marker + reformulated turn
- A: original turn + hesitation marker + ``I meant'' + reformulated turn

3 Results

Interventions were performed symmetrically on both members of a dyad. No participants reported detecting the experimental manipulation. Examining the transcripts showed that participants who received these transformed turns used more abstract Cartesian location descriptions than participants in a baseline condition. This pattern was already apparent after 5 minutes in the task. Task performance followed a different pattern initially participants who received these interventions performed worse completing fewer mazes and requiring more moves to solve each maze. However, by the end of the task, participants who received the interventions performed at the same level as participants in the baseline condition. Crucially, participants who received the transformed turns continued to use more abstract descriptions.

4 Discussion & Conclusions

The patterns observed in the maze game dialogues show that the interventions have a beneficial effect on semantic coordination. However, it is currently unclear how the constituent components of the self-repairs contributed: It could be that this effect is due entirely to the hesitation markers. Conversely, it is possible that this effect is due solely to participants reading the deleted text. If so, it is possible that the deleted text provides additional information about the other's level of (mis)understanding. It could also be that the deleted text makes the dialogue *less* coherent, forcing participants to compensate for the perturbation caused by the interventions.

Since participants encountered multiple interventions per trial, it is not possible to distinguish between the effects of the individual components. However, in aggregate we argue that the artificial self-repairs having a beneficial effect of amplifying naturally occurring signals of miscommunication: the artificially generated disfluencies and reformulations are used by participants as cues that their partner is having difficulty coordinating on the semantics of referring expressions. Consequently, participants expend more effort to address these problems and once these problems have been identified and resolved, dyads are able to converge quicker on more stable and more abstract referring schemas.

5 References

Clark, H. H (1996). Using language. 1996. Cambridge University Press: Cambridge

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tyln, K. (2012). Coming to terms quantifying the benefits of linguistic coordination. Psychological science

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. Cognitive Science, 29(5), 737.

Garrod, S. C., Anderson, A., (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. Cognition, 27, 181218

Garrod, S., & Doherty, G. (1994). Conversation, coordination and convention: An empirical investigation of how groups establish linguistic conventions. Cognition, 53(3), 181-215.

Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. Cognitive Science, 31(2), 285-30

Healey, P. G. (2004). Dialogue in the degenerate case? Peer commentary on Pickering & Garrod (2004). Behavioural and Brain Sciences, 27(2), 201.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure. In Simulating the evolution of language (pp. 121-147). Springer London.

Krauss, R. M., Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. JPSP, 4, 343346.

Larsson, S. (2007). Coordinating on ad hoc semantic systems in dialogue. In Proceedings of DECALOG

Mills, G. J. (2011). The emergence of procedural conventions in dialogue. Proceedings of the Cognitive Science Society (pp. 210-211).

Pickering, M. J., Garrod, S. (2004).Towards a mechanistic psychology of dialogue. Behavioural and Brain Sciences, 27(2), 169190.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Cognition, 141, 52-66.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. The Journal of the Learning Sciences, 4(3), 321-354.

Shirouzu, H.,, Miyake, N., Masukawa, H. (2002). Cognitively active externalization for situated reflection. Cognitive science, 26, 469501

Verhoef, T., Walker, E., Marghetis, T., (2016) Cognitive biases and social coordination in the emergence of temporal language. Proceedings of Cog Sci

Towards a Formal Semantics of Verbal Irony

Julian J. Schlöder
Institute for Logic, Language and Computation
University of Amsterdam
julian.schloeder@gmail.com

Abstract

This paper presents a formal semantics of verbal irony in assertions. In particular, it makes precise what is meant by the common intuition that an ironic utterance expresses *the opposite* of its literal meaning. We start by considering cases of verbal irony that are marked by a particular prosodic tune in English. We then demonstrate that an extant model for intonational meaning can be extended to capture ironic prosody. Afterwards, we discuss how to expand this semantics to cases of irony that are not marked by prosody.

1 Irony and Prosody

The goal of this paper is to formally model some intuitions about *verbal irony*. We first approach the topic as a problem in the *semantics of intonation* and assign a semantics to one particular prosodic *tune* that appears to mark an utterance as ironic. This semantics in particular specifies how to compute *the opposite* of the literal content of an ironic utterance. We then investigate how this semantics could generalise.

Formalising intonational meaning faces many difficulties. One central problem is that tunes fall on a *spectrum* that resists comprehensive sorting into *discrete* categories. This presents a problem for symbolic approaches in general, as then intonational meaning also resists discretisation (Ladd, 1980; Calhoun, 2007). However, intonation *can* be studied formally by considering clear, exaggerated tunes where the intuitions about the associated meanings are uncontroversial (see e.g. Steedman (2014), Schlöder and Lascarides (2015)).

Empirical data suggests that irony is linked to *prosodic contrasts* (Bryant, 2010). In this paper we will consider one such tune: a *steep fall* followed by a *stretched, sustained low pitch*, which robustly leads to ironic readings. We will annotate this tune with a downward arrow ↓ at the fall. Example (1) with the tune in Fig. 1 is an example of an ironic utterance with this tune (note in particular the stretched vowel in "so").[1]

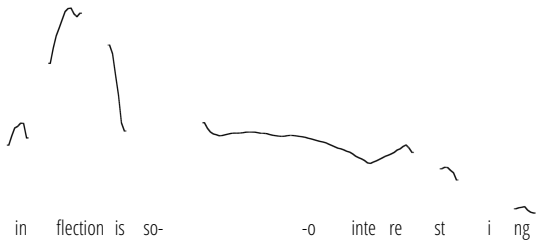


Figure 1: Tune of example 1.

(1) a. Inflection is ↓so-o interesting.
⇝ *inflection is very uninteresting*.

The significance of prosody with respect to irony can be appreciated by considering a minimal pair where the tune of an utterance makes the difference between acceptance and rejection. (2) is one such case (↗ marks a high pitched accent).

(2) a. A: Are you going to Mike's show tonight?
b. B: I'll ↗definitely go to that. ⇝ *will go*
b.′ B: I'll ↓de-finitely go to that. ⇝ *won't go*

Tune is the only variable that distinguishes (2b) from (2b′). The goal of this paper is to isolate a semantics (in the sense of Ladd's (1980) *intonational lexicon*) for the tune of (2b′) that makes the right predictions. Note that we are not claiming that the tune in (1) and (2b′) is the *only* tune that marks irony—or that irony requires *any* particular intonation. We return to this in section 4.

2 Irony and Negation

There are many competing explanations of verbal irony, but they—by and large—revolve around a

[1]The scene containing (1) can be listened to at https://www.youtube.com/watch?v=zIavvxoqxvs

common intuition: that the speaker of an ironic utterance means *the opposite* or *the inverse* of the literal content of their utterance.

On a Gricean account, verbal irony is the flouting of the Maxim of Quality, i.e. the speaker asserts *something recognisably false* and therefore means the opposite; on the *echoic* account an ironic utterance *mentions* a sentence and indicates *dissent* from it (Sperber and Wilson, 1981); on the *joint pretense* account, an ironic utterance invites one's interlocutors to consider a situation in which the utterance would be true and notice how absurd this situation is (Kumon-Nakamura et al., 1995); our own model will follow Martin (1992) in considering irony to be a form of *implicit negation* (also see (Giora, 1995)).

Schlöder and Lascarides (2015) present a formal model of intonation that already goes some way towards a semantics of irony; they assign a semantic term to ironic tunes that expresses *dissent from the literal content of an utterance* and allow this dissent to be strengthened to *assent to the negated utterance*. The predictions of their account are too weak, however. It is not sufficient to just add a negation to the ironic utterance's propositional content. Recall (2b′).

(2) b.′ B: Yeah, I'll ↓de-efinitely go to that.

The negation of the literal content of (2b′) is *it is not the case that B will definitely go to the show* which resolves to *B might not go to the show*. The correct reading is however that *B will (definitely) not go*. Giora (1995) can potentially account for this: she makes the *implicit* negations stemming from irony be subject to a preference for *contrary* negation over *contradictory* negation (Horn, 1989, chs. 4–5). This is in particular realised as a preference for narrow over wide scoped negation: *definitely not* is contrary to *definitely*, but *not definitely* is contradictory to *definitely*.

This preference, however, must be formalised to be predictive. One at least has to indicate how to select from the multiple different possibilities to form a contrary negation. In the case of (2b′), at least two such readings are available.

(3) a. B will definitely not go to Mike's show.
b. B will definitely go somewhere that is not Mike's show.

Both (3a) and (3b) are contrary to the literal content of (2b′). However, (3b) overstates what B expresses in (2b′) because it entails that B will go somewhere—but this does not seem to be part of the meaning of (2b′). It is unclear what privileges (3a) over (3b) (or other options) if all we have is a general preference for contrariness.

While there might be ways to spell this out, there is another option. The *position of the fall* in the ironic tune we are considering seems to indicate the placement of the negation. For instance, (3b) would be the appropriate interpretation for a third possible answer to (2a).

(4) a. A: Are you going to Mike's show tonight?
b.″ B: Yeah, I'll definitely go to ↓tha-at.

(4b″) is appropriate in a context where there is a salient alternative activity that B could attend and the context moreover suggests that B would prefer this one (e.g. because there is a much better show overlapping with Mike's).

The pattern that the contrary negation is scoped on the word immediately following the fall seems to be robust. Consider some variants of (1).

(5) a. Inflection is ↓so-o interesting.
a.′ Inflection is so ↓i-interesting.
a.″ ↓Infle-ection is so interesting.

One can equally well place the fall before the intensifier or before the predicate to obtain the contrary negation *intensely uninteresting*; (5a″) is felicitous, but gets an additional implicature like in (4b″). However, in (6), placement on 'movie' instead of 'amazing' sounds odd.

(6) a. A: *Showgirls* is a nice movie.
b. B: It's an ↓ama-azing movie. $\rightsquigarrow$ *terrible*
#b.′ B: It's an amazing ↓mo-ovie.

We will now model this as follows. We consider the word following the fall to be the prosodic *focus* of the ironic utterance and amend an existing theory for focus to include an implicit negation.

3 Irony and Focus

Geurts and van der Sandt (2004) provide us with a *minimal* theory of prosodic focus. Their *background–presupposition rule* (BPR) states that whenever prosodic focus gives rise to a background $\varphi(x)$, there is a presupposition that $\exists x.\varphi$. In a dynamic model for discourse update, we can specify this as follows (the intended notion of presupposition is van der Sandt's *presupposition as anaphora* model (van der Sandt, 1992)).

Background–Presupposition Rule.
The focus placement separates an utterance into a

foreground f and a *background* $\varphi(x)$. The variable x occurs freely in the formula φ, and the constituent f is of the type required by x. Write $\langle f, \varphi(x)\rangle$ for a foreground–background pair.

Updating a discourse with $\langle f, \varphi(x)\rangle$ is to update with the *proffered* content $(\lambda x.\varphi)(f)$ and the *presupposed* content $\exists x.\varphi$.

This rule is *minimal* in the sense that other models for focus make *at least* the predictions of the BPR. Two prominent models are the QUD model (Roberts, 2012) and Alternative Semantics (Rooth, 1992); the former stipulates that $\langle f, \varphi(x)\rangle$ presupposes the wh-question $?\lambda x.\varphi$ and the latter that $\langle f, \varphi(x)\rangle$ raises the set of alternatives $\{x \mid \varphi(x)\}$. Under the reasonable assumptions that wh-questions presuppose that there is a true answer, and that sets of alternatives are not empty, both models include the BPR.[2] Thus, for instance, the discourse in (7) is treated as follows.

(7) a. A Who does Rachel like?
presupposes: *Rachel likes someone*.
b. B: Rachel likes Michael.
presupposes: *Rachel likes someone*.

In this case, the presupposition of (7b) is bound to the presupposition of A's question (7a). In contrast, there are cases where the presupposition effected by the BPR is accommodated.

(8) a. A Does Rachel like anyone?
b. B: Rachel likes Michael.
presupposes: *Rachel likes someone*.

In (8) the presupposition of (8b) cannot be bound and must be accommodated; the contribution of B's utterance can be paraphrased as *Rachel does like someone—specifically, she likes Michael*.

However, while the BPR models (prosodic) *focus*, it is not sensitive to the overall *tune* of an utterance. The tune can potentially affect the content of both the foreground and the background; see (Beaver and Clark, 2009, p. 47) for a discussion in the context of Alternative Semantics.

Thus, it is not surprising that we need to make amendments to the BPR when attempting to model ironic intonation. Schlöder and Lascarides (ms) argue that fall-rise tunes work by placing an implicit negation in the background. Here, we adapt the BPR to include an implicit negation in the *foreground*. This negation is placed to result in a *contrary* reading.

[2]Some (e.g. Dryer (1996)) have challenged the idea that focus is directly related to presupposition; Geurts and van der Sandt offer responses that we cannot repeat or evaluate here.

Irony Rule.

If an utterance is intonated with the ironic tune, and the fall is immediately preceding the constituent f then the foreground–background pair of the utterance is $\langle \sim f, \varphi(x)\rangle$ where: (i) $\varphi(x)$ is the background resulting from considering f the foreground of the utterance and (ii) $\sim$ is a meta-operator[3] that specifies *contrary negation*:

- if f is a modal or quantifier, $\sim f$ is $f\neg$.
- if f is on a scale, $\sim f$ is an item from the opposite end of the scale;
- if f is a bivalent predicate, then $\sim f$ is $\neg f$;
- if f is an entity, then $\sim f$ is a meta-variable such that for any predicate P, $P(\sim f) = \sim P(f)$.

That is, updating a discourse with this utterance is to update (by usual methods) with the presupposition $\exists x.\varphi$ and the proffer $(\lambda x.\varphi)(\sim f)$.

This rule can be regimented in SDRT (Asher and Lascarides, 2003) with an appropriate semantics for presuppositions (Asher and Lascarides, 1998).

The Irony Rule models some of the examples we have seen so far as follows. Consider first (6b):

(6) a. A: *Showgirls* is a nice movie.
b. B: Yeah, it's an ↓ama-azing movie.
presupp: $\exists x_{predicate}.x(s) \wedge \texttt{movie}(s)$
proffers: $\sim\texttt{amazing}(s) \wedge \texttt{movie}(s)$
$\equiv \texttt{terrible}(s) \wedge \texttt{movie}(s)$

The presupposition of (6b) indicates that B's utterance matches the current topic of the discussion, i.e. the properties of *Showgirls*. Roberts (2012) provides an account of what the current topic is and what it means to match it; this too can be regimented in the present model (Schlöder and Lascarides, ms) and we will not go into the details here. Then, the Irony Rule modifies what B is taken to proffer by adding a *contrary* negation.

By specifying the relative scope of the contrary negation $\sim$ we avoid ambiguities. We show this for (2b′) and (4b″): (In the logical forms we simplify or ignore a number of ancillary details, including tense, possessive case, and the presuppositions associated with proper names.)

(2) a. A: Are you going to Mike's show tonight?
presupp: $\exists s.\texttt{of}(m, s) \wedge \texttt{show}(s)$
proffers: $?\texttt{go}(b, s)$
b.′ B: Yeah, I'll ↓de-efinitely go to that.
presupp: $\exists x_{aux}.x(\texttt{go}(b, s))$
proffers: $\sim\Box\texttt{go}(b, s) \equiv \Box\neg\texttt{go}(b, s)$

[3]That is, it is an operator on logical forms; its application is computed when logical form is constructed.

(4) b.″ B: Yeah, I'll definitely go to ↓tha-at.
presupp: $\exists x_{entity}.\Box\texttt{go}(b, x)$
proffers: $\Box\texttt{go}(b, \sim s) \equiv \Box\neg\texttt{go}(b, s)$

Thus, the proffered contents of (2b′) and (4b″) are the same: both are—by way of irony—negative answers to A's question in (2a). But while in (2b′) the presupposition effected by the Irony Rule is a tautology (because for any p there is a modality ∇ such that ∇p), the presupposition of (4b″) entails that B is going somewhere—just not to Mike's show. This is precisely the difference between the two competing contrary negations in (3).

Overall, the placement of the steep fall in the tune we are considering here appears to be quite flexible, and the Irony Rule predicts this. Similarly to (2) and (4), the utterances in (5) are all assigned the same proffered content, but the presupposition varies in (5a″). The anomalous (6b′) on the other hand is assigned the absurd interpretation that *Showgirls is not a movie*.

(6) #b.′ B: It's an amazing ↓mo-ovie.
presupp: $\exists x_{predicate}.\texttt{amazing}(s) \wedge x(s)$
proffers: $\texttt{amazing}(s) \wedge \sim\texttt{movie}(s)$
$\equiv \texttt{amazing}(s) \wedge \neg\texttt{movie}(s)$

4 Beyond Intonation

As said, we do not claim that every ironic utterance must carry a particular tune. The following is an example by Cutler (1977) for irony that is not marked by prosody.

(9) *Upon entering a restaurant devoid of custom.*
a. A: Looks like a really popular place.

In cases like (9) it is the salient contrast between what is said and what is actually the case that leads us to an ironic interpretation. Cutler does not provide a tune to go with the utterance, but it seems to us that (10a) would be natural, and moreover that one *can* use ironic intonation as in (10b).

(10) a. A: Looks like a ↗really popular place.
b. A: Looks like a ↓rea-ally popular place.

Now note that the Irony Rule makes the correct prediction for (10b). With this in mind, there does not seem to be anything that would stop us from saying that we use the Irony Rule instead of the BPR in *any* situation where an utterance is ironic. That is, we generalise the Irony Rule to also capture utterances like (10a), and take the foreground f to be the focus of the utterance.[4]

[4]There seems to be a tacit consensus that *every* assertion has a focus; McNally (1998) spells this out.

Similar extensions could be made to cases of *written* verbal irony, e.g. as marked by scare quotes. (11) is cited from Predelli (2003).

(11) a. this remarkable piece of 'art' consists of a large canvas covered with mud (...)

Again, applying the Irony Rule to (11a) under $f = \texttt{art}$ makes the correct predictions here. Similar things can be said about written irony marked by some form of irony punctuation.

However, one needs to spell out such extensions of the Irony Rule with great care. Not every false utterance is ironic, and neither is every instance of scare-quoting (Predelli, 2003). There are many potential cues that speakers can employ to signal irony, including intonation, facial expression, gesture, hyperbole *etc.*—and then irony is still frequently misunderstood (Cutler, 1974; Kreuz and Roberts, 1995; Bryant and Fox Tree, 2005). We cannot offer a formalisation of all these cues here.

Then there are still cases of irony where it takes the form of playful mockery instead of the implicit assertion of a negative. Wilson (2006) demonstrates this with (12a,b), said to a very careful driver that always makes sure the tank is filled; the utterances mock this behaviour ironically.[5]

(12) a. A: Do you think we should stop for petrol?
b. A: I really appreciate cautious drivers.

(12a) is no assertion, and it does not seem to be the case that (12b) means that A dislikes (contrary to *appreciate*) cautious drivers. Thus, the cases in (12) go beyond what our Irony Rule captures.

5 Conclusion

We have given a fully formal model for verbal irony, insofar as irony is understood as meaning the opposite of what one has asserted literally. The contribution of the model is in particular to make formally precise what we mean by 'the opposite'. The model starts out as a model of ironic intonation and embeds seamlessly into an extant model of intonation in discourse, but it stands to reason that it may extend to ironic assertions that are not specifically marked by intonation.

[5]A reviewer points out that there is also pragmatic irony: *Thanks for holding the door* after a door has *not* been held. The Irony Rule can potentially explain this; it yields *thanks for not holding the door*. This is a proffer that fulfills preparatory conditions for thanking, so we can continue with standard pragmatic reasoning. Note that one can *explicitly* utter *thanks for not holding the door* to make, by and large, the same speech act as an ironic *thanks for holding the door*.

Acknowledgements

I am grateful to Raquel Fernández, Alex Lascarides, and three anonymous reviewers for their helpful comments.

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no. 607062.

References

Nicholas Asher and Alex Lascarides. 1998. The semantics and pragmatics of presupposition. *Journal of Semantics*, 15(3):239–300.

Nicholas Asher and Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press.

David I Beaver and Brady Z Clark. 2009. *Sense and sensitivity: How focus determines meaning*. Wiley–Blackwell.

Gregory A Bryant and Jean E Fox Tree. 2005. Is there an ironic tone of voice? *Language and speech*, 48(3):257–277.

Gregory A Bryant. 2010. Prosodic contrasts in ironic speech. *Discourse Processes*, 47(7):545–566.

Sasha Calhoun. 2007. *Information structure and the prosodic structure of English: A probabilistic relationship*. Ph.D. thesis, University of Edinburgh.

Anne Cutler. 1974. On saying what you mean without meaning what you say. In *tenth regional meeting of the Chicago Linguistic Society*, pages 117–127.

Anne Cutler. 1977. The context-dependence of intonational meanings. In *13th regional meeting. Chicago Linguistic Society, Chicago*, pages 104–115.

Matthew S Dryer. 1996. Focus, pragmatic presupposition, and activated propositions. *Journal of Pragmatics*, 26(4):475–523.

Bart Geurts and Rob van der Sandt. 2004. Interpreting focus. *Theoretical Linguistics*, 30:1–44.

Rachel Giora. 1995. On irony and negation. *Discourse processes*, 19(2):239–264.

Laurence Horn. 1989. *A Natural History of Negation*. University of Chicago Press.

Roger J Kreuz and Richard M Roberts. 1995. Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and symbol*, 10(1):21–31.

Sachi Kumon-Nakamura, Sam Glucksberg, and Mary Brown. 1995. How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, 124(1):3.

D Robert Ladd. 1980. *The structure of intonational meaning: Evidence from English*. Indiana University Press.

Robert Martin. 1992. Irony and universe of belief. *Lingua*, 87(1-2):77–90.

Louise McNally. 1998. On the linguistic encoding of information packaging instructions. *Syntax and Semantics*, pages 161–184.

Stefano Predelli. 2003. Scare quotes and their relation to other semantic issues. *Linguistics and philosophy*, 26(1):1–28.

Craige Roberts. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Mats Rooth. 1992. A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116.

Julian J Schlöder and Alex Lascarides. 2015. Interpreting english pitch contours in context. In *19th SemDial Workshop on the Semantics and Pragmatics of Dialogue*, pages 131–139.

Julian J Schlöder and Alex Lascarides. ms. Understanding focus: Tune, placement and coherence. Manuscript.

Dan Sperber and Deirdre Wilson. 1981. Irony and the use-mention distinction. *Philosophy*, 3:143–184.

Mark Steedman. 2014. The surface-compositional semantics of english intonation. *Language*, pages 2–57.

Rob A van der Sandt. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9(4):333–377.

Deirdre Wilson. 2006. The pragmatics of verbal irony: Echo or pretence? *Lingua*, 116(10):1722–1743.

Poster Presentations

Dynamic Social Choice for Anaphora Resolution

Sumiyo Nishiguchi
Department of Liberal Arts, Faculty of Science Division I, Tokyo University of Science
1-3 Kagurazaka, Shinjuku, Tokyo 162-8601 Japan
Email: nishiguchi@rs.tus.ac.jp

Abstract

Disambiguation of pronoun reference has been an important issue for both theoretical and computational linguists. While linguistic theories on binding conditions eliminate impossible readings to a certain extent, many inter-sentential anaphora remain ambiguous. Nishiguchi (2011, 2012a,b, 2014, 2016a,b) consider pronoun resolution as a social choice among discourse participants which obeys Arrow's Impossibility Theorem (Arrow 1963). This paper further discusses discourse update of Social Welfare Function which provides updated variable assignment.

In (1), *she* has multiple candidates for its antecedent— Emma, Lisa and Lisa's mom. Proximity and saliency of antecedents have been considered to be key factors to decide (Leass 1991). In (1), the most proximate antecedent *her (Lisa)'s mom* is identified to be the antecedent for *she*.

(1) Frances: ...Not while Emma's not here. You know Emma
Billy: Mm.
Frances: she's, she was walking with Lisa and I weren't there and her Mum sh– jus– , like *she* muc– , *she* mucks about a lot and she told Leigh that if he don't serve her he's gonna die, she's gonna punch him right!

However, proximity does not always resolve referential ambiguity of pronouns. *Him* in (2a) unambiguously means someone other than the closest *John*—some discourse-salient entity. In (2b), the pronoun is ambiguous.

(2) a. John$_i$ likes him$_{\{*i/j\neq i\}}$.

b. John$_i$ said he$_{\{i/j\}}$ likes himself$_{\{i/j\}}$.

In linguistic binding theory (Chomsky 1981, Reinhart 1983), antecedents are called *binders*, which bind bindees that are anaphoric pronouns, e.g., *him* or *himself*. Condition B is that pronouns must be free in their local domain, meaning that they are not bound by the antecedent by means of coindexing and c-commanding relation. *C-command* is roughly equivalent to precedence, with some restrictions.

However, (3) is ambiguous in four ways and can have either one of the following interpretations: i) John broke John's leg, ii) John broke Bill's leg, iii) Bill broke Bill's leg, or iv) Bill broke John's leg. *He* and *his* can be bound by either *John* or *Bill*. The binding theories have no way of disambiguating these pronouns since there is no way of knowing speaker intention. Proximity does not predict the different readings in (3) either.

(3) Anna: Bill$_j$ is a good goalkeeper.
Kim: John$_i$ said he$_{i/j}$ broke his$_{i/j}$ leg recently.

1 Social Choice Theory

Although Social Choice Theory (Arrow 1963, Moulin 1988, Taylor 2005, Gaertner 2009) has only been briefly mentioned in van Rooij (2011) in relation with interadjective comparison, Arrow's Impossibility Theorem is obeyed in a social choice of pronominal reference. Typically, social choice theory explains collective decision making in case of voting and has solved the problems with majority decision. Preferences are ordering between alternatives and should satisfy the following axioms. When R stands for a knowledge of all pairs and x, y and z for alternatives,

Axiom 1. *For all x and y, either xRy or yRx.*

Axiom 2. *For all x, y, and z, xRy and yRz imply xRz.*

Axiom 1 states that the relation R is connected—every candidate is related to each other. Relations that satisfy Axiom 2 are transitive. In (4), N, a finite set of individuals or voters, consists of five individuals and χ, a nonempty set of alternatives or candidates, has three members. Let L(χ) denote the set of all linear orders on χ. A profile *R* is a vector of linear orders, or preferences. R_i is a vector of preferences of an individual i. $N^R_{x>y}$ denotes the set of individuals that prefer the candidate x to y. Supposing R the profile given in this model, $N^R_{o>c}$ is a set of people who prefers Obama to Clinton, that are, Anna, Heather and George (cf. Endriss 2016).

(4) a. N = {a, k, h, g, n}

b. χ = {o, c, m}

c. R $\in$ L(χ)^N

d. $N^R_{o>c}$ = {a, h, g}

e. SWF F: L(χ)^N $\rightarrow$ L(χ)

A *social welfare function* (SWF) *F* is a function which takes individual's preferences and returns collective preference. Arrow demonstrated that any SWF for three or more alternatives the following conditions must be a dictatorship. Condition 2 states that the relative ranking of two candidates remains unchanged regardless of other candidates.

Theorem 1 (General Possibility Theorem (Impossibility Theorem))**.** *If there are at least three alternatives which the members of the society are free to order in any way, then every social welfare function satisfying Conditions 1 and 2 and yielding a social ordering satisfying Axioms 1 and 2 must be either imposed or dictatorial.*

Condition 1 (Pareto condition)**.** *A SWF* F *satisfies the Pareto condition if, whenever all individuals rank x above y, then so does society:* $N^R_{x>y} = N$ *implies xF(R)y*

Condition 2 (Independence of irrelevant alternatives (IIA))**.** *A SWF* F *satisfies IIA if the relative social ranking of two alternatives only depends on their relative individual rankings:* $N^R_{x>y} = N^{R'}_{x>y}$ *implies xF(R)y* $\Leftrightarrow$ *xF(R')y*

Condition 3 (Nondictatorship)**.** *There is no individual i such that for every element in the domain of rule f,* $\forall x, y \in X$: $xP_iy \rightarrow xPy$ *(Sen 1979)*

2 Application to Pronoun Resolution

SWF for pronoun resolution satisfies Arrow's Impossibility Theorem, or General Possibility Theorem, by satisfying Axioms 1, 2, Pareto Condition and IIA but demonstrating dictatorship. Pronoun resolution is compared with voting by multiple voters, discourse participants. The candidates or choices would be different interpretation of the sentence. In (5), the referent of *he* is ambiguous. Chris meant *he* to be *Bob*, while Naomi interpreted *him* to be *John*. As the disagreement on pronominal reference is consolidated in the discourse, pronoun resolution is certainly a social choice and Social Choice Function (SCF) decides the antecedent.

(5) Chris: John said he broke his leg.
Naomi: Did he? John looked fine when I saw him this morning.
Chris: It is Bob who broke his leg.
Naomi: I thought you were talking about John.

When individuals I = {c, n}, candidates χ = {j, b}, Chris and Naomi's ordering is jR_cb $\wedge$ bR_nj, denote the set of linear orders on χ by $L(\chi)$. Preferences (or ballots) are taken to be elements of $L(\chi)$. A profile R $\in L(\chi)^I$ is a vector of preferences. SCF or voting rule is a function F : $L(\chi)^I$ $\rightarrow 2^\chi \setminus \emptyset$ mapping a given profile to a nonempty set of winners; e.g., a singleton set {b} for (5). SWF is a function F : $L(\chi)^I \rightarrow L(\chi)$ mapping any given profile to a (single) collective preference order. Although the preferences between the candidates vary between the individuals, SWF returns a single preference order and ambiguities are resolved during the conversation.

There are three possible antecedents for *she* in (1)—Emma (e), Lisa (l) and Lisa's mother (m). Let us say that Billy (b) prefers e to l, and also l to m to be the antecedent. On the other hand, the speaker Francis (f) prefers m to l, and l to e according to the proximity. All three candidates are ordered in accordance with Axiom 1, i.e., eR_bl $\wedge$ lR_bm and mR_fl $\wedge$ lR_fe. Transitivity also holds for pronoun antecedent preferences. Each of them implies eR_blR_bm and mR_flR_fe. SWF for pronoun resolution also meets Pareto condition. When the interpretation of the addressees agrees with the one of the speaker, the decision of the society follows. It is unlikely that pronouns refer to someone

else other than speaker's intention and hearer's interpretation. A SWF *F* satisfies IIA if the relative social ranking of two alternatives only depends on their relative individual rankings. Let us say that the preference relations are denoted by R and R'. Assume that IIA does not hold and consider a dialogue in (7) where the relative rankings between Bob and John is affected by irrelevant candidate Victor's ranking. The social decision differs from the relative ranking between John and Bob of speaker and hearer, which does not happen, in (8).

(6) Chris: Bob is a good skier. But John said he broke his leg.
Naomi: Did he? Poor Bob!

(7) Chris: Victor is a good skier and so is Bob. But John said he broke his leg.
Naomi: Did he? Poor Bob!

(8) bR'$_c$vR'$_c$j $\wedge$ bR'$_n$jR'$_n$v $\not\rightarrow$ jF(R)b
Then, $N^{R}_{b>j} = N^{R'}_{b>j}$ implies bF(R)j $\Leftrightarrow$ bF(R')j

The speaker's decision on pronominal reference dictates the social preference. Even when there is disagreement or misunderstanding, the speaker corrects unifies interpretation in general, as in (9). Pronoun resolution is dominated, or dictated, by the speaker's meaning.

(9) Chris: Bob is a good skier. But John said he broke his leg.
Naomi: Did he? Poor Bob!
Chris: No. I mean John broke his leg.

(10) xP$_c$y $\rightarrow$ xPy

Proof. Suppose: xP$_c$y $\rightarrow$ $\sim$xPy, that is, xP$_c$y $\rightarrow$ yRx, where R is weak preference. However, the dialogue normally proceeds jP$_c$b $\rightarrow$ jPb as in (10). Contradiction.

□

Lemma 1. *The social welfare function for pronoun resolution is IIA and Pareto but is dictatorial.*

3 Dynamic Update of SCF

In linguistic literature, a variable assignment function g has been assumed to assign the referent to indices indexed to pronouns. For example, g may assign John to the variable x: g(x) = John. Now, g can be considered to be SCF which selects a referent for a pronoun socially. Let us define g and the space as in (11). The assignment function g is updated throughout the discourse as in (12).

(11) a. g = {<x, i>: x refers to i}

b. Information state σ consists of Social Welfare Function F, Social Choice Function g for variable assignment, individual's preferences R, individuals in the discourse X, a set of indices such as i, a set of discourse participants V, and relation between decisions B.

Σ = < F, G, R, X, I, V, B>

(12) σ1 There were ooh's and aah's when he$_{x1}$ finished, and some unbridled laughter. Aileen$_a$ was looking dubiously at her$_{y1}$ husband$_h$ but he$_{x2}$ was in no mood to disapprove.

σ2 He$_{x3}$ winked at the Duke$_d$ and called across to him$_{x4}$, 'What a grand thing, your Honour, to have a wedding without a minister!' The Duke$_d$ did his$_{x5}$ stately bow at that and then Donald$_m$ was calling for another song.

σ3 Some of the veterans$_v$ were on the point of giving tongue but young Donald McCulloch$_m$ was on his$_{x6}$ feet and moving into the middle of the ring, he$_{x7}$ was full of himself$_{x8}$, sparkling with mischief but with an undertow of ardour.

σ4 'Duncan Ban MacIntyre$_b$ wrote a song for his$_{x9}$ wife Mary$_r$.

σ5 I do not know if Alex$_l$ used it to court his$_{10}$ $Mary_r$ – he$_{x11}$ must have used something —'The joke was unconscious but crowing laughter came from the young men$_n$ beside the whisky jar. (BNC A0N1311-1315, *King Cameron*)

(13) a. g1= {<y1, a>, <x2, h>}
I = {a, r} (a: author, r: reader)
S = {a, h}

b. g2 = {<x3, h>, <x4, d>}
S = {a, h, d, m}

c. g3 = { <x6, m>, <x7, m>, <x8, m>}
S = {a, h, d, v, m}

d. g4 = {<x9, b>}
S = {a, h, d, v, m, b, r}

e. g5 = {<x10, l>, <x11, l>}
S ={a, h, d, v, m, b, r, l, n}

f. $[\![her_y]\!]^{g1}$ = a

G is regarded as SCF. Also, the set of best elements S' can be called its choice set of the whole set of alternaties, and is denoted g(S', R) (cf. Sen 1979) R is a sequence of individual's preferences where R_x is a preference ordering of x.

(14) g1(S, R) = {a, h}
g2(S, R) = {h, d}
g3(S, R) = {m}
g4(S, R) = {b}
g5(S, R) = {l}

As the author's dynamic preferences change in the discourse as in (15a), g is updated throughout the discourse by means of a relation B.

(15) a. σ2: hR$_a$d for *he*$_{x3}$ $\wedge$ dR$_a$h for *he*$_{x4}$ $\wedge$ dR$_a$hI$_a$m for *he*$_{x5}$ (aI$_x$b: x is indifferent between a and b, $\wedge$: dynamic conjunction)

b. Social Decision: hRd $\wedge$ dRh $\wedge$ dRhIm

c. B(g_n,g_{n+1})

(16) Dynamic Social Welfare Function:
F_nBF$_{n+1}$BF$_{n+2}$,...

4 Comparison with Other Studies

Dynamic Predicate Logic (Groenendijk and Stokhof 1991) consider update semantics where two states differ with respect to variable assignment. When h[x]g, the state g is updated with respect to the assignment to x. The current paper consider an abstract function B between two SCFs. Parkes and Procaccia (2013) model dynamic decision making under constantly changing preferences using Markov decision processes, in which the states coincide with preference profiles and a policy corresponds to a social choice function.

5 Detection of Speaker Intention

In order to implement Dynamic Social Choice for pronoun disambiguation, speaker's intention needs to be detected from the text. The phrases such as "I mean" are used to resolve ambiguity of pronominal reference in the discourse. as in (17).

(17) '...And Sarah Morgan likes the idea of Angela marrying someone in the government.' McLeish considered this cold and rational assessment. 'When did you last see **her**? **Miss Angela Morgan**, **I mean**.' (BNC AB9)

Out of 18 instances of "I mean PNP" (PNP stands for proper name") found with the query "I mean N" in BNC, 7 instances had a preceding pronoun, the caraphor.

References

Arrow, K. J.: 1963, *Social Choice and Individual Values*, 2 edn, Yale University Press, New Haven.

Chomsky, N.: 1981, *Lectures on Government and Binding*, Foris Publications, Dordrecht.

Endriss, U.: 2016, Judgment aggregation, *in* H. Moulin, F. Brandt, V. Conitzer, U. Endriss, J. Lang and A. D. Procaccia (eds), *Handbook of Computational Social Choice*, Cambridge University Press, pp. 399–426.

Gaertner, W.: 2009, *A Primer in Social Choice Theory. Revised edition*, LSE Perspectives in Economic Analysis, Oxford University Press.

Groenendijk, J. and Stokhof, M.: 1991, Dynamic predicate logic, *Linguistics and Philosophy* **14**, 39–100.

Leass, H. J.: 1991, Anaphora resolution for machine translation: A study, *IWBS Report* **187**.

Moulin, H.: 1988, *Axioms of Cooperative Decision Making*, Econometric Society Monographs, Cambridge University Press, Cambridge.

Nishiguchi, S.: 2011, Computational social choice for pronoun resolution, *Ipsj sig-nl*.

Nishiguchi, S.: 2012a, Social choice for anaphora resolution, *NLP2012 Proceedings*, pp. 97–100.

Nishiguchi, S.: 2012b, Social choice for anaphora resolution, Society for Social Choice, New Delhi.

Nishiguchi, S.: 2014, Application of social choice theory in pronoun resolution, Society for Social Choice, Boston College.

Nishiguchi, S.: 2016a, Social choice for anaphora resolution, *Studies in Liberal Arts and Sciences*, Vol. 48, Tokyo University of Science, pp. 147–156.

Nishiguchi, S.: 2016b, Social choice for disambiguation of pronominal reference, Society for Social Choice, Lund University.

Parkes, D. C. and Procaccia, A. D.: 2013, Dynamic social choice with evolving preferences, *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, Palo Alto, CA, pp. 767–773.

Reinhart, T.: 1983, *Anaphora and Semantic Interpretation*, The University of Chicago Press, Chicago.

Sen, A. K.: 1979, *Collective Choice and Social Welfare*, North-Holland, Amsterdam.

Taylor, A.: 2005, *Social Choice and the Mathematics of Manipulation*, Cambridge University Press, Cambridge.

van Rooij, R.: 2011, Measurement and interadjective comparisons, *Journal of Semantics* **28**, 335–358.

A Process Algebra Account of Speech-gesture Interaction

Hannes Rieser,
Bielefeld University, Germany
Hannes.Rieser@Uni-Bielefeld.de

Abstract

The paper is based on extensive corpus work dealing with the interaction of gesture and speech in natural route-description dialogues. The issue discussed is how non-regimented gesture and speech processes can be modelled in a formal system. The main argument is that this cannot be achieved in structural paradigms currently in use. The proposal is to turn instead to process algebras in the tradition of Milner's π-calculus. The special algebra discussed is a newly developed hybrid λ-ψ calculus which can transport typed λ-expressions over communicating input-output channels. Central for the account is the notion of agent. Speech-gesture interaction is implemented via i-o-channel interactions. Interactions are allowed, postponed or blocked using a typing system. Terminating communication among agents leads to a multi-modal meaning representation.

1 Relevance for the workshop

The key-concepts in the workshop title "Formal approaches to the dynamics of linguistic interaction" are implemented in this paper in the following way: The type of linguistic interaction handled is the interface between speech processes and co-verbal iconic gestures. The dynamics comes in due to the incremental modelling of the speech and gesture processes and the interaction among these which results in multi-modal meaning. (Concerning interaction in the CA sense, as for example treated in Kempson et al. 2016, see Section 4.) Finally, the formal side is provided by the hybrid λ-ψ-calculus used. The paper builds upon former work on a λ-π-account of speech-gesture coordination in Rieser (2014, 2015, 216).

2 Speech-gesture Interaction

The paper starts from the corpus-based observation that gestures are semantically related to the speech they accompany. In light of this, the question arises how they interact with speech and how this interaction can be modelled. Virtually all gesture research assumes that gestures have form and meaning. Following Kendon and McNeill (Kendon 2004, McNeill 1992, 2005), a gesture's structure is characterised by the three consecutive stages preparation, stroke, and retraction. Gestures span from rest position to rest position. Rest positions hence determine a gesture's individuation. The stroke extends over a time span measured in systematic annotation. Only the gesture's stroke must be present to represent a gesture; the meaning of a gesture resides in its stroke. In natural conversation, however, rest positions vary. Moreover, the stroke position is often held by communication participants producing so-called "post-stroke-holds" which can be operative within and across turns, see Example 1. The effect of this is usually that given information is kept and hence visually present on the gesture channel while currently new information is produced on the speech channel: Next speaker may already produce her turn, possibly accompanied by her own gesticulation while the previous speaker still holds stroke information. So gesture and speech have their own modes of encoding and yielding information. Gesture information as understood here is encoded in a formal language specifying topological entities like points, lines, planes or solids and intersections of these. It is drawn from systematic annotation of gesture occurrences using hand positions, palm- wrist- and back-of-hand orientation etc. (as developed in Rieser 2010). Gesture information is always partial. The partiality feature is not treated in this paper, a first account of it is given in (Lawler, Hahn,

and Rieser 2017) contained in these proceedings.

3 Gesture Speech Asynchrony

Systematic annotation of multi-modal data shows that when interacting with speech, gestures do not perfectly synchronise with their "privileged" semantic coordination point. Although received knowledge, this is still a research problem for current descriptive and formal gesture research (Alahverdzhieva and Lascarides 2010, Giorgolo 2010, Lascarides and Stone 2006, 2009, Lücking 2013, Oviatt *et al* 1997, Rieser 2014, Röpke *et al* 2013) as the discussion of gesture-attachment issues shows. Gestures can come entirely before the aligned speech, entirely after it or overlap it. Gesture information can be totally independent of speech information, thus providing additional content as in the examples sketched below. Especially this last case is taken as evidence for the independence of the gesture system from the speech system and will largely determine the style of modelling. As a consequence, the description of speech-gesture coordination cannot be given fitting the gesture meaning representation into the speech meaning representation in some naïve compositional way using e.g. unification. Doing so would violate the independence of gestural information and unduly regiment natural data; especially its non-perfect synchronisation with speech would then escape reconstruction. To clarify this last point, assume that a gesture indicating a square comes entirely before or entirely after an utterance of "window" which does not provide the square-information. Then fusing the square property directly with the "window"-representation would avoid to reconstruct non-perfect synchrony. Motivated by corpus data (Lücking *et al* 2012) and concentrating on referential and iconic gestures, we propose to view gesture and speech as independent processes which interact if it is semantically apt, expressed more technically, if their typings fit. Seen from one point of view, speech is gesture's main companion: gesture may "offer" its information to speech and speech may take it up. If taken up, we get multi-modal information, information assembled from two different sources. If rejected, the gesture stroke can be held waiting for a more appropriate communication opportunity, which, however, could fail to arise: Gesture was put on an outgoing channel but could not enter an ingoing port. There are also more subtle types of gesture-speech communication where speech provides the immediate context for gesture interpretation and the result then again interfaces with speech. It is an open question whether we always have this dependence on the speech context. This will not be discussed in this paper (however, see Lawler *et al*. 2016, where that is the central topic).

4 Outline of Process Algebra Used

Before we give some indication of how to model the gesture speech asynchrony described above, we briefly sum up the empirical findings: Empirical data suggest the need for

- channels on which information (data, agents or procedures) can be sent,
- procedures operating concurrently,
- interfaces enabling communication among processes,
- active and non-active processes, and
- communication among agents organised *via* an i-o-mechanism.

The shift to considering communicating processes necessitates the move to a methodology featuring a process ontology instead of a purely domain-of-objects one as usual in linguistics, logics and philosophy. The one we will use is the ψ-calculus (Bengtson et al. 2011, Johansson, 2010), a recent extension of Milner's π-calculus (Milner, 1999, Parrow, 2001, Sangiorgi and Walker, 2001), belonging to the field of Process Algebra (Fokkink 2000, Hennessy 1988, Bergstra et al. 2000). The ψ-calculus works with processes (so-called agents) and data structures which can be transmitted among agents *via* structured channels using an i-o-facility. Essentially, gesture and speech are viewed as such ψ-agents in this paper.

We provide here and comment upon the central definition for the behaviour of ψ-agents P, Q, … , following (Bengtson et al. 2011):

22 Foll: Hold on. Well, you-CUTOFF. Well, you walk now into this
23 WINDING GESTURE
24 street and then where is the sculpture? Is it at the front or to
GESTURE GESTURE
25 the left or to the right
GESTURE GESTURE
26 WINDING GESTURE HELD

Example 1: English translation of a German transcript from the Bielefeld SaGA corpus (Follower). Right-hand winding gesture in green, left-hand indexing gestures in yellow. The winding gesture (stroke and post-hold) extends throughout turns 22 to 26.

Definition:

0	Nil, the empty agent
$\overline{M}$N.P	Output
$\underline{M}$(λx)N.P	Input
τ	Silent agent
case φ_1: P_1 ‖ … ‖ φ_n: P_n	Case construct
(*νa*)P	Restriction
P | Q	Parallel
!P	Replication
(|Ψ|)	Assertion
"."	Sequential composition

The 0 agent is inactive. "$\overline{M}$N.P" (M overbar, N dot P) puts a data structure N onto an outgoing channel M, and continues with process P, possibly a 0 process. "$\underline{M}$(λx)N.P" (M under-bar) indicates that a data structure is received on the input channel M and substituted for the λ-variable *x* in N and P. In the case construct one alternative P_i is chosen given that φ_i is true. The case construct is also used to model the non-deterministic *or*. The restriction *ν* means that the scope of "a" is local to "P". The parallel operator "|" enables P and Q to expand independently or to communicate with each other via the i-o-operators, possibly after several independent expansions. Replication is defined as P|!P which means that P can be repeated arbitrarily often.

Before we present an informal description of how the λ-ψ- calculus can be put to operation, example 1 shows the English translation of a German transcript from the Bielefeld Speech-and-Gesture-Alignment corpus (SAGA, Lücking *et al*, 2012) used for this purpose.

The example is a section of a multi-modal dialogue between a route-giver and a follower. We briefly sketch how the dialogue excerpt can be modelled using the ψ-technology: The follower uses a winding gesture when starting her contribution with "well". On one reading, she wants to modify "street", so the gesture stroke precedes the optimal interface point. Other possible integration points not discussed here are "walk", "into", and most notably, the event of walking-into itself. After, e.g., interaction with "street" and production of a multi-modal meaning "bendy street" the winding gesture is still held. In the end, ψ's i-i-i-o-facility is taken to model speech-gesture coordination. Due to the incremental grammar hypothesized, the logic of the data structures involved (typed λ-calculus) and the logic of ψ we arrive at a complex hybrid tool, the λ-ψ-calculus.

5 Definition of the Speech-gesture Interaction Agent SGIA in the λ-ψ-Calculus

The λ-ψ-agent SGIA that

- handles incrementality,
- implements the intuitively correct scopes, and
- achieves the speech-gesture integration

is defined in the following protocol (0-agents being sometimes omitted):

SGIA $=_{def}$ $!\overline{ch1}$<λf λu (f(u) ∧ bendy'(u))>
$|\overline{ch7}$ (we'). <λp(well'(p))(we')>
$|\overline{ch4}$ (w'). $\underline{ch5}$ (i'). $\underline{ch3}$ (ts′). $\underline{ch6}$ (nw'). $\overline{ch7}$. nw'<<< λf λru (λx (f(x, you') ∧ r(x, u))e)w'>i'>ts'>
| $\overline{ch4}$. < walk'>.0
| $\overline{ch6}$. <λp now'(p).0>
| $\overline{ch5}$. into'.0
| $\underline{ch2}$(s').$\overline{ch3}$. <<λg (this x (g(x)) s'>>
| $\underline{ch1}$(b'). $\overline{ch2}$. <b' <λx(street'(x))>>.0

The agent consists of eight concurrent processes, indicated by "|" of which only the gesture-simulating one is recursive due to !. Sequentiality (order among constituents) is achieved by types, not given here. It is helpful to keep in mind that we have o-i-channels indicated by overbar and under-bar, respectively: A winding gesture is produced concurrently with the words <"well", "you", "walk", "now", "into", "this", "street">. Using $\overline{ch1}$ it sends its information to "street", yielding thus "bendy street". The property "bendy street" in turn sends its information via $\overline{ch2}$ to "this" and we get the referring expression "this bendy street". This information is set aside for a while, since the output channel does not immediately find a matching input channel. The information tied to "you" is a propositional function and needs several constants inserted *via* channels $\underline{ch4}$ (w'), $\underline{ch5}$ (i') and $\underline{ch3}$ (ts′), respectively, in particular a relation "walk" defined on an event e and a subject "you" and a relation "into" defined on the same event and the multi-modal referring expression "this bendy street" already compiled. The resulting term is the proposition "There is an event of you walking into this street" that "now" looks for due to its $\overline{ch6}$ and with which it combines moving into $\underline{ch6}$ to yield another proposition, "Now there is an event of you walking into this bendy street", in more colloquial terms (cf. the annotation of the dialogue-part in Example 1), "Now you walk into this bendy street". This new proposition is put on an outgoing channel $\overline{ch7}$ and combines with "well" using input channel $\underline{ch7}$, again generating a proposition "Well, now you walk into this bendy street" while the winding gesture continues to be held due to $!\overline{ch1}\langle\lambda f\,\lambda u\,(f(u) \wedge \text{bendy}'(u))\rangle$. Hence, the formula to be interpreted is in the end

$!\overline{ch1}\ \langle\lambda f\lambda u(f(u) \wedge \text{bendy}'(u))\rangle.0 \mid \text{well}'(\text{now}'(\text{walk}'(e, \text{you}') \wedge \text{into}'(x, \text{this}' \text{ x } (\text{street}'(x) \wedge \text{bendy}'(x))))).0$

of which only the second closed process $\text{well}'(\text{now}'(\text{walk}'(e, \text{you}') \wedge \text{into}'(x, \text{this}' \text{ x } (\text{street}'(x) \wedge \text{bendy}'(x))))).0$ is satisfied.

6 Future Research

The account given handles the property + noun semantics case using λ-ψ-processes. The shortcoming of this particular example is that the initial introduction of the bendy-street-gesture combination into the dialogue is not shown. This opens up the question at which level existing dialogue theories can be married with the process architecture. In talks I already sketched that the basic λ-ψ-i-o-facility can also be used to model split utterances with in-turn-acknowledgements as discussed, e.g., in Eshghi et al. (2015):

A. The doctor.
B. Chorlton?
A. No, Fitzgerald.
B. uh-huh.

In order to do so, one establishes "turn channels" transporting the respective dialogue contributions of A and B. These have to satisfy A's and B's tests modelled with the case-construct. Furthermore, by way of generalisation it can be argued that ψ can be used to model any type of multi-modal information which was subjected to rigid annotation.

Acknowledgements

Thanks for the comments of three ESSLLI-reviewers which helped to improve the Ms. Due to space restrictions not all of the reviewers' suggestions could be taken up. Some would also require much additional research.

References

Alahverdzhieva, K. and Lascarides, A. (2010). Analysing language and co-verbal gesture in constraint-based grammars. In Müller, St., editor, *Proceedings of the 17th International Conference on Head-Driven Phase Structure Grammar (HPSG)*, pp. 5–25, Paris.

Bengtson, J., Johansson, M., Parrow, J., and Björn, V. (2011). Psi-Calculi: A framework for mobile processes with nominal data and logic. *Logical Methods in Computer Science.* Vol. 7 (1:11), 2011, pp. 1-44.

Bergstra, J. A., Fokkink, W. J., Ponse, A. (2000). Process algebra with recursive operations. In Bergstra et al. (editors), *Handbook of Process Algebra*. Amsterdam: Elsevier, pp. 333-391.

Eshghi, A., Howes, C., Gregoromichelaki, E., Hough, J., Purver, M. (2015). Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics*, pp. 261-271.

Fokkink, W. (2000). *Introduction to Process Algebra*. Berlin, Heidelberg: Springer.

Giorgolo, G. (2010). *Space and Time in Our Hands.* UIL-OTS, Universiteit Utrecht, 2010.

Hennessy, M. (1988). *Algebraic Theory of Processes.* Cambridge, Mass.: The MIT Press.

Johansson, M. (2010). *Psi-calculi: a framework for mobile process calculi*. Diss. from the Faculty of Science and Technology 94. 184 pp. Upsala.

Kempson, R., Gregoromichelaki, E., Cann, R., Chatzikyriakidis, S., (2016). Language as mechanisms for interaction. *Theoretical Linguistics*, Bd. 42, Heft 3-4.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: CUP.

Lascarides, A. and Stone, M. (2006). Formal semantics of iconic gesture. In Schlangen, D. and Fernández R., editors, *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue, (Brandial)*, pp. 64–71, Potsdam.

Lascarides, A. and Stone, M. (2009). A formal semantic analysis of gesture. *Journal of Semantics*, 26(4), pp. 393-449.

Lawler, I., Hahn, F., and Rieser, H. (2016). Multi-modal context-dependency. Extended abstract. *Workshop on Situations, Information, and Semantic Content*, LMU Munich.

Lawler, I., Hahn, F., and Rieser, H. (2017). Gesture meaning needs speech meaning to denote - A case of speech-gesture meaning interaction. In *Proceedings of the Workshop on Formal Approaches to Linguistic Interaction; ESSLLI 2017*.

Lücking, A. (2013). *Ikonische Gesten. Grundzüge einer linguistischen Theorie.* De Gruyter Mouton, Germany.

Lücking, A., Bergmann, K., Hahn, F., Kopp, S., and Rieser, H. (2012). Data-based analysis of speech and gesture: The Bielefeld speech and gesture alignment corpus (SaGA) and its Applications. *Journal on Multimodal User Interfaces* 7(1-2), pp. 5-18.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought.* Chicago: Chicago University Press.

McNeill, D. (2005). *Gesture and Thought.* Chicago: Chicago University Press.

Milner, R. (1999). *Communicating and mobile systems: the π-calculus.* Cambridge: CUP.

Oviatt, S., DeAngeli, A., and Kuhn, K. (1997). Integration and synchronisation of input modes during multimodal human-computer interaction. *CHI*, pp. 415-422.

Parrow, J. (2001). An introduction to the π-calculus. In Bergstra, J. A., Ponse, A., and Smolka, S. A., editors, *Handbook of Process Algebra*. Amsterdam: Elsevier, pp. 479–545.

Rieser H. (2010). On factoring out a gesture typology from the Bielefeld speech-and-gesture-alignment corpus (SAGA). In: Kopp S., Wachsmuth I. (eds) *Proceedings of GW 2009*, pp 47–60.

Rieser, H. (2014). Gesture and speech as autonomous communicating processes. Talk at *Workshop on Embodied meaning goes public,* Stuttgart University.

Rieser, H. (2015). When hands talk to mouth. Gesture and speech as autonomous communicating processes. *Proceedings of the 19th Workshop on the Semantics and Pragmatics of Dialogue, (goDIAL),* Gothenburg.

Rieser, H. (2016). A process-algebra account of speech-gesture interaction. Extended abstract, *Workshop on Situations, Information, and Semantic Content*, LMU Munich.

Röpke, I., Hahn, F., and Rieser, H. (2013). Interface constructions for gestures accompanying verb phrases. In *Abstracts of the 35th Annual Conference of the German Linguistic Society*, Potsdam, pp. 295–296.

Sangiorgi, D. and Walker, D. (2001). *The π-calculus. A Theory of Mobile Processes.* Cambridge: CUP.

Rational Interaction and the Pragmatics of the Slippery Slope and 'Guilt by Association'

Gerhard Schaden
Université Lille SHS
CNRS UMR 8163 STL
59000 Lille Cedex
gerhard.schaden@univ-lille3.fr

Abstract

This paper proposes a pragmatic analysis of two so-called fallacies in argumentation, namely the 'Slippery Slope' and 'Guilt by Association'. I will examine their rational use, and argue that they exemplify at least partially non-cooperative, but still inference-based conversational moves.

1 Introduction: Discourse Participants and Cooperation

Pragmatic theories of the (neo|post)Gricean type typically assume that conversation and inferences can be modeled by the (rational) interaction of a speaker and a hearer, and also, that pragmatic inferences are based somehow on a cooperative interaction between speaker and hearer.

Both of these assumption have been challenged. So-called 'argumentative' theories of pragmatics (Ducrot, 1980; Merin, 1999) provide a different way of rationalizing pragmatic inferences, based on the rational pursuit of opposing goals. And other fields of the study of argumentation use a considerably richer notion of 'discourse participants' than the standard speaker vs. hearer dichotomy (Groarke and Tindale, 2004; Tindale, 2007; Tindale, 2015): there is an ARGUER, advancing some argument, which goes against the OPPONENT, and is directed to convince an AUDIENCE.[1,2] In this paper, I will show that such a richer representation is required, and that without it, deciding issues like the question of the degree of cooperation involved in a verbal exchange cannot be properly addressed.

Gricean pragmatics operates within the assumption that speaker and hearer are cooperative, as embodied by the cooperative principle. While rarely stated explicitly (but see Fox (2014) on this issue), one implication one can draw from this idea is that in non-cooperative contexts, there should not be any pragmatic inference. The resulting 'Kumbaya'-pragmatics may not correspond to Grice's intentions, but it makes certain empirical predictions. These predictions, however, rely on the fact that the participants the speaker is cooperative with are correctly identified.

Argumentative communication provides a testing ground for the idea that pragmatic inferences necessarily necessitate cooperative interaction, or necessarily involve a common ground (see Clark, 1996). In order for meaningful argumentation to be possible, the issue argued about cannot be part of the common ground (this would be 'preaching to the choir'). In many instances, argumentation will include opposed preferences of the participants. In extreme cases, like debates between candidates for a presidential election, argumentation boils down to a zero-sum game (as assumed explicitly for all kinds of linguistic interaction by Merin, 1999): whatever benefits one of the candidates will hurt the other to the same extent. Yet, the audience of the arguers is not their respective opponent, but rather the electorate (or the part of it that is watching the debate), and the arguer clearly would want to be cooperative with respect to (at least parts of) that audience. If we could show that argumentation in such contexts is dependent on pragmatic inference, and does not involve a fully cooperative setup even with the audience, this would be a strong argument against the Kumbaya-vision of pragmatics.

Now, the idea that argumentative communication is based on inference is not exactly new; it can be traced back at least to Aristotle. He noted that 'normal' argumentation does not contain com-

[1] And while it is not often developed, the audience in itself can overlap to various degrees with the DECIDERS.

[2] Levinson (1988) or Clark (1996) have proposed to enrich the speaker-hearer dichotomy in other ways. Their move is motivated by issues such as turn-taking and participation. I — like the literature on argumentation — focus however on strategic interaction. As far as I see, these proposals are perfectly orthogonal.

prehensive and logically valid demonstrations, but rather abbreviated versions of it. Aristotle points out — in an explanation that has a Gricean ring to it — that generally, speakers in argumentation do not use full syllogisms, but rather the shorter *enthymemes*:[3]

> The Enthymeme must consist of few propositions, fewer often than those which make up the normal syllogism. For if any of these propositions is a familiar fact, there is no need even to mention it; *the hearer adds it himself.* [...] we must not carry its reasoning too far back, or the length of our argument will cause obscurity: nor must we put in all the steps that lead to our conclusion, or we shall waste words in saying what is manifest. [4]

Avoid obscurity and *be brief* can be included in a cooperative setup (as in Grice's maxims of manner), or they can be seen as self-interested, rational ways of dealing with the (limited) attention span of the audience. But crucially, these maneuvers entail making use of inference by the audience. While Aristotle seems to consider here essentially inferences from the common ground, this is not always the case. Consider "*Make America great again. Vote Trump.*" Assuming this to be an instance of argumentative communication, this relies (at least) on the unstated proposition "*Donald Trump can make America great again*". This proposition is arguably not part of the common ground of American voters.

My intention here is to follow up on remarks by Volokh (2003), namely that mechanisms that are generally considered by logicians to be fallacies (for instance, *ad hominem* arguments) are better conceived of as *heuristics* for real-time decision-making by rationnally ignorant agents.[5]

I will also assume that there are principles of rational argumentation, such as giving the strongest argument at the arguer's disposal (see, e.g. Anscombre & Ducrot, 1983), and that the audience of an argument interprets not only what has been said, but also what could have been said. I will try to show that these principles lead to inferences that can be explained by the interaction of rational agents, and therefore, that they remain in the realm of pragmatics.

More precisely, I will show with the examples of the Slippery Slope and 'Guilt by Association' that these are contexts where the notions of *audience* and *opponent* need to be kept apart, but where it is doubtful that we face a fully cooperative setting with either of these.

2 The Verbal Mechanisms of the Slippery Slope

The argument of the slippery slope involves advancing an argument against some proposition *A* based not on the intrinsic merits or deficiences of *A*, but rather on the assumption that once *A* is in place, there would be no way of meaningfully opposing *B*, which is assumed by the arguer to be undesirable. Arguments of this kind often appear in discussions concerning gay marriage (*A*), which is opposed not as such, but which is argued to lead to a state where one could not oppose further develo legalization of adoption by gay couples, or even the legalization of polygamy, incest, bestiality, etc. This argument is fallacious (that is, in need of additional inference), since there is no logical, entailment-based link between, e.g., polygamy and gay marriage.[6]

Volokh (2003) provides a very complete study of the mechanism of the slippery slope, and makes clear that at least in some cases, the risk of going down a slippery slope is real. While his paper contains many observations that are relevant to linguists, his main problem — as a legal scholar — is to identify how the slippery slope works *in the real world*, which are the mechanisms of slippage, and how it can be avoided. His article, however, does not address directly the issue as to when the *argument* of the slippery slope is rational or appropriate, which is the focus of the present paper.

Volokh identifies several mechanisms that can cause slippery slopes, of which several entail mixed motives (what Volokh calls "multi-peaked" preferences). An example Volokh gives is the proposal to install video surveillance in a town, where there are in principle three alternatives: i) oppose it, and remain in current state (note this 0); ii) vote for a version where cameras are not connected to facial-recognition software and tapes are rapidly destroyed (*A*); or iii) vote for a version where cameras are connected to facial-recognition software and tapes are kept for a long time (*B*). In a context where voters are not only motivated by concerns

[3] A enthymeme is often considered to be a truncated syllogism — that is, a syllogism where one premise is lacking.

[4] Aristotle, *Rhetoric, 1357a 16, 1395b 25*, as cited in Hamblin (1970, 71); my emphasis.

[5] Rational ignorance means that we often have to make decisions while ignoring their precise outcome.

[6] I take it that even the staunchest opponent of gay marriage would have to concede that there might be, *in principle*, a society which bans polygamy, but nevertheless allows gay marriage. I furthermore take it that the disagreement hinges on the question whether such a state of affairs is attainable and maintainable for the real world, given the current state.

about privacy vs. security, but also by the financial cost of the system, some people will oppose video surveillance for cost-reasons, even if they are in principle favorable. Therefore, there may be no way of directly going to B from 0. However, if A is enacted, the cost motive for opposing B will be removed (since tapes and software are much less costly than the installation of the cameras in the first place), and in a subsequent vote, B could be adopted. Therefore, people with a preference profile of $A > 0 > B$ should rationnally oppose the move to A, even though it is their preferred option, because they would end up with B, which they strongly oppose.

Let us come back to the issue of cooperativity. Volokh (2003, 1034f.) makes the following observation with respect to slippery slopes.

> Slippery slopes may occur even when a principled distinction can be drawn between decisions A and B. The question shouldn't be "*Can* we *draw the line between A and B?*", but rather "*Is it likely that* other *citizens, judges, and legislators will draw the line there*?" [...] Societies are composed of people who have different views, so one person or group of people may want to oppose A for fear of what *others* will do if A is accepted. And these others need not constitute a majority of society: slippery slopes can happen even if A will lead only a significant minority of voters to support B, if that minority is the swing vote.

According to Volokh, thus, a slippery slope will not occur if one can trust the deciders (the *other citizens, judges, and legislators*). Therefore, an argument of the slippery slope is a sign of lack of trust, and not of principled and uncompromising cooperation. In order to investigate when it is rational to use an argument of the slippery slope, I will try to make explicit the decision process in terms of conditional probabilities, along what has been done by Merin (1999).

First of all, the argument of the slippery slope is an indirect argument. Going down this route should only be done if the direct approach — that is, directly opposing A — does not appear to be viable. This in itself is a sign of a weak position, and it is rational only if — given the arguer's information state, there is a sufficient majority of deciders in favor of A, such that the change towards A can be enacted. Furthermore, in order for the argument of the slippery slope to make sense, there are two further requirements: first, it must be the case that the probability of implementing B given A is higher then the probability of implementing B. This can be written as follows: $P(B|i[A])^7 > P(B|i)$. This is probably too weak a requirement, since it must be also the case that the the deciders are in majority opposed to B.

Second, the slippage towards B will only work as an argument if B is considered as sufficiently repulsive to motivate a rejection of A even if if A is the preferred option. Let us note the expected utility of some action or state S $EU(S)$. One way of thinking about this is the following: The expected utility for state A, given the probability of slippage from A to B $P(B|A)$, and the expected utility of B, will be the expected utility of A plus the probability of slippage multiplied by the expected utility of B.

$$(1) \qquad EU(A) + P(B|A) \times EU(B)$$

If (1) is negative, a rational agent should reject a move to A. When will this be the case? The lower the (positive) expected utility of A, and the higher the risk of slippage and the (negative) expected utility of B, the stronger the trend to rejection.

Therefore, the slippery slope will be most appropriate if A is too popular to be attacked directly, if B is as repulsive as possible, and if, at the same time the risk of slippage from A to B is considered to be high among the audience. It seems obvious that in most circumstances, these conditions will not be met — especially if the passage from A to B is under full control of the audience, and is not impinged on by issues of applicability (possibly under the control of third parties — like the justice or the police). Hence, if the audience are the deciders, the mere suggestion of the possibility of a slippage can be interpreted as a vote of non-confidence towards (at least) a majority of the deciders. Therefore, the argument of the slippery slope can be detrimental to the arguer and his thesis.

Notice, though, that the argument of the slippery slope can only work if there is a justifiable lack of trust with respect to the deciders, and that it is therefore not an argument that is built on unconditional cooperation.

3 Guilt By Association

Guilt by association is an argumentative move where the opponent's position is rejected based on the assertion that this position was also held by other, less-than-recommendable people (noted henceforth as *bogeyman*). For instance, *reductio ad Hitlerum* is an instance of guilt by association, but it englobes also *red-baiting* on the other end of the

[7] I note as $i[A]$ the information state i augmented with A — which may cause changes other than merely adding A.

political spectrum. Once again, this move is classified as a fallacy, since the fact (or still less, the assertion) that, say, Hitler held some view (against smoking, or for vegetarianism, for instance) cannot generally be taken as a reason for dismissing this view without additional arguments (or contextual inference, for that matter). More precisely, guilt by association depends on a relevance-implicature.

Now, when is such an argumentative move rational? Notice that guilt by association may have two different aims: first, if the opponent is (part of) the audience, the opponent is to be shamed into accepting the arguer's view. Second, if the opponent is not part of the audience, it aims to exclude the opponent's arguments from consideration by the audience.

Let us start by considering the first strategy. The wished-for reaction in the opponent would be the following: *The arguer asserts that only* bogeyman *would hold opinion ϕ.*[8] *I asserted that position, but I do not wish to be identified as* bogeyman*; therefore, I (possibly publicly) abandon my position, and adopt the position of my opponent.*

In the second scenario, where the opponent is not a member of the audience, the basic process is like above — but since members of the audience have not brought forward any claim, they will not have to publicly retract.

Generally, the move is based on social exclusion, and is intended to remove the opponent from the people that are entitled to present counterarguments with respect to some theme to the audience. Therefore, the audience should assume that the proportion of people holding the opinion is low (whether this is true or not is another question) — since otherwise, it will not provide a good means of social stigmatization. If the stigmatized opinion is widespread, and even if the *bogeyman* as such is strongly rejected, guilt by association may backfire, and provoke rejection towards the arguer.

The latter thought process can be explicited as follows: *I hold opinion ϕ, and I know that I am not* bogeyman. *Furthermore, I know that a considerable part of the audience hold opinion ϕ, and are no* bogeymen. *Therefore, the argument of the arguer does not hold. The arguer could have presented another type of argument, but he chose this one. Therefore, this must be what he thinks to be his strongest argument at that point. Since it is not correct, the case must be dismissed.*

Guilt by association operates with a strongly negative social emotion (which is the prototypical non-cooperative move), and pits that against whatever evidence the opponent has for his position. If the opponent thinks that that evidence holds up well, or feels strongly against being publicly shamed, guilt by association will fail.

Finally, there is an intrinsic problem with guilt by association arguments: the more evil *bogeyman* in the opinion of the audience,[9] the stronger the argument. And as the *bogeyman* is evil, this implies that the opponent's argument should not even be acknowledged. However, if the audience identifies this as the intention behind the use of the guilt by association argument — and if (at least a considerable part of) the audience holds that opinion — it will be received as a refusal to discuss that particular issue, and communication may break down. In any case, guilt by association is a highly polarizing type of argument, whose aim is rather to mobilize the own camp within the audience than to bring around people holding the opponent's view. So, if the audience and the opponent are identical (or if the opponent is a critical part of the audience), guilt by association should in most cases be avoided.

4 Conclusion

I have tried to show that in argumentative discourse, the familiar speaker-hearer dichotomy is too simple to meaningfully describe the rational interaction of discourse participants. I also tried to show that there are inference-based discursive moves which are clearly non-cooperative, not only with respect to the opponent, but also with respect to parts of the audience. In case of an argument of the slippery slope, the basic outline of the argument is based on the idea that the deciders cannot be sufficiently trusted not to go down the slippery slope. In case of guilt by association, I have argued that the argument is based on (the threat of) social ostracism, and therefore equally non-cooperative.

[8] This may appear as a unnecessary strengthening of an argument of guilt by association, which is likely to go rather like "*bogeyman thought|said that, too*". However, even the scariest bogeyman will have countless opinions that are perfectly mainstream in the considered community, such as "*It is right to drive on the right hand side*", etc. Therefore, for the argument even to be relevant, it has to be the case that that particular kind of opinion must have some link to what makes that bogeyman a bogeyman.

[9] In an assembly of neonazis, the *reductio ad Hitlerum* would obviously not qualify as an argument of guilt by association, but rather as an argument by authority.

Acknowledgments

I would like to thank the three anonymous reviewers for their comments and suggestions on a previous version of this paper. All remaining errors are mine alone.

References

Jean-Claude Anscombre and Oswald Ducrot. 1983. *L'argumentation dans la langue*. Philosophie et langage. Pierre Mardaga Éditeur, Liège.

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

Oswald Ducrot. 1980. *Les échelles argumentatives*. Minuit, Paris.

Danny Fox. 2014. Cancelling the maxim of quantity: Another challenge for a Gricean theory of scalar implicatures. *Semantics and Pragmatics*, 7(5):1–20, April.

Herbert Paul Grice. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics. Speech Acts*, volume 3, pages 41–58, New York. Academic Press.

Leo A. Groarke and Christopher W. Tindale. 2004. *Good Reasoning Matters!: A Constructive Approach to Critical Thinking*. Oxford University Press, Oxford, 3 edition.

Charles Hamblin. 1970. *Fallacies*. Methuen & Co, London.

Stephen C. Levinson. 1988. Putting linguistics on a proper footing: Explorations in Goffman's participation framework. In Paul Drew and Tony Wootton, editors, *Erving Goffman: Exploring the Interaction Order*, chapter 7, pages 161–227. Polity Press, Oxford.

Arthur Merin. 1999. Information, relevance, and social decisionmaking: Some principles and results of Decision-Theoretic Semantics. In Lawrence S. Moss, Jonathan Ginzburg, and Maarten de Rijke, editors, *Logic, Language, and Computation*, volume 2, pages 179–221. CSLI Publications, Stanford.

Christopher W. Tindale. 2007. *Fallacies and Argument Appraisal*. Cambridge University Press, Cambridge.

Christopher W. Tindale. 2015. *The Philosophy of Argument and Audience Reception*. Cambridge University Press, Cambridge.

Eugene Volokh. 2003. The mechanisms of the slippery slope. *Harvard Law Review*, 116(4):1026–1137.