PROCEEDINGS OF THE 15^{th} STUDENT SESSION

OF THE EUROPEAN SUMMER SCHOOL FOR LOGIC, LANGUAGE AND INFORMATION

Marija Slavkovik (editor)

AUGUST 9-20, COPENHAGEN, DENMARK

Program committee:

Marija Slavkovik (chair)

Logic and Computation

Jens Ulrik Hansen

Szymon Klarman

Logic and Language

Ekaterina Lebedeva

Mingya Liu

Logic and Computation

Natalia Vinogradova

Pierre Lison

Area Experts:

Logic and Computation

Andreas Herzig

Stephane Demri

Logic and Language

Benjamin Spector

Chris Barker

Language and Computation

Alexander Clark

Antal van den Bosch

Preface

This year we celebrate 15 years of Student Session at the European Summer School of Logic, Language and Information (ESSLLI). We received forty-nine submissions, both in the long and short paper tracks, from which we selected the papers included here. We owe a great gratitude to the many reviewers for taking the time to provide insightful reviews that helped us in the selection process. The same gratitude is owed to all the students who submitted their work for the Student Session. It is these exquisite contributions that made the Student Session such an interesting event.

I would like to thank the program committee for the excellent teamwork and dedication, as well as the area experts for their contribution. The organizing committee of ESSLLI 2010 has to be commended for meeting all our logistic needs in Copenhagen. Other individuals, from FoLLI and ESSLLI Organizing Committee, provided many helpful insights and contributions, in particular Paul Dekker and Sophia Katrenko. As in previous years, Springer-Verlag has generously offered prizes for Best Paper awards, and for this we are very grateful.

Marija Slavkovik,
Chair of the 2010 Student Session

August, 2010

Contents

Language and Computation - Long Papers 161

Language and Computation - Posters 223

Reviewers 247

1

Logic and Computation

Long Papers

Epistemic Logic and Relevant Alternatives

Wesley H. Holliday

Department of Philosophy, Stanford University

Abstract. According to the Relevant Alternatives (RA) Theory of knowledge, to know that something is the case is to have *ruled out the relevant alternatives*. The conception of knowledge in epistemic logic also involves the elimination of possibilities, but without an explicit distinction, among the possibilities consistent with an agent's information, between those *relevant* possibilities that an agent must rule out in order to have knowledge and those remote, far-fetched or otherwise irrelevant possibilities. This paper proposes a formalization of the RA theory. Using a variation of modal preference logic, we formalize two influential versions of the theory, due to Fred Dretske and David Lewis respectively. The formalization clarifies a famous debate in epistemology, pitting Dretske against Lewis, about whether the RA theorist should accept the principle that *knowledge is closed under known implication*, familiar as the K axiom in epistemic logic. Dretske's thesis that knowledge is a "semi-penetrating" operator is shown to be in tension with his version of the RA theory, while Lewis's version is naturally understood in terms of *dynamic* operations that change what is relevant through shifts in context.

1 Introduction

Imagine that two medical students are subjected to a test. Their professor introduces them to the same patient, who presents various symptoms, and the students are to make a diagnosis of the patient's condition. After some independent investigation, both students conclude that the patient has a common condition C. In fact, they are correct. Yet only the first student passes the test. The professor wished to see if the students would check for another common condition C', which causes the same visible symptoms as C. While the first student ran laboratory tests to rule out C' before making the diagnosis of C, the second student made the diagnosis of C after only a physical exam, having never considered the possibility of C'. As a result, the second student fails the test.

In evaluating the students, the professor concludes that although both students gave the correct diagnosis of C, the second student did not *know* that the patient's condition was C, since he did not rule out the alternative of C'. Had the patient's condition been C', the second student might still have made the diagnosis of C, since the physical exam would not have revealed a difference. In a sense, the second student *got lucky*—the condition he associated with the patient's visible symptoms happened to be the condition the patient had, but if the professor had chosen a patient with C', the second student would have made a misdiagnosis. By contrast, the first student secured against this possibility of error by running the laboratory tests. For this reason, according to the professor, the first student knew the patient's condition and passed the test.

Of course, the first student did not secure against *every* possibility of error. Suppose there is an extremely rare disease X such that people with disease X appear to have C on many laboratory tests, even though people with X are completely *immune* to C, and only extensive further testing can detect the presence or absence of X in its early stages. Should we say that the first student did not know that the patient's condition was C after all, since she did not rule out the possibility of X? The requirement that one rule out *all* possibilities of error makes knowledge impossible, since there are always some possibilities of error—however remote and far-fetched—that are not eliminated by one's evidence and experience. Yet if the student has no reason to think that the patient may have the rare disease X, then it should not be necessary to rule out such a remote possibility in order to know that the patient has some common condition.

The previous example and analysis provides the classic kind of argument for the Relevant Alternatives (RA) Theory of knowledge. According to this theory, to know that something is the case is to have *ruled out the relevant alternatives*. The necessary condition, that one must rule out all of the relevant alternatives, is a kind of *anti-luck* condition on knowledge, as suggested in the first part of the example. The sufficient condition, that one need only rule out the *relevant* alternatives, is a kind of *anti-skeptical* position about knowledge, as suggested in the second part of the example. What makes a particular alternative relevant is a controversial issue in epistemology. However, in the case of medical diagnosis, doctors routinely make judgments about what are the relevant alternatives that must be ruled out and what are the remote possibilities that may be properly ignored, unless and until new information makes them relevant.

Like the conception of knowledge according to the RA theory, the conception of knowledge in epistemic logic [11,8,13] also involves the elimination of possibilities. Yet in standard epistemic logic there is no explicit distinction, among the possibilities consistent with an agent's information, between those that are relevant and those that are not. Of course, we may simply think of the set of states accessible from a given state in a standard epistemic model not as the set of all possibilities consistent with the agent's information in the given state, but rather as the set of *relevant* possibilities consistent with the agent's information. This appears to be the view of Hendricks and Symon [10]:

> Contemporary mainstream epistemologists choose to speak of *relevant* possible worlds as a subset of all possible worlds. The epistemic logician considers an *accessibility* relation between scenarios in a designated class out of the entire universe of possible scenarios. There is no principled difference between relevance and accessibility. (p. 144)

However, thinking of all accessible states as relevant will not allow us to model examples like the medical diagnosis case, which requires that we make distinctions between possibilities that are consistent with the agent's information but not relevant, possibilities that are relevant but are not consistent with the agent's information (having been ruled out by tests), etc. To model what is distinctive about the RA theory, we must go beyond standard epistemic logic.

This paper proposes a formalization of the RA theory. Using a variation of modal preference logic [3,2], we formalize two influential versions of the theory, one due to Fred Dretske [6,7], the earliest proponent of the RA theory, and the other to David Lewis [12], who among others gave the RA theory a "contextualist" twist. Not only does the formalization display the modeling capability of the modal logic framework, it also yields a philosophical payoff. In particular, the formalization clarifies a famous debate in epistemology [6,16,12,9,14], pitting Dretske against Lewis, about whether the RA theorist should accept the principle that *knowledge is closed under known implication* (hereafter "closure"), familiar as the K axiom $K(\varphi \to \psi) \to (K\varphi \to K\psi)$ of normal epistemic logics.

2 Philosophical Background

The K axiom has been much discussed in epistemic logic in connection with the "problem of logical omniscience." Together with the the rule of necessitation—if φ is a theorem, then $K\varphi$ is a theorem—the K axiom implies that agents know all the logical consequences of what they know. As Fagin et al. [8] put the problem, "While this property may be reasonable for some applications, it is certainly not reasonable in general. After all, we cannot really hope to build logically omniscient robots" (p. 9). In epistemology, the closure principle expressed by the K axiom has also been much discussed, but for a different reason. According to Dretske [6], closure would not hold in general even for computationally unlimited agents: "Were we all ideally astute logicians, were we all fully appraised of all the necessary consequences...of every proposition, perhaps then epistemic operators would [satisfy closure]. It is this...claim that I mean to reject" (p. 1010).

Dretske identifies a spectrum of *fully penetrating*, *semi-penetrating*, and *non-penetrating* sentential operators. An operator O is fully penetrating if whenever P entails Q, $O(P)$ entails $O(Q)$. Fully penetrating operators include 'it is true that', 'it is a fact that', 'it is necessary that', etc. For if P entails Q, then 'it is necessary that P' entails 'it is necessary that Q', as in normal modal logic. Non-penetrating operators include 'it was strange that', 'it was a mistake that', 'it was accidental that', etc. For it may be strange that P *and* Q but not strange that P. Such an operator fails "to penetrate to some of the most elementary logical consequences of a proposition" (p. 1009). Dretske's question is where epistemic operators such as 'S knows that' fall on this spectrum. His thesis is that they are *semi*-penetrating. That epistemic operators are not non-penetrating is the "trivial side of my thesis...because it seems...fairly obvious that if someone knows that P and Q...he thereby knows that Q" and "If he knows that P is the case, he knows that P or Q is the case" (ibid.). However, neither are epistemic operators fully penetrating, for Dretske denies the general closure principle that if one knows that P and knows that P implies Q, then one knows that Q.

Dretske's rejection of closure is based, on the one hand, on purportedly intuitive examples of closure failure, and on the other hand, on a theory of knowledge, the RA theory, which is supposed to explain why closure fails. Consider again the medical diagnosis example. If one accepts that in order to know that the patient's condition is C, it is not necessary to rule out the possibility of the extremely rare disease X that produces C-like test results, then one is close to a denial of closure. For suppose that the student knows that people with X have complete immunity to C, as assumed above, so $K(c \to \neg x)$. Since the student did not run any tests that could possibly detect the presence or absence of X, it would be unreasonable to claim that she knows that the patient does not have X, so $\neg K \neg x$. Then together with our judgment that the student knows that the patient has condition C, Kc, we have a clear violation of closure. To retain closure, one must either conclude that the student does not know the patient's condition after all or that one can know that a patient does not have a rare disease without running any tests at all. The first possibility leads to epistemic skepticism, while the second seems to lead to epistemic irresponsibility.

According to Dretske, the RA theory explains why closure fails. In order to know the premises c and $c \to \neg x$, the agent must rule out certain relevant alternatives. In order to know the conclusion $\neg x$, the agent must also rule out certain relevant alternatives. But the sets of relevant alternatives for the premises and the conclusion *are not the same*. We have already argued that X is not a relevant alternative that must be ruled out in order for Kc to hold. But X certainly is a relevant alternative that must be ruled out in order for $K\neg x$ to hold. It is because the relevant alternatives may be different for the premises and the conclusion that the epistemic closure principle does not hold in general.

In an influential article objecting to Dretske's claims of closure failure, G.C. Stine [16] argued that to allow for the set of relevant alternatives to be different for the premises and the conclusion of a modus ponens argument "would be to commit some logical sin akin to equivocation" (p. 256). Yet as Mark Heller [9] has pointed out in a more recent defence of closure denial on the basis of the RA theory, a similar charge of equivocation could be made (incorrectly) against accepted counterexamples to the principles of transitivity or antecedent strengthening of counterfactual conditionals. If we take a counterfactual conditional $\varphi \Rightarrow \psi$ to be true just in case the "closest" φ-worlds are ψ-worlds, then the inference from $\varphi \Rightarrow \psi$ to $\varphi \wedge \chi \Rightarrow \psi$ fails because the closest $\varphi \wedge \chi$-worlds may not be the same as the closest φ-worlds. Heller argues that there is no equivocation in these counterexamples since we use the same, fixed similarity ordering on the set of worlds to evaluate the different conditionals. Similarly, in the example of closure failure, the (most) relevant $\neg c$-worlds may differ from the (most) relevant x-worlds, even assuming a fixed relevance ordering over the set of worlds.

In the next section, starting from Heller's assumption of an ordering over the set of worlds that determines which alternatives are relevant, we will formalize two version of the RA theory, one due to Dretske [6,7], the other to Lewis [12]. Dretske's thesis that knowledge is a semi-penetrating operator is shown to be in tension with his version of the RA theory, while Lewis's version is naturally understood in terms of *dynamic* operations that change what is relevant through shifts in context. Through the formalization, we raise a question about the RA theory that has been relatively neglected in the epistemological literature, namely how the RA theorist should handle higher-order knowledge.

3 Formalizing the Relevant Alternatives Theory

There are many versions of the RA theory, differing with respect to the controversial issues of *what determines relevance* and *what it means to rule out* an alternative. Our formalization will be neutral on these issues. In our models, the relations that encode which possibilities are relevant and which have been ruled out are primitives, just as the indistinguishability relation is in epistemic logic.

One important distinction between different versions of the RA theory, which our formalization will capture, has to do with their logical structure. Dretske [7] introduces the following definition in developing his theory: "let us call the set of possible alternatives that a person must be in an evidential position to exclude (when he knows P) the *Relevancy Set*" (p. 371). It is clear from Dretske's definition and the discussion that follows that the choice of the relevancy set depends on P. By contrast, Heller [9] considers (and rejects) an interpretation of the RA theory according to which "there is a certain set of worlds selected as relevant, and S must be able to rule out the not-p worlds within that set" (p. 197). In this case, the choice of the set of relevant worlds does not depend on P.

The distinction here is of course one of the logician's favorites, the distinction between $\forall\exists$ and $\exists\forall$. Let us distinguish the following versions of the RA theory, where W is a set of possible worlds and a proposition P is understood as a subset of W (so that the complement of P in W, W/P, is the proposition not-P):

- $\mathsf{RA}_{\forall\exists}$: (for every context $\mathcal{C}$ and) for every proposition P, there is a relevancy set $R_{\mathcal{C}}(P) \subseteq (W/P)$ such that in order to know P one must rule out $R_{\mathcal{C}}(P)$.

- $\mathsf{RA}_{\exists\forall}$: (for every context $\mathcal{C}$) there is a set of relevant worlds $R_{\mathcal{C}}$ such that for every proposition P, in order to know P one must rule out $R_{\mathcal{C}} \cap (W/P)$.

Although we leave aside the details here, it is clear from the classic papers of Dretske [7] and Lewis [12] that Dretske was assuming $\mathsf{RA}_{\forall\exists}$, while Lewis was assuming $\mathsf{RA}_{\exists\forall}$. As we will see, this difference turns out to be at the heart of their disagreement about the closure of knowledge under known implication.

To begin our formalization, we introduce the basic language used throughout.

Definition 1. *Let* At *be a set of atomic sentence symbols. The language* $\mathcal{L}_{\mathsf{RA}}$ *is defined by*

$$\mathcal{L}_{\mathsf{RA}} := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box^{\sim}\varphi \mid \Box^{\prec}\varphi \mid U\varphi,$$

where $p \in \mathsf{At}$.

The intended readings of the modal formulas are: for $\Box^{\sim}\varphi$, "φ is true in every possibility consistent with the agent's information"; for $\Box^{\prec}\varphi$, "φ is true in every strictly-more-relevant possibility"; and for $U\varphi$, "φ is true in every possibility." We write $\Diamond^{\sim}$, $\Diamond^{\prec}$, and E for the duals of the three primitive modalities, respectively.

Definition 2. *A* basic RA model *is a tuple* $\mathcal{M} = \langle W, @, \sim, \preceq, V \rangle$ *where* W *is a non-empty set,* $@ \in W$*,* $\sim$ *is an equivalence relation on* W*,* $\preceq$ *is a total preorder on* W*,* $@$ *is maximal in* $\preceq$*, and* $V : \mathsf{At} \to \mathcal{P}(W)$ *is a valuation function.*

A basic RA model can be seen as an ordinary epistemic (partition) model with a distinguished "actual world" $@$ and a relevance (pre-)ordering $\preceq$ over the set of worlds. We refer to elements of W as "worlds," "possibilities," and "alternatives," interchangeably. We interpret $w \sim v$ to mean that possibility v is consistent with the agent's information at w; $\sim(w) = \{v \in W \mid w \sim v\}$ is the set of such possibilities. We interpret $w \preceq v$ to mean that alternative v is *at least as relevant* as alternative w; $\max_{\preceq}(S) = \{v \in S \mid u \preceq v \text{ for all } u \in S\}$ is the set of most relevant alternatives in a set $S \subseteq W$. For abbreviation, $w \prec v := w \preceq v \;\&\; v \not\preceq w$.

Ignoring the distinguished world $@$, our RA models are the same as the epistemic preference models of [3]. Note that in modeling preference, there is no reason to suppose that the actual world is maximal in the preference ordering. However, in the case of the RA theory, it is reasonable to assume that the actual world is maximal in the relevance ordering. Indeed, Lewis calls this the "*Rule of Actuality*...actuality is always a relevant alternative" [12, p. 554].

Definition 3. *Given an* RA *model* $\mathcal{M}$ *and* $\varphi \in \mathcal{L}_{\mathsf{RA}}$*, we define* $\mathcal{M}, w \vDash \varphi$ *and* $[\![\varphi]\!]_{\mathcal{M}} = \{v \in W \mid \mathcal{M}, v \vDash \varphi\}$ *as follows* (*with propositional cases as usual*):

$$\begin{array}{lll} \mathcal{M}, w \vDash \Box^{\sim}\varphi & \text{iff} & (W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset; \\ \mathcal{M}, w \vDash \Box^{\prec}\varphi & \text{iff} & \forall v : w \prec v \Rightarrow \mathcal{M}, v \vDash \varphi; \\ \mathcal{M}, w \vDash U\varphi & \text{iff} & \forall v : \mathcal{M}, v \vDash \varphi. \end{array}$$

We write $\vDash \varphi$ *for ordinary validity and* $\vDash_{@} \varphi$ *if for all models,* $\mathcal{M}, @ \vDash \varphi$.

The $\Box^{\sim}$ modality has the standard semantics of knowledge in epistemic logic. However, we write the clause for $\Box^{\sim}\varphi$ in a non-standard way, for comparison with our RA knowledge operators, defined below. While $\Box^{\sim}\varphi$ requires that the agent's information eliminates *all* $\neg\varphi$-possibilities, the RA knowledge operators will only require that the agent's information eliminates the *relevant* $\neg\varphi$-possibilities, which may be a proper subset of all $\neg\varphi$-possibilities. The modality $\Box^{\sim}$ could be interpreted as a kind of "Cartesian knowledge," in sense that Descartes (it is often said) held that knowledge requires eliminating all possibilities of error.

The language $\mathcal{L}_{\mathsf{RA}}$ and its semantics are very close to the language and semantics of "epistemic preference logic" in [3] and "modal preference logic" in [2], though the intuitive interpretations are different. A complete axiomatization of the validities in $\mathcal{L}_{\mathsf{RA}}$ is easily obtained from the axiomatizations for these logics, but we will not go into the details here.

3.1 Dretske-knowledge vs. Lewis-knowledge

In this section we investigate a Dretske-knowledge modality K_d and a Lewis-knowledge modality K_l, both of which turn out to be already definable in $\mathcal{L}_{\mathsf{RA}}$.

Definition 4. *The truth definitions for Dretske- and Lewis-knowledge are*:

$$(\mathsf{RA}_{\forall\exists})\ \ \mathcal{M}, w \vDash K_d\varphi \ \text{ iff } \ \max_{\preceq}(W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset;$$
$$(\mathsf{RA}_{\exists\forall})\ \ \mathcal{M}, w \vDash K_l\varphi \ \text{ iff } \ \max_{\preceq}(W) \cap (W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset.$$

In the definition of Dretske-knowledge, for any φ there is a set of relevant alternatives, namely *the most relevant* $\neg\varphi$*-worlds*, $\max_{\preceq}(W \setminus [\![\varphi]\!]_{\mathcal{M}})$, that an agent must rule out in order to know φ. For Lewis-knowledge, there is a set of relevant alternatives, $\max_{\preceq}(W)$, such that for any φ, to know φ an agent must rule out the $\neg\varphi$-worlds *in that set*. Recall the distinction between $\mathsf{RA}_{\forall\exists}$ and $\mathsf{RA}_{\exists\forall}$ above.

Proposition 1. (i) *$K_d\varphi$ is definable in $\mathcal{L}_{\mathsf{RA}}$ as $\Box^{\sim}(\neg\varphi \to \Diamond^{\prec}\neg\varphi)$.* (ii) *$K_l\varphi$ is definable in $\mathcal{L}_{\mathsf{RA}}$ as $\Box^{\sim}(\Box^{\prec}\bot \to \varphi)$.*

Proof. (i) $\mathcal{M}, w \vDash \Box^{\sim}(\neg\varphi \to \Diamond^{\prec}\neg\varphi)$ iff for all $v \in \sim(w)$, if $\mathcal{M}, v \vDash \neg\varphi$ then $\mathcal{M}, v \vDash \Diamond^{\prec}\neg\varphi$. This is equivalent to: for all $v \in \sim(w)$, $v \notin \max_{\preceq}(W \setminus [\![\varphi]\!]_{\mathcal{M}})$, i.e., $\max_{\preceq}(W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset$, which is the condition for $\mathcal{M}, w \vDash K_d\varphi$.

(ii) $\mathcal{M}, w \vDash \Box^{\sim}(\Box^{\prec}\bot \to \varphi)$ iff for all $v \in \sim(w)$, if $\mathcal{M}, v \vDash \Box^{\prec}\bot$ then $\mathcal{M}, v \vDash \varphi$. This is equivalent to: for all $v \in \sim(w)$, if $v \in \max_{\preceq}(W)$, then $v \in [\![\varphi]\!]_{\mathcal{M}}$, i.e., $\max_{\preceq}(W) \cap (W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset$, the condition for $\mathcal{M}, w \vDash K_l\varphi$.

We will now verify in parts (ii)-(iv) of the following proposition that K_d and K_l have the bare minimum of properties one would expect from a knowledge operator according to the RA theory: if the agent knows φ, then φ is true (ii); if the agent's information eliminates every $\neg\varphi$-possibility, then the agent knows that φ (iii); but—what is crucial for the RA theory—if the agent knows that φ, it need not be the case that the agent's information eliminates *every* $\neg\varphi$-possibility (iv), since some are not relevant.

Proposition 2. *Let K_* be either K_d or K_l. Then* (i) $\vDash K_d\varphi \to K_l\varphi$*,* (ii) $\vDash_{@} K_*\varphi \to \varphi$*,* (iii) $\vDash \Box^{\sim}\varphi \to K_*\varphi$*, and* (iv) $\nvDash K_*\varphi \to \Box^{\sim}\varphi$*.*

Proof. (i) $\max_{\preceq}(W) \cap (W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) \subseteq \max_{\preceq}(W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w)$.

(ii) By (i), it suffices to consider K_l: if $\mathcal{M}, @ \nvDash \varphi$, then given $@ \in \max_{\preceq}(W)$ and $@ \in \sim(@)$, we have $@ \in \max_{\preceq}(W) \cap (W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w)$, so $\mathcal{M}, @ \nvDash K_l\varphi$.

(iii) By (i), it suffices to consider K_d: if $\mathcal{M}, w \vDash \Box^{\sim}\varphi$ then $(W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset$, so $\max_{\preceq}(W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset$ and hence $\mathcal{M}, w \vDash K_d\varphi$.

(iv) By (i), it suffices to consider K_d: in a model $\mathcal{M}$ with $W = \{@, w, v\}$, $\sim(@) = \{@, v\}$, $v \prec w \prec @$, and $V(p) = \{@\}$, we have $\max_{\preceq}(W \setminus [\![p]\!]_{\mathcal{M}}) = \{w\}$ but $w \notin \sim(@)$, so $\mathcal{M}, @ \vDash K_d p$; yet $v \in (W \setminus [\![p]\!]_{\mathcal{M}}) \cap \sim(@)$, so $\mathcal{M}, @ \nvDash \Box^{\sim}p$.

To obtain a stronger version of (ii) such that $\vDash K_*\varphi \to \varphi$, it suffices to add to Definition 4 the requirement for $\mathcal{M}, w \vDash K_*\varphi$ that $w \in [\![\varphi]\!]_{\mathcal{M}}$. Yet it is not necessary to build the general veridicality of knowledge into the logic in this way, since the semantic modifications necessary to appropriately handle higher-order knowledge, discussed in Section 3.3 below, also give the result that $\vDash K_*\varphi \to \varphi$.

Given Dretske's claim that knowledge is not in general closed under known implication and Lewis's claim that it is (which we will consider further below), we should expect this difference to appear in our formalization of their different versions of the RA theory. The next proposition confirms that this is the case.

Proposition 3. *Closure under known implication fails for K_d and holds for K_l:*

1. $\not\vDash K_d\varphi \rightarrow (K_d\,(\varphi \rightarrow \psi) \rightarrow K_d\psi)$
2. $\vDash K_l\varphi \rightarrow (K_l\,(\varphi \rightarrow \psi) \rightarrow K_l\psi)$

Proof. Part 2 is immediate from the truth definition for K_l, since K_l simply acts as a normal modal operator with restricted access to $\max_{\preceq}(W)$. For part 1, consider the model in Figure 1, where arrows indicate the $\sim$ relation (with reflexive loops omitted) and the relevance ordering is indicated between points (it is not specified whether the relation between w_1 and w_2 is strict, since it does not matter for the argument). The formula K_dp is true at w_1, since the most relevant $\neg p$-world (w_2) is not accessible from w_1 via the $\sim$ relation. The formula $K_d(p \rightarrow q)$ is also true at w_1, since the most relevant $\neg(p \rightarrow q)$-world (w_4) is not accessible from w_1 via the $\sim$ relation. However, the most relevant $\neg q$-world (w_3) is accessible from w_1 via the $\sim$ relation, so K_dq is false at w_1.

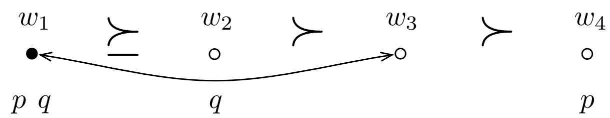


Fig. 1.

Proposition 3 captures the claim made by Dretske and defended by Heller that according to the RA theory, the knowledge operator is not a *fully penetrating* operator. It seems not to have been recognized in the literature that according to their version of the theory, the knowledge operator is not even *semi*-penetrating.

Proposition 4. *K_d is a non-penetrating operator*:

1. $\not\vDash K_d\,(\varphi \wedge \psi) \rightarrow K_d\varphi \wedge K_d\psi$
2. $\not\vDash K_d\varphi \rightarrow K_d\,(\varphi \vee \psi)$

Proof. The model in Figure 1 also shows the non-validity of the above formulas. For 1, the formula $K_d(p \wedge q)$ is true at w_1, since the most relevant $\neg(p \wedge q)$-world (w_2) is not accessible from w_1 via the $\sim$ relation. However, the most relevant $\neg q$-world (w_3) is accessible from w_1 via the $\sim$ relation, so K_dq is false at w_1. For 2, we have already seen that K_dp is true at w_1, yet the most relevant $\neg(p \vee q)$-world (w_3) is accessible from w_1 via the $\sim$ relation, so $K_d(p \vee q)$ is false at w_1.

It is important to realize that Proposition 4 is not an artifact of a particular choice of formal system. Although Dretske's [7] discussion of relevancy sets leaves open whether he would accept our picture in terms of a relevance ordering over possibilities, Heller's [9] defense of Dretske explicitly appeals to this picture. In light of Proposition 4, we may conclude that the Dretske-Heller version of the RA theory cannot sustain the claim that knowledge is semi-penetrating, which Dretske took to be the "trivial" side of his thesis. It remains a challenge for defenders of Dretske's views to formulate a plausible version of the RA theory according to which closure of knowledge under known implication fails while closure under conjunction elimination and disjunction introduction holds.

3.2 The Dynamics of Context Change

According to Lewis [12], knowledge is closed under known implication, but it can easily *seem* as if it is not. For sometimes one can truly attribute to an agent knowledge of some propositions, but one cannot go on to truly attribute to the agent knowledge of (what are known by the agent to be) consequence of those very propositions; the reason is that by raising the very issue of those consequences, one can *shift the context* of knowledge attribution from an "everyday" context to a context in which the bar for knowledge is suddenly higher.

For example, in our medical diagnosis case, although we initially said that the student knew the patient's condition to be C (Kc) and that the student had the background knowledge that the rare disease X confers complete immunity to C ($K(c \to \neg x)$), by *raising the possibility* of the rare disease X, which the student had not ruled out ($\neg K \neg x$), we thereby (according to Lewis) shifted the context in such a way that it would no longer be appropriate to attribute knowledge of C to the student. This looks like a failure of closure, but Lewis argues it is not:

> Knowledge *is* closed under implication.... Implication preserves truth—that is, it preserves truth in any given, fixed context. But if we switch contexts, all bets are off.... Dretske gets the phenomenon right...it is just that he misclassifies what he sees. He thinks it is a phenomenon of logic, when really it is a phenomenon of pragmatics. Closure, rightly understood, survives the rest. If we evaluate the conclusion for truth not with respect to the context in which it was uttered, but instead with respect to the different context in which the premise was uttered, then truth is preserved. (p. 564)

Lewis discusses other mechanisms that may shift the context of knowledge attribution, besides the explicit raising of hitherto ignored possibilities. He claims, for example, that if *what is at stake* in knowing increases, then the standards of knowledge might increase accordingly. As a result, alternatives that were previously irrelevant may become relevant. Fewer alternatives may be ignored.

To model the dynamic process whereby context shifts change the relevant alternatives, we will extend our language with dynamic modalities. Let $\mathcal{L}_{\mathsf{RA}^+}$ be obtained by adding to the grammar from Definition 1 a clause for formulas of the form $[+\varphi]\varphi$. We will read $[+\varphi]\psi$ as "after the issue of (whether or not) φ is raised, ψ is the case." We now define the corresponding model transformation.

Definition 5. *Given $\mathcal{M} = \langle W, @, \sim, \preceq, V\rangle$, $\mathcal{M}_{+\theta} = \left\langle W, @, \sim, \preceq^{+\theta}, V\right\rangle$ is obtained by changing the relevance relation $\preceq$ to $\preceq^{+\theta}$ as follows. The most relevant θ-states and the most relevant $\neg\theta$-states become equally relevant and most relevant overall, but otherwise the old ordering remains, i.e., for any $w, w' \in W$:*

1. If $w' \in \max_{\preceq}(\llbracket\theta\rrbracket_{\mathcal{M}}) \cup \max_{\preceq}(W \setminus \llbracket\theta\rrbracket_{\mathcal{M}})$, then $w \preceq^{+\theta} w'$.
2. If $w, w' \notin \max_{\preceq}(\llbracket\theta\rrbracket_{\mathcal{M}}) \cup \max_{\preceq}(W \setminus \llbracket\theta\rrbracket_{\mathcal{M}})$, then $w \preceq^{+\theta} w'$ iff $w \preceq w'$.

Since $\max_{\preceq}(W) \subseteq \max_{\preceq}(\llbracket\theta\rrbracket_{\mathcal{M}}) \cup \max_{\preceq}(W \setminus \llbracket\theta\rrbracket_{\mathcal{M}})$, if $w' \in \max_{\preceq}(W)$, then $w \preceq^{+\theta} w'$ for all w by 1. Hence $\max_{\preceq}(W) \subseteq \max_{\preceq^{+\theta}}(W)$.

Definition 6. *The truth definition for formulas with a dynamic modality is*:

$$\mathcal{M}, w \vDash [+\theta]\, \varphi \text{ iff } \mathcal{M}_{+\theta}, w \vDash \varphi.$$

As with the extension of $\mathcal{L}_{\mathsf{RA}}$ with K_d and K_l, the extension of $\mathcal{L}_{\mathsf{RA}}$ with these dynamic modalities turns out not to confer any additional expressivity.

Proposition 5. *Every* $\varphi \in \mathcal{L}_{\mathsf{RA}^+}$ *is reducible to an equivalent* $\varphi' \in \mathcal{L}_{\mathsf{RA}}$.

Proof. The following valid reduction axioms give a system for rewriting any $\varphi \in \mathcal{L}_{\mathsf{RA}^+}$ as an equivalent $\varphi' \in \mathcal{L}_{\mathsf{RA}}$:

$$\begin{array}{ll}
[+\theta]\, p \leftrightarrow p & [+\theta]\, I\varphi \leftrightarrow I\, [+\theta]\, \varphi \\
[+\theta]\, \neg\varphi \leftrightarrow \neg\, [+\theta]\, \varphi & [+\theta]\, U\varphi \leftrightarrow U\, [+\theta]\, \varphi \\
[+\theta]\, (\varphi \wedge \psi) \leftrightarrow [+\theta]\, \varphi \wedge [+\theta]\, \psi &
\end{array}$$

$$\begin{aligned}
[+\theta]\, \Box^{\prec}\varphi \leftrightarrow\; & \Box^{\prec} \perp \vee (\theta \wedge \Box^{\prec}\neg\theta) \vee (\neg\theta \wedge \Box^{\prec}\theta) \\
& \vee \Big[\Box^{\prec}\, [+\theta]\, \varphi \wedge U\, ((\theta \wedge \Box^{\prec}\neg\theta) \rightarrow [+\theta]\, \varphi) \\
& \quad \wedge U\, ((\neg\theta \wedge \Box^{\prec}\theta) \rightarrow [+\theta]\, \varphi)\Big]
\end{aligned}$$

We comment only on the last reduction axiom, as the others are standard (see, e.g., [3]). The lhs expresses that in the new model obtained by raising the issue of θ, all strictly more relevant worlds satisfy φ. The rhs expresses an equivalent conditions in terms of what is true in the original model. First, if the current world is among (a) the most relevant worlds in the original model (in which case $\Box^{\prec} \perp$ holds), then by the observation following Definition 5, it is also among the most relevant worlds in the new model, in which case the lhs is trivially true for any φ. The same reasoning applies if the current world is among (b) the most relevant θ-worlds (in which case $\theta \wedge \Box^{\prec}\neg\theta$ holds) or among (c) the most relevant $\neg\theta$-worlds (in which case $\neg\theta \wedge \Box^{\prec}\theta$ holds). On the other hand, if the current world is not among (a)-(c) in the original model, then the worlds that are more relevant than the current world in the *new* model are, first, those worlds that were more relevant than the current world in the original model, and second, worlds (b)-(c). For the lhs to be true, all of these worlds must satisfy φ in the new model, and this is exactly what the last disjunct on the rhs expresses.

It is immediate from the proposition that one obtains a complete axiomatization of validities in the language $\mathcal{L}_{\mathsf{RA}^+}$ from a complete axiomatization for the language $\mathcal{L}_{\mathsf{RA}}$ together with the reduction axioms.

Equipped with our new dynamic modalities, we can now state in the following proposition Lewis's position on the issue of closure under known implication.

Proposition 6. K_l *is closed under known implication with respect to a fixed context, but not closed under known implication across context changes:*

$$\begin{array}{l}
\vDash K_l\varphi \rightarrow (K_l\, (\varphi \rightarrow \psi) \rightarrow K\psi) \\
\nvDash K_l\varphi \rightarrow [+\psi]\, (K_l\, (\varphi \rightarrow \psi) \rightarrow K\psi)
\end{array}$$

Proof. For the second part, use Figure 1 and substitute p for φ and q for ψ.

The second part of the proposition captures Lewis's view that by raising the issue of whether (what the agent knows to be) a consequence of what the agent knows is true, one can change the context and increase the standards of knowledge in such a way that we cannot ascribe knowledge of the consequence to the agent. Of course, it follows from the first part of the proposition that in the new, more demanding context, we can no longer attribute to the agent knowledge of the original fact of which the second was a consequence. Hence the title of Lewis's paper [12], "Elusive Knowledge," for "when we do epistemology, and we attend to the proper ignoring of possibilities, we make knowledge vanish" (p. 566).

3.3 Higher-order Knowledge

By formalizing the RA theory in the framework of epistemic logic, we are led naturally to the question how of the theory handles *higher-order* knowledge. Despite much discussion in epistemology of the general "KK thesis", corresponding to the **4** axiom $K\varphi \to KK\varphi$ in epistemic logic, the question of how the RA theory in particular handles higher-order knowledge has received less attention.

Proposition 7. *K_d and K_l satisfy positive ($\vDash K_*\varphi \to\ K_*K_*\varphi$) and negative ($\vDash \neg K_*\varphi \to\ K_*\neg K_*\varphi$) introspection.*

Proof. Something stronger is true, namely $\vDash K_*\varphi \leftrightarrow \Box^{\sim} K_*\varphi$ and $\vDash \neg K_*\varphi \leftrightarrow \Box^{\sim}\neg K_*\varphi$, from which the proposition follows using Proposition 2(iii). By the truth definitions and the fact ($\star$) that if $v \in \sim(w)$, then $\sim(v) = \sim(w)$, the case of $\vDash K_l\varphi \leftrightarrow \Box^{\sim} K_l\varphi$ is given by: $\mathcal{M}, w \vDash \Box^{\sim} K_l\varphi \Leftrightarrow \forall v \in \sim(w) : \mathcal{M}, v \vDash K_l\varphi \Leftrightarrow \forall v \in \sim(w)\, \forall u \in \max_{\preceq}(\sim(v)) : \mathcal{M}, u \vDash \varphi \Leftrightarrow_{(\star)} \forall u \in \max_{\preceq}(\sim(w)) : \mathcal{M}, u \vDash \varphi \Leftrightarrow \mathcal{M}, w \vDash K_l\varphi$. The other cases also follow easily using ($\star$).

For reasons noted below, one should not expect such introspection properties for knowledge in the RA theory. They are the result of a simplifying assumption in our semantics, which we remove in the definition of *general* RA models below. The proof of Proposition 7 also shows that the basic semantics is inappropriate if we wish to consider *information about knowledge.* From Proposition 2(iv), $K_*\varphi \wedge \Diamond^{\sim}\neg\varphi$ is satisfiable, as it should be in the RA theory. Yet together with $\vDash K_*\varphi \leftrightarrow \Box^{\sim} K_*\varphi$, this shows that $\Diamond^{\sim}(K_*\varphi \wedge \neg\varphi)$ is satisfiable, which is highly counterintuitive; there should not be any possibility consistent with the agent's information—however remote—in which the agent knows a falsehood. The technical source of the problem is that when a formula containing K_* is evaluated at a world, the K_* modality can only access the relevant worlds in the model, even if such a formula is evaluated at an *irrelevant* world. Yet the $\Diamond^{\sim}$ modality can also access irrelevant worlds, which explains the satisfiability of $\Diamond^{\sim}(K_*\varphi \wedge \neg\varphi)$.

The conceptual question is how an agent's knowledge is to be determined in irrelevant worlds. A simple proposal is that while in relevant worlds, knowing φ only requires that φ be true in all relevant worlds consistent with one's information, in irrelevant worlds, knowing φ requires that φ be true in *irrelevant* worlds as well. Let us modify Lewis-knowledge in this way, adding a new modality K_l.

Definition 7. *The truth definition for modified Lewis-knowledge is*:

$$\mathcal{M}, w \vDash \mathsf{K}_l\varphi \text{ iff } \begin{cases} w \in \max_{\preceq}(W) \Rightarrow \max_{\preceq}(W) \cap (W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset \\ w \notin \max_{\preceq}(W) \Rightarrow (W \setminus [\![\varphi]\!]_{\mathcal{M}}) \cap \sim(w) = \emptyset \end{cases}$$

Proposition 8. $\mathsf{K}_l\varphi$ *is definable in* $\mathcal{L}_{\mathsf{RA}}$ *as* $(\Box^{\prec} \bot \wedge K_l\varphi) \vee (\Diamond^{\prec}\top \wedge \Box^{\sim}\varphi)$.

With the modified definition of knowledge, $\vDash \mathsf{K}_l\varphi \rightarrow \varphi$ (c.f. Proposition 2(ii)), and the interaction between knowledge and information becomes very natural.

Proposition 9. *The following hold for* K_l *as defined in* Definition 7:

(1) $\vDash \mathsf{K}_l\varphi \leftrightarrow \mathsf{K}_l\mathsf{K}_l\varphi$	(3) $\vDash \Box^{\sim}\varphi \leftrightarrow \Box^{\sim}\mathsf{K}_l\varphi$	(5) $\nvDash \mathsf{K}_l\varphi \rightarrow \Box^{\sim}\mathsf{K}_l\varphi$
(2) $\nvDash \neg\mathsf{K}_l\varphi \rightarrow \mathsf{K}_l\neg\mathsf{K}_l\varphi$	(4) $\vDash \Box^{\sim}\varphi \leftrightarrow \mathsf{K}_l\Box^{\sim}\varphi$	(6) $\nvDash \neg\mathsf{K}_l\varphi \rightarrow \Box^{\sim}\neg\mathsf{K}_l\varphi$

While K_l interacts naturally with $\Box^{\sim}$, from the point of view of the RA theory, it is questionable whether positive introspection (1) should hold for knowledge. (To see why negative introspection already fails, suppose all worlds satisfy φ except an "irrelevant" $w \notin \max_{\preceq}(W)$, and evaluate $\mathsf{K}_l\neg\mathsf{K}_l\varphi$ at w.) On a natural interpretation of the RA theory applied to higher-order knowledge, in order for an agent to know that she knows φ, she should have to know that she has ruled out the relevant alternatives to φ. By having only a single relevance pre-order $\preceq$ in our basic models, we implicitly assume that the agent knows exactly what the relevant alternatives are. Yet reflection on the medical diagnosis case shows that agents may be uncertain about the relevant alternatives and hence uncertain about whether they have ruled out the alternatives necessary for knowledge.

To model higher-order knowledge in a way that is faithful to the RA theory, we must consider models in which each world has an associated relevance preorder $\preceq_w$ over the set of all worlds, so that agents may be uncertain about what are the *actual* relevant alternatives. Formally, a *general* RA model is a tuple $\mathcal{M} = \langle W, \sim, \{\preceq_w\}_{w \in W}, V \rangle$ where $\mathcal{M} = \langle W, \sim, V \rangle$ is an epistemic (partition) model and each $\preceq_w$ is a total pre-order on W such that w is maximal in $\preceq_w$. (Very similar models, without the assumption that w is maximal in $\preceq_w$, are used for logics of belief revision [4,1] and logics of counterfactual conditionals [5].) Modifying the semantics for K_d and K_l in Definition 4 such that $\preceq$ is replaced by $\preceq_w$, it is not difficult to see that $K_d\varphi$ and $K_l\varphi$ are no longer definable in $\mathcal{L}_{\mathsf{RA}}$. We leave it for future work to give an axiomatization in this setting.

4 Conclusion

Despite the expressive limitations of $\mathcal{L}_{\mathsf{RA}}$ with respect to general models for higher-order knowledge, this paper shows that it is possible to go a long way in modeling different epistemological notions from a rather minimal logical base. It also shows how a formalization in the framework of epistemic logic can clarify a contentious epistemological issue concerning RA theories and closure. The philosophical payoff does not end here. Extending the system to model the interaction of *belief* and knowledge leads to other areas of application, such as the "problem of missed clues" [15] for RA theories, to be investigated in future work.

Acknowledgements. I wish to thank Tomohiro Hoshi and Thomas Icard for helpful comments on an earlier draft of this paper.

References

1. Johan van Benthem. Dynamic logic of belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–156, 2007.
2. Johan van Benthem, Patrick Girard, and Olivier Roy. Everything else being equal: A modal logic for *Ceteris Paribus* preferences. *Journal of Philosophical Logic*, 38:83–125, 2009.
3. Johan van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17(2):157–182, 2007.
4. Oliver Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49:49–80, 2004.
5. John P. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22(1):76–84, 1981.
6. Fred Dretske. Epistemic operators. *The Journal of Philosophy*, 67(24):1007–1023, 1970.
7. Fred Dretske. The pragmatic dimension of knowledge. *Philosophical Studies*, 40:363–378, 1981.
8. Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.
9. Marc Heller. Relevant alternatives and closure. *Australasian Journal of Philosophy*, 77(2):196–208, 1999.
10. Vincent F. Hendricks and John Symons. Where's the bridge? Epistemology and epistemic logics. *Philosophical Studies*, 128:137–167, 2006.
11. Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. College Publications, 2005.
12. David Lewis. Elusive knowledge. *Australasian Journal of Philosophy*, 74(4):549–567, 1996.
13. J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*, volume 41 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1995.
14. Duncan Pritchard. Closure and context. *Australasian Journal of Philosophy*, 78(2):275–280, 2000.
15. Jonathan Schaffer. Knowledge, relevant alternatives and missed clues. *Analysis*, 62(3):202–208, 2001.
16. G.C. Stine. Skepticism, relevant alternatives, and deductive closure. *Philosophical Studies*, 29:249–261, 1976.

Comparing Inconsistency Resolutions in Multi-Context Systems

Antonius Weinzierl*

Knowledge-based Systems Group, Vienna University of Technology
weinzierl@kr.tuwien.ac.at

Abstract. Inconsistency in heterogeneous knowledge-integration systems with non-monotonic information exchange is a major concern as it renders systems useless at its occurrence. The problem of finding all possible resolutions to inconsistency has been addressed previously and some basic steps have been proposed to find most preferred resolutions. Here, we refine the techniques of finding preferred resolutions of inconsistency in two directions. First, we extend qualitative methods using domain knowledge about the intention and category of information exchange to minimize the number of categories that are affected by a resolution. Second, we present ways to compute a quantitative inconsistency measure on resolutions, each being suitable for certain application scenarios.

1 Introduction

Knowledge integration frameworks are essential for combining information from different knowledge bases. Multi-Context Systems (MCSs) introduced in [1] are a powerful framework for non-monotonic information exchange between heterogeneous knowledge bases. They extend MultiLanguage systems of [2] by allowing non-monotonic information exchange. Information in the MCS framework is exchanged via bridge rules of the form

$$(k : s) \leftarrow (c_1 : p_1), \ldots, (c_j : p_j), \mathbf{not}\ (c_{j+1} : p_{j+1}), \ldots, \mathbf{not}\ (c_m : p_m) \qquad (1)$$

which state that information s is added to knowledge base k whenever information p_i is present in knowledge base c_i (for $1 \leq i \leq j$) and information p_l is not present in knowledge base p_l (for $j < l \leq m$).

In this work we advance and refine previously introduced methods of finding preferred resolutions to inconsistency in MCSs (cf. [3] and [4]). Inconsistency is of major interest as it can render logic-based systems useless and, it occurs easily due to unanticipated side effects of information exchange Therefore we consider faulty information exchange, i.e., bridge rules, as reason of inconsistency. Several strategies to cope with inconsistency have been developed. For example,para-consistent reasoning, where inconsistency is often treated purely technically, i.e., any way to resolve the inconsistency is considered good enough. For real scenarios,

* Sponsored by the Vienna Science and Technology Fund (WWTF) grant: ICT-08-020

e.g., an MCS employed in a hospital, this is not acceptable as inconsistencies have to be resolved respecting additional domain knowledge.

Consider the case of an MCS used in a hospital to give decision-support on patient medication and handling the billing process. Assume there is a patient needing a certain medication, say human insulin, because she has severe hyperglycemia (high blood sugar) and she is allergic to animal insulin. If the billing system refuses treatment with human insulin, because the patient's medical insurance does not cover it then the system becomes inconsistent. There are several resolutions of that inconsistency, either modify the information flow to the billing system or ignore the patient's needs and treat her with the wrong medication. Technically both resolutions are fine, but the patient may feel different.

Following common terminology, we call the resolution to inconsistency a diagnosis. In [4] the problem of finding preferred diagnoses is addressed in general by a) specifying ways to compare diagnoses using domain knowledge, b) defining a quantitative inconsistency value for bridge rules. We advance this work by:

a) introducing a preference relation on diagnoses using domain knowledge on the intention of bridge rules. We categorize bridge rules by their intention, e.g. rules that exchange information about medication make up one category, billing is another one; we also track the dependency of categories. Preferred diagnoses are those that modify the least amount of categories.

b) extending the quantitative inconsistency value from bridge rules to diagnoses. We discuss and motivate different ways to achieve this.

By that, we give a) a useful method employing domain knowledge which is (implicitly) present in each system, and b) notions to use the quantitative inconsistency value for selection of preferred diagnoses. Both steps are towards a policy language for (semi-)automatic inconsistency handling.

The remainder of this paper is organized as follows: Section 2 defines MCS and diagnoses, Section 3 introduces bridge rule categorization and a comparison relation for diagnoses based on this categorization, Section 4 defines quantitative measures of inconsistency on diagnoses, and in Section 5 we conclude and discuss related and future work.

2 Preliminaries

A heterogeneous non-monotonic MCS [1]. consists of *contexts*, each composed of a knowledge base with an underlying *logic*, and a set of *bridge rules* which control the information flow between contexts.

A logic $L = (\mathbf{KB}_L, \mathbf{BS}_L, \mathbf{ACC}_L)$ consists, in an abstract view, of the following components:

- $\mathbf{KB}_L$ is the set of well-formed knowledge bases of L. We assume each element of $\mathbf{KB}_L$ is a set (of "formulas").
- $\mathbf{BS}_L$ is the set of possible belief sets, where the elements of a belief set are "formulas".
- $\mathbf{ACC}_L : \mathbf{KB}_L \rightarrow 2^{\mathbf{BS}_L}$ is a function describing the "semantics" of the logic by assigning to each knowledge base a set of acceptable belief sets.

This concept of a *logic* captures many monotonic and non-monotonic logics, e.g., classical logic, description logics, modal, default, and autoepistemic logics, circumscription, and logic programs under the answer set semantics.

We only use disjunctive answer-set programs (ASP) [5] as context logics, therefore $\mathbf{KB}$ is the set of disjunctive logic programs over Σ, $\mathbf{BS}$ is the set of sets of atoms over Σ, and $\mathbf{ACC}(kb)$ returns the set of answer sets of kb.

A *bridge rule* can add information to a context, depending on the belief sets which are accepted at other contexts. Let $L = (L_1, \ldots, L_n)$ be a sequence of logics. An L_k-bridge rule r over L is of the form (1) where $1 \leq c_i \leq n$, p_i is an element of some belief set of L_{c_i}, k refers to the context receiving information s. We denote by $h_{hd}(r)$ the belief formula s in the head of r.

A *Multi-Context System* $M = (C_1, \ldots, C_n)$ is a collection of contexts $C_i = (L_i, kb_i, br_i)$, $1 \leq i \leq n$, where $L_i = (\mathbf{KB}_i, \mathbf{BS}_i, \mathbf{ACC}_i)$ is a logic, $kb_i \in \mathbf{KB}_i$ a knowledge base, and br_i is a set of L_i-bridge rules over $(L_1, \ldots, L_n)$. For each $H \subseteq \{h_{hd}(r) \mid r \in br_i\}$ it holds that $kb_i \cup H \in \mathbf{KB}_{L_i}$, i.e., bridge rule heads are compatible with knowledge bases.

A *belief state* of an MCS $M = (C_1, \ldots, C_n)$ is a sequence $S = (S_1, \ldots, S_n)$ such that $S_i \in \mathbf{BS}_i$. A bridge rule (1) is applicable in S iff for $1 \leq i \leq j$: $p_i \in S_i$ and for $j < l \leq m$: $p_l \notin S_l$. $br_M = \bigcup_{i=1}^{n} br_i$ denotes the set of bridge rules of M.

Example 1. Let M be an MCS handling patient treatments and billing in a hospital; it contains the following contexts: a patient database C_1, a program C_2 suggesting proper medication, and a program C_3 handling the billing. Knowledge bases for these contexts are:

$$
\begin{aligned}
kb_1 =& \{\mathit{hyperglycemia.}\ \mathit{insurance\_B.}\},\\
kb_2 =& \{\mathit{give\_human\_insulin} \vee \mathit{give\_animal\_insulin} \leftarrow \mathit{hyperglycemia.}\\
& \quad \bot \leftarrow \mathit{give\_animal\_insulin}, \mathit{not}\ \ \mathit{allow\_animal\_insulin}\}.\\
kb_3 =& \{\mathit{bill} \leftarrow \mathit{bill\_animal\_insulin.}\ \mathit{bill\_more} \leftarrow \mathit{bill\_human\_insulin.}\\
& \quad \bot \leftarrow \mathit{insurance\_B}, \mathit{bill\_more.}\}
\end{aligned}
$$

Context C_1 provides information that the patient has severe hyperglycemia, and her health insurance is from company B. Context C_2 suggests to apply either human or animal insulin if the patient has hyperglycemia and requires that the applied insulin does not cause an allergic reaction. Context C_3 does the billing and encodes that insurance B only pays animal insulin. Bridge rules of M are:

$$
\begin{aligned}
r_1 &= (2 : \mathit{hyperglycemia}) && \leftarrow (1 : \mathit{hyperglycemia}).\\
r_2 &= (2 : \mathit{allow\_animal\_insulin}) && \leftarrow \mathbf{not}\ (1 : \mathit{allergic\_animal\_insulin}).\\
r_3 &= (3 : \mathit{bill\_animal\_insulin}) && \leftarrow (2 : \mathit{give\_animal\_insulin}).\\
r_4 &= (3 : \mathit{bill\_human\_insulin}) && \leftarrow (2 : \mathit{give\_human\_insulin}).\\
r_5 &= (3 : \mathit{insurance\_B}) && \leftarrow (1 : \mathit{insurance\_B}).
\end{aligned}
$$

Equilibrium semantics selects certain belief states of an MCS M as acceptable. Intuitively, an equilibrium is a belief state S, where each context C_i takes the heads of all bridge rules that are applicable in S into account, and accepts S_i. Formally, $S = (S_1, \ldots, S_n)$ is an equilibrium of M, iff for all $1 \leq i \leq n$:

$$S_i \in \mathbf{ACC}_i\left(kb_i \cup \{h_{hd}\left(r\right) \mid r \in br_i \text{ applicable in } S\}\right)$$

Inconsistency in an MCS is the lack of an equilibrium.

Example 2. In our example, one equilibrium S exists:

$$\begin{aligned} S = (&\{\mathit{hyperglycemia}, \mathit{insurance\_B}\},\\ &\{\mathit{give\_animal\_insulin}, \mathit{allow\_animal\_insulin}, \mathit{hyperglycemia}\},\\ &\{\mathit{bill}, \mathit{bill\_animal\_insulin}, \mathit{insurance\_B}\}). \end{aligned}$$

Rules r_1, r_2, r_3, and r_5 are applicable in S.

Example 3. As running example, we consider a slightly modified version of Example 1, with the patient being allergic to animal insulin:

$$kb_1 = \{\mathit{allergic\_animal\_insulin}, \mathit{hyperglycemia}, \mathit{insurance\_B}\}.$$

The MCS is inconsistent as r_2 becomes applicable, forcing C_2 to treat the patient with human insulin, which makes r_4 applicable and finally C_3 inconsistent.

We will use the following notation. Given an MCS M and a set R of bridge rules (compatible with M), by $M[R]$ we denote the MCS obtained from M by replacing its set of bridge rules br_M with R (e.g., $M[br_M] = M$ and $M[\emptyset]$ is M with no bridge rules). By $M \models \bot$ we denote that M has no equilibrium, i.e., is inconsistent, and by $M \not\models \bot$ the opposite. For any set of bridge rules A, $\mathit{heads}(A) = \{\alpha \leftarrow \top \mid \alpha \leftarrow \beta \in A\}$ are the rules in A in unconditional form.

Diagnoses. A diagnosis identifies a part of the bridge rules that need to be changed to restore consistency. In non-monotonic reasoning, adding or removing knowledge can both cause and prevent inconsistency. Therefore, a diagnosis is a pair of sets of bridge rules such that if the rules in the first set are removed, and the rules in the second set are added in unconditional form, the MCS becomes consistent (i.e., it admits an equilibrium). Formally: given an MCS M, a *diagnosis* of M is a pair (D_1, D_2), $D_1, D_2 \subseteq br_M$, s.t. $M[br_M \setminus D_1 \cup \mathit{heads}(D_2)] \not\models \bot$; by $D^{\pm}(M)$ we denote the set of all diagnoses. To obtain a more relevant set of diagnoses, pointwise subset-minimal diagnoses are preferred: we denote by $D^{\pm}_m(M)$ the set of all such diagnoses of an MCS M.

Example 4. In our running example,

$$D^{\pm}_m(M) = \{(\{r_1\}, \emptyset), (\{r_4\}, \emptyset), (\{r_5\}, \emptyset), (\emptyset, \{r_2\})\}.$$

Accordingly, deactivating r_1, r_4, r_5, or adding r_2 unconditionally, respectively, results in a consistent MCS. This means ignoring the illness of the patient, ignoring the application of human insulin, ignoring the insurance company or considering the patient to be not allergic.

3 Assessment with Categories

In this section we introduce categories as a method to compare diagnoses on qualitative terms. We rely on preference orders as defined in [4] for their realization, i.e., partial orders over diagnoses. Preference orders allow to compare diagnoses in general, based on the rules they modify. This covers statements like "proper treatment of patients is more important than correct billing", trust relations, or any other preference relation over diagnoses.

Definition 1 (Eiter et al. (cf. [4])). *Let M an MCS, a* preference order *for M is a transitive binary relation $\prec$ on $2^{br_M} \times 2^{br_M}$.*

Such orders may be realized as follows: Transform the given MCS M into an MCS M^t which is such that any diagnosis D of M corresponds one-to-one to a diagnosis D^t of M^t and for two diagnoses D_1^t, D_2^t of M^t holds that $D_1^t \subset D_2^t$ iff $D_1 \prec D_2$, i.e., each most preferred diagnosis wrt. $\prec$ corresponds to a subset-minimal diagnosis of M^t. For space reasons, we omit the details of the transformation and refer the reader to [4]. Notably, the transformation is efficient for many preference orders.

Categories. In logic programming a rule by itself most often is not useful, but only several rules together form a specific behaviour and cover an intended meaning. We assume that bridge rules are used similarly. In our example rules r_1 and r_2 carry the information of how to treat a patient correctly and we call this a category of the MCS. Category names are arbitrary, including the possibility of a syntactic derivation from the MCS, e.g., by a partitioning of beliefs.

Definition 2. *Let $\mathcal{C}$ be the set of category names, M an MCS, and for each $r \in br_M$ let $cat(r) \subseteq \mathcal{C}$ be an association of bridge rules to category names. Then $Cat_M = \bigcup_{r \in br_M} cat(r)$ are the categories of M.*

If a bridge rule is modified by a diagnosis, it is very likely that the semantics of all categories where the bridge rule is part of, are modified and possibly corrupted. Reconsider our example, if r_2 is modified, the patient not only is given a different treatment, but also the billing gives other results than expected – although correct under the modified assumptions. Therefore categories can depend on each other, e.g., category "billing" depends on category "treatment" and modifications of rules of the latter also change the result of the former. So we also consider dependencies among categories. Let Cat_M be the categories of an MCS M. Each $c \in Cat_M$ is associated a set of categories $P_c \subseteq Cat_M$ on which it depends. We write $dep(x, y)$ iff $x \in Cat_M$ and $y \in P_c$.

Note that the dependency of categories as well as their names and associations are semantic information, so for an MCS several categorizations may be adequate. As different categorizations may lead to other diagnoses being preferred, we assume in the following that a categorization deemed correct for the given MCS is applied. Finding such a categorization is to be adressed in the future.

Definition 3. *Let M be an MCS, Cat_M its categories with dependencies dep, and D be a diagnosis. The set of* possibly corrupted categories *of M wrt. D is the smallest set $C_D \subseteq Cat_M$ s.t. for all $r \in D$ holds $cat(r) \in C_D$ and whenever $y \in C_D$ and $dep(x, y)$ then $x \in C_D$.*

Obviously, a diagnosis which modifies a smaller set of categories is always desirable, as it guarantees that more parts of the diagnosed system still yield the expected results. This induces a preference order which is such that preferred diagnoses modify only a minimal set of categories

Definition 4. *Let $D, D' \in D^{\pm}(M)$ be diagnoses of an MCS M. D is preferred over D' iff $C_D \subset C_{D'}$. We denote this preference order by $D \prec_{sd} D'$.*

Example 5. We assign to our example MCS the set of category names $Cat_M = \{treatment, billing\}$ where $cat(r_1) = cat(r_2) = treatment$, $cat(r_3) = cat(r_4) = cat(r_5) = billing$ and dependency is given by $dep(billing, treatment)$. This categorization naturally follows from what the bridge rules are intended to do. Such information is actually present for every MCS; at least at the time of its creation.

Assuming that all categories are of equal importance, one can strengthen the above notion by requiring that a preferred diagnosis modifies only the least amount of categories, i.e., select by cardinality minimality. Cardinality-based preference can drastically reduce the number of diagnoses to be considered. So it is easier for human operators to select the best from the remaining diagnoses.

Definition 5. *Let $D, D' \in D^{\pm}(M)$ be diagnoses of an MCS M. D is preferred over D' iff $|C_D| < |C_{D'}|$. This is denoted by $D \prec_{|sd|} D'$.*

Example 6. With $\prec_{sd}$ (or $\prec_{|sd|}$), preferred diagnoses are $(\{r_4\}, \emptyset)$ and $(\{r_5\}, \emptyset)$.

4 Assessment with Quantitative Measures

The quantitative inconsistency measure for bridge rules is based on the notion MIV_C from [6], which employs cardinalities of the minimal inconsistent sets a certain formula belongs to. For MCSs an equivalent notion of a minimal inconsistent set is defined in [3] as inconsistency explanation. It is a pair of sets of bridge rules, whose presence resp. absence causes a relevant inconsistency.

An *inconsistency explanation* of an MCS M is a pair $(E_1, E_2) \in br_M \times br_M$ s.t. for all (R_1, R_2) where $E_1 \subseteq R_1 \subseteq br_M$ and $R_2 \subseteq br_M \setminus E_2$, it holds that $M[R_1 \cup heads(R_2)] \models \bot$. The set of all pointwise subset-minimal such (E_1, E_2) is denoted by $E^{\pm}_m(M)$.

The intuition is that $M[E_1]$ is inconsistent, and this inconsistency is relevant for M, as adding more bridge rules from br_M never resolves that inconsistency. Moreover, the inconsistency of M entailed by E_1 cannot be avoided by adding bridge rules unconditionally, unless bridge rules from E_2 are used.

As a bridge rule r may introduce and prevent inconsistency, we define the inconsistency value m_{br} of r as a pair (I_1, I_2) where I_1 and I_2 measure the amount of inconsistency caused, respectively, prevented by r.

Definition 6 (Eiter et al. (cf. [4]))**.** *Let M a MCS and $r \in br_M$, and let $A^i_r(M) = \{(E_1, E_2) \in E^{\pm}_m(M) \mid r \in E_i\}$, $i = 1, 2$. Then*

$$m_{\mathrm{br}}(M, r) = \Big(\sum_{(E_1,E_2) \in A^1_r(M)} \frac{1}{|E_1|}, \sum_{(E_1,E_2) \in A^2_r(M)} \frac{1}{|E_2|} \Big).$$

Example 7. There is one minimal inconsistency explanation: $(\{r_1, r_4, r_5\}, \{r_2\})$. So the inconsistency values are: $m_{\mathrm{br}}(M, r_1) = \left(\frac{1}{3}, 0\right)$, $m_{\mathrm{br}}(M, r_2) = (0, 1)$, $m_{\mathrm{br}}(M, r_3) = (0, 0)$, $m_{\mathrm{br}}(M, r_4) = \left(\frac{1}{3}, 0\right)$, and $m_{\mathrm{br}}(M, r_5) = \left(\frac{1}{3}, 0\right)$.

Given a quantitative measure on bridge rules, we derive quantitative measures on diagnoses. This allows to select preferable diagnoses without additional domain knowledge as well as to select preferable diagnoses that are considered incomparable or equal by measures based on domain knowledge.

On the one hand, subset-minimal diagnoses which remove the most inconsistency are preferable as they yield a most "clean" system. This may be the method of choice for "stable" systems that should not give rise to inconsistency when further modifications are applied. On the other hand, potential inconsistencies still carry some kind of information which therefore should not be removed without the need to, i.e., subset-minimal diagnoses removing the least amount of inconsistency are then preferable.

Definition 7. *Let M be an MCS, π_i be a projection to the ith element of tuples, and $m_{\mathrm{br}} : br_M \to \mathbb{R} \times \mathbb{R}$ a measure on bridge rules, then $m_d : 2^{br_M} \times 2^{br_M} \to \mathbb{R}$ is a measure on diagnoses with*

$$m_{\mathrm{d}}(D_1, D_2) = \sum_{r \in D_1} \pi_1 \left(m_{\mathrm{br}}\left(r\right)\right) + \sum_{r \in D_2} \pi_2 \left(m_{\mathrm{br}}\left(r\right)\right).$$

If diagnoses that remove the most inconsistency are preferred, one gets

$$D_{C^+} = \arg\max_{D \in D_m^{\pm}} \{m_{\mathrm{d}}\left(D\right)\}.$$

Preferring diagnoses that remove a least amount of inconsistency gives

$$D_{C^-} = \arg\min_{D \in D_m^{\pm}} \{m_{\mathrm{d}}\left(D\right)\}.$$

Example 8. In our running example $D_{C^+} = \{(\emptyset, \{r_2\})\}$ and $D_{C^-} = \{(\{r_1\}, \emptyset), (\{r_3\}, \emptyset), (\{r_5\}, \emptyset)\}$. Note that $(\emptyset, \{r_2\})$ removes the most inconsistency as r_2 alone makes the system consistent, while for D_{C^-} each rule r_1, r_3, r_5 only contributes one third to the cause of inconsistency.

Observe that we take into account how a bridge rule appears in a diagnosis, so either the value for removing or the value for adding it unconditional is counted. As removing a bridge rule also removes the ability to restore consistency with that rule, one may combine both values to obtain a different measure: $m'_{\mathrm{d}}(D_1, D_2) = \sum_{r \in D_1 \cup D_2} \pi_1 \left(m_{\mathrm{br}}\left(r\right)\right) - \pi_2 \left(m_{\mathrm{br}}\left(r\right)\right)$.

5 Related Work and Conclusion

In this paper we advanced the available methods of finding preferred resolutions of inconsistency in knowledge-integration systems by a) introducing categorizations of bridge rules for qualitative assessment, and b) establishing notions of quantitative inconsistency measures.

Related work: In [7] the problem of inconsistency in MCSs is addressed using defeasible rules which are applicable only if they do not cause inconsistency. In the presence of inconsistency defeasible rules are deactivated using additional trust information. Four algorithms to compute trust are given where the first uses a total order over contexts while the others also employ provenance information of increasing depth. Provenance, however, requires insight to context internals which is in conflict with our requirements of information hiding and privacy.

Work on distributed ontologies bears some similarities to our work, consider e.g., [8] where bridge rules represent ontology mappings and a notion of minimal diagnosis is used to repair inconsistent mappings. While our categorizations are similar to concepts in ontologies, the scopes of the works are different as we seek preference criteria for diagnoses on bridge rules between heterogeneous logics.

Inconsistency tolerance in peer-to-peer systems (e.g., [9]) considers homogeneous logics only and inconsistency resolution is local to each peer while our notion of minimal diagnosis is globally minimal.

Concerning the MCS framework and global inconsistency management, no further work addressing inconsistency management is known to us, although finding preferred diagnoses is an ubiquitous task in inconsistency handling.

Future work: Inconsistency could be resolved by seeking an equilibrium that is as close as possible to the original semantics of the inconsistent system. A first step towards this is an investigation of how the semantics of an inconsistent system can be described. Other aims are to improve the notion of categories for the aspects of: allowing arbitrary importance of categories, how categories can be derived from given MCSs, and how multiple categorizations of a system can be handled.

References

1. Brewka, G., Eiter, T.: Equilibria in heterogeneous nonmonotonic multi-context systems. In: AAAI, AAAI Press (2007) 385–390
2. Giunchiglia, F., Serafini, L.: Multilanguage hierarchical logics or: How we can do without modal logics. Artif. Intell. **65**(1) (1994) 29–70
3. Eiter, T., Fink, M., Schüller, P., Weinzierl, A.: Finding explanations of inconsistency in multi-context systems. In Lin, F., Sattler, U., Truszczynski, M., eds.: KR, AAAI Press (2010)
4. Eiter, T., Fink, M., Weinzierl, A.: Preference-based inconsistency assessment in multi-context systems. In: 12th European Conference on Logics in Artificial Intelligence. (2010) To appear.
5. Przymusinski, T.: Stable semantics for disjunctive programs. New generation computing **9**(3) (1991) 401–424
6. Hunter, A., Konieczny, S.: Measuring inconsistency through minimal inconsistent sets. In Brewka, G., Lang, J., eds.: KR, AAAI Press (2008) 358–366
7. Bikakis, A., Antoniou, G., Hassapis, P.: Alternative strategies for conflict resolution in multi-context systems. In: AIAI. (2009) 31–40
8. Meilicke, C., Stuckenschmidt, H., Tamilin, A.: Repairing ontology mappings. In: AAAI, AAAI Press (2007) 1408–1413
9. Calvanese, D., Giacomo, G.D., Lembo, D., Lenzerini, M., Rosati, R.: Inconsistency tolerance in p2p data integration: An epistemic logic approach. In Bierman, G.M., Koch, C., eds.: DBPL. Volume 3774 of Lecture Notes in Computer Science., Springer (2005) 90–105

Justification Counterpart of Distributed Knowledge Systems

Meghdad Ghari

Department of Mathematical Sciences, Isfahan University of Technology
Isfahan, 84156-83111, Iran
mghari@math.iut.ac.ir

Abstract. In this paper, we will introduce justification counterparts for some distributed knowledge systems. Our justified systems have explicit knowledge operators of the form $[\![t]\!]_i F$ and $[\![t]\!]_{\mathcal{D}} F$, which are interpreted respectively as "t is a justification that agent i accepts for F", and "t is a justification that all agents implicitly accept for F". We then present Kripke style models, called pseudo-Fitting, and prove the completeness theorem. Finally, using labelled sequent calculus of distributed knowledge systems and the fully explanatory property of pseudo-Fitting models, we establish the realization theorem.

1 Introduction

Justification logics (cf. [2]) are a new generation of epistemic logics in which the knowledge operators $\mathcal{K}_i F$ (agent i knows F) are replaced with evidence-based knowledge operators $[\![t]\!]_i F$ (agent i accepts t as an evidence for F), in which t is a justification term. The first justification logic, Logic of Proofs **LP**, was introduced by Artemov in [1] as an one-agent justification counterpart of the epistemic modal logic **S4**. The exact correspondence between **LP** and **S4** is given by the *Realization Theorem*: all occurrences of knowledge operator $\mathcal{K}$ (or $\Box$) in a theorem of **S4** can be replaced by suitable terms in a way that the resulting formula is a theorem of **LP**, and vise versa. To prove the realization theorem, Artemov used a cut-free sequent calculus for **S4** (cf. [1]).

Logic of proofs is a justification logic with a new operator $[\![\cdot]\!]$ for one agent. In [10] Yavorskaya (Sidon) studied two-agent justification logics that have interactions, e.g., evidences of one agent can be verified by the other agent, or evidences of one agent can be converted to evidences of the other agent. Renne also introduced dynamic epistemic logic with justification, a system for multi-agent communication (see e.g. [9]). Later, Bucheli, Kuznets and Studer in [3] suggested an explicit evidence system with common knowledge, an attempt to find justification counterpart of common knowledge systems (although proving the realization theorem for this system is still an open problem). None of the aforementioned papers deal with the notion of distributed knowledge.

In this paper, we study multi-agent evidence-based systems in a distributed environment. Distributed knowledge is the knowledge that is implicitly available in a group, and can be discovered through communication (cf. [4, 8]). We introduce an evidence-based knowledge operator for distributed knowledge $[\![t]\!]_{\mathcal{D}} F$ with the meaning "t is an evidence that all agents implicitly accept for F". To capture this notion, we present axiomatic systems $\mathbf{JK_n^D}$, $\mathbf{JT_n^D}$, $\mathbf{JS4_n^D}$ and $\mathbf{JS5_n^D}$ as well as possible world semantics for these systems. In the present paper, we consider $[\![\cdot]\!]_{\mathcal{D}}$ as an agent, and give pseudo-Fitting models with additional accessibility relation $\mathcal{R}_{\mathcal{D}}$ and evidence function $\mathcal{E}_{\mathcal{D}}$.

Finally, by proving the *Realization Theorem*, we show that our systems are the justification counterparts of the known distributed systems $\mathbf{K_n^D}$, $\mathbf{T_n^D}$, $\mathbf{S4_n^D}$ and $\mathbf{S5_n^D}$. To this end, we employ the labelled sequent calculus of distributed knowledge presented by Hakli and Negri in [7], and the *Fully Explanatory* property of pseudo-Fitting models. Our realization algorithm is essentially the algorithm presented by Artemov in [1] for **S4**, adapted for labelled sequent systems. Nevertheless, our method, unlike the Artemov's one, is not constructive.

2 Distributed Knowledge Systems

In this paper, we fix a set of n agents $G = \{1, 2, \ldots, n\}$. The language of distributed knowledge systems is obtained by adding the modal operators $\mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{D}$ to propositional logic. Hence, if A is a formula then $\mathcal{K}_i A$, for $i = 1, \ldots, n$, and $\mathcal{D}A$ are also formulas. The intended meaning of $\mathcal{K}_i A$ is "agent i knows A", and $\mathcal{D}A$ means "A is distributed knowledge". Next, we recall well known distributed knowledge systems (for more expositions cf. [4, 8]).

Definition 1. *The axioms of* $\mathbf{K_n^D}$ *are (where* $i = 1, \ldots, n$*):*

Taut. *Finite set of axioms for propositional logic,*
K. $\mathcal{K}_i(A \to B) \to (\mathcal{K}_i A \to \mathcal{K}_i B)$,
KD. $\mathcal{D}(A \to B) \to (\mathcal{D}A \to \mathcal{D}B)$,
$\mathbf{K_iD}$**.** $\mathcal{K}_i A \to \mathcal{D}A$.

The rules of inference are:

Modus Ponens: *from* A *and* $A \to B$*, infer* B*,*
Necessitation rule: *from* A *infer* $\mathcal{K}_i A$*.*

If the number of agents $n = 1$*, then we add the additional axiom:*

$\mathcal{D}A \to \mathcal{K}_1 A.$

Extensions of $\mathbf{K_n^D}$ *obtain by adding some axioms as follows:*

- $\mathbf{T_n^D} = \mathbf{K_n^D} + (\mathcal{K}_i A \to A) + (\mathcal{D}A \to A)$,
- $\mathbf{S4_n^D} = \mathbf{T_n^D} + (\mathcal{K}_i A \to \mathcal{K}_i \mathcal{K}_i A) + (\mathcal{D}A \to \mathcal{D}\mathcal{D}A)$,
- $\mathbf{S5_n^D} = \mathbf{S4_n^D} + (\neg\mathcal{K}_i A \to \mathcal{K}_i \neg\mathcal{K}_i A) + (\neg\mathcal{D}A \to \mathcal{D}\neg\mathcal{D}A)$.

In what follows, $\mathbf{L^D}$ is any of the logics defined above. Next we recall Kripke models for the systems $\mathbf{L^D}$.

Definition 2. *A Kripke model* $\mathcal{M}$ *for* $\mathbf{K_n^D}$ *is a tuple* $\mathcal{M} = (\mathcal{W}, \mathcal{R}_1, \ldots, \mathcal{R}_n, \Vdash)$ *where* $\mathcal{W}$ *is a non-empty set of worlds (or states), each* $\mathcal{R}_i$ *is a binary accessibility relation between worlds, and the forcing relation* $\Vdash$ *is a relation between pairs* $(\mathcal{M}, w)$ *and propositional letters, that can be extended to all formulas as follows:*

1. $\Vdash$ *respects classical Boolean connectives,*
2. $(\mathcal{M}, w) \Vdash \mathcal{K}_i A$ *iff for every* $v \in \mathcal{W}$ *with* $w\mathcal{R}_i v$, $(\mathcal{M}, v) \Vdash A$,
3. $(\mathcal{M}, w) \Vdash \mathcal{D}A$ *iff for every* $v \in \mathcal{W}$ *with* $w\mathcal{R}_{\mathcal{D}} v$, $(\mathcal{M}, v) \Vdash A$, *where* $\mathcal{R}_{\mathcal{D}} = \bigcap_{i=1}^{n} \mathcal{R}_i$.

For Kripke models of $\mathbf{T_n^D}$, $\mathbf{S4_n^D}$ *and* $\mathbf{S5_n^D}$ *each* $\mathcal{R}_i$ *should be reflexive, reflexive and transitive and an equivalence relation, respectively.*

Theorem 1. *[4, 8]* $\mathbf{K_n^D}$, $\mathbf{T_n^D}$, $\mathbf{S4_n^D}$ *and* $\mathbf{S5_n^D}$ *are sound and complete with respect to their models.*

Hakli and Negri in [7] presented cut-free labelled sequent calculus for distributed knowledge systems $\mathbf{L^D}$. Consider a Kripke model $\mathcal{M} = (\mathcal{W}, \mathcal{R}_1, \ldots, \mathcal{R}_n, \Vdash)$. First we need a countably infinite set L of labels $w, v, u, \ldots$, that are used as the possible worlds in Kripke models. Then we extend the language of the sequent calculus by forcing atoms (or labelled atoms) $w \Vvdash A$ and relational atoms (or accessibility atoms) $wR_i v$, where w and v are labels and A is a formula in the language of distributed knowledge. These atoms respectively denote the statements $(\mathcal{M}, w) \Vdash A$ and $w\mathcal{R}_i v$ in Kripke models. The axioms and rules for $\mathbf{G3K_n^D}$ are given in Table 1. In the premise of the rules $(R\mathcal{K}_i)$ and $(R\mathcal{D})$, the label v (called an *eigenlabel*) should

Table 1. Labelled sequent calculus for $\mathbf{G3K_n^D}$.

Axioms:

$w \Vvdash P, \Gamma \Rightarrow \Delta, w \Vvdash P$ $\quad$ $w \Vvdash \bot, \Gamma \Rightarrow \Delta$ $\quad$ $wR_i v, \Gamma \Rightarrow \Delta, wR_i v$

where P is a propositional letter.

Rules:

$$\frac{\Gamma \Rightarrow \Delta, w \Vvdash A \quad w \Vvdash B, \Gamma \Rightarrow \Delta}{w \Vvdash A \to B, \Gamma \Rightarrow \Delta}(L \to) \qquad \frac{w \Vvdash A, \Gamma \Rightarrow \Delta, w \Vvdash B}{\Gamma \Rightarrow \Delta, w \Vvdash A \to B}(R \to)$$

$$\frac{v \Vvdash A, w \Vvdash \mathcal{K}_i A, wR_i v, \Gamma \Rightarrow \Delta}{w \Vvdash \mathcal{K}_i A, wR_i v, \Gamma \Rightarrow \Delta}(L\mathcal{K}_i) \qquad \frac{wR_i v, \Gamma \Rightarrow \Delta, v \Vvdash A}{\Gamma \Rightarrow \Delta, w \Vvdash \mathcal{K}_i A}(R\mathcal{K}_i)$$

$$\frac{v \Vvdash A, w \Vvdash \mathcal{D}A, wR_1 v, \ldots, wR_n v, \Gamma \Rightarrow \Delta}{w \Vvdash \mathcal{D}A, wR_1 v, \ldots, wR_n v, \Gamma \Rightarrow \Delta}(L\mathcal{D}) \qquad \frac{wR_1 v, \ldots, wR_n v, \Gamma \Rightarrow \Delta, v \Vvdash A}{\Gamma \Rightarrow \Delta, w \Vvdash \mathcal{D}A}(R\mathcal{D})$$

not appear in the conclusion of the rules. Then we can define a labelled sequent

Table 2. Rules for accessibility relations.

$$\frac{wR_i w, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}(Ref_i) \quad \frac{vR_i w, wR_i v, \Gamma \Rightarrow \Delta}{wR_i v, \Gamma \Rightarrow \Delta}(Sym_i) \quad \frac{wR_i u, wR_i v, vR_i u, \Gamma \Rightarrow \Delta}{wR_i v, vR_i u, \Gamma \Rightarrow \Delta}(Trans_i)$$

calculus for various distributed knowledge systems according to properties of their Kripke models (the rules corresponding to accessibility relations are given in Table 2):

- $\mathbf{G3T_n^D} = \mathbf{G3K_n^D} + (Ref_i)$,
- $\mathbf{G3S4_n^D} = \mathbf{G3K_n^D} + (Ref_i) + (Trans_i)$,
- $\mathbf{G3S5_n^D} = \mathbf{G3K_n^D} + (Ref_i) + (Trans_i) + (Sym_i)$,

where $i \in G$. By $\mathbf{G3L^D}$ we will denote the above labelled sequent systems. Hakli and Negri proved that all the structural rules (weakening, contraction and cut) are admissible in the systems $\mathbf{G3L^D}$ (see [7]).

3 Distributed Justification Logics

In the rest of the paper, we denote by $*$ one of the agents in G or distributed knowledge operator $\mathcal{D}$, i.e. $* \in \{1, \ldots, n, \mathcal{D}\}$. Similar to the language used in [3] and [10], for each $* \in \{1, \ldots, n, \mathcal{D}\}$, we shall define a set of terms as justifications for agent $*$. We start by defining a set of justification variables and constants:

$$Var^* = \{x_1^*, x_2^*, \ldots\} \qquad Cons^* = \{c_1^*, c_2^*, \ldots\}.$$

Now define the set of admissible terms Tm_* (for each $*$) as follows[1]:

– $Var^* \subseteq Tm_*$,
– $Cons^* \subseteq Tm_*$,
– if $s, t \in Tm_*$, then $s +_* t$, $s \cdot_* t$, $!_* t$, $?_* t \in Tm_*$,
– $Tm_i \subseteq Tm_{\mathcal{D}}$, for each $i \in G$.

Note that by the last clause, $Tm_i \subseteq Tm_{\mathcal{D}}$, we do not need to define variables and constants for operator $\mathcal{D}$. Another different alternative is to define only one set of terms Tm that is admissible for all agents as well as for distributed knowledge operator. A slightly different approach used by Renne in [9] to construct a multi-agent update language for justification logic with multi-agent communication and evidence elimination.

Formulas of the justification counterpart of distributed systems are constructed as follows:

$$F := P \,|\, \bot \,|\, F \to F \,|\, [\![t]\!]_* F,$$

where $t \in Tm_*$. The intended meaning of $[\![t]\!]_i F$ is "t is a justification that agent i accepts for F", and $[\![t]\!]_{\mathcal{D}} F$ means "t is a justification that all agents implicitly accept for F". We begin by defining the language and axioms of basic distributed justification logic.

Definition 3. *The language of* $\mathbf{JK_n^D}$ *contains only the operators* $\cdot_*$ *and* $+_*$. *The axioms of* $\mathbf{JK_n^D}$ *are:*

A0. *Finite set of axioms for propositional logic,*
A1. $[\![s]\!]_* A \vee [\![t]\!]_* A \to [\![s +_* t]\!]_* A$,

[1] We define the set of terms Tm_* for the logic $\mathbf{JS5_n^D}$, other distributed justification logics, as will defined in the Definition 3, do not contain operators $!_*$ or $?_*$. Thus, each distributed justification logic requires those clauses in the construction of terms that contains the corresponding operator in its language.

A2. $[\![s]\!]_*(A \to B) \to ([\![t]\!]_* A \to [\![s \cdot_* t]\!]_* B)$,
A3. $[\![t]\!]_i A \to [\![t]\!]_{\mathcal{D}} A$.

The rules of inference are:

R1. *Modus Ponens: from A and $A \to B$, infer B,*
R2. *Iterated Axiom Necessitation rule:* $\vdash [\![c^{i_m}_{j_m}]\!]_{i_m} \ldots [\![c^{i_1}_{j_1}]\!]_{i_1} A$, *where A is an axiom, $c^{i_k}_{j_l}$'s are justification constants and $i_1, \ldots, i_m$ are in G.*

If the number of agents $n = 1$, then we add the additional axiom:

A4. $[\![t]\!]_{\mathcal{D}} A \to [\![t]\!]_1 A$.

The justification system $\mathbf{JT^D_n}$ *is obtained from* $\mathbf{JK^D_n}$ *by adding the following axioms:*

A5. $[\![t]\!]_* A \to A$.

The justification system $\mathbf{JS4^D_n}$ *is obtained from* $\mathbf{JT^D_n}$ *by first extending the language with operators* $!_*$ *and then adding the following axioms:*

A6. $[\![t]\!]_* A \to [\![!_* t]\!]_* [\![t]\!]_* A$.

The justification system $\mathbf{JS5^D_n}$ *is obtained from* $\mathbf{JS4^D_n}$ *by first extending the language with operators* $?_*$ *and then adding the following axioms:*

A7. $\neg [\![t]\!]_* A \to [\![?_* t]\!]_* \neg [\![t]\!]_* A$.

Notice that, in the axioms **A1**, **A2**, **A6** and **A7** all occurrences of $*$ are the same agent or $\mathcal{D}$. By $\mathbf{JL^D}$ we denote logics defined in the Definition 3. Following [10], we define constant specifications as follows:

Definition 4. *A Constant Specification $\mathcal{CS}$ for $\mathbf{JL^D}$ is a set of formulas of the form $[\![c^{i_m}_{j_m}]\!]_{i_m} \ldots [\![c^{i_1}_{j_1}]\!]_{i_1} A$, where A is an axiom of $\mathbf{JL^D}$, $c^{i_l}_{j_l}$'s are justification constants and $i_1, \ldots, i_m$ are in G, and moreover it is downward closed: if $[\![c^{i_m}_{j_m}]\!]_{i_m} \ldots [\![c^{i_1}_{j_1}]\!]_{i_1} A \in \mathcal{CS}$, then $[\![c^{i_{m-1}}_{j_{m-1}}]\!]_{i_{m-1}} \ldots [\![c^{i_1}_{j_1}]\!]_{i_1} A \in \mathcal{CS}$. A constant specification $\mathcal{CS}$ is axiomatically appropriate if for each axiom A there is a constant $c^i \in Tm_i$ such that $[\![c^i]\!]_i A \in \mathcal{CS}$ and also $\mathcal{CS}$ is upward closed: if $[\![c^{i_m}_{j_m}]\!]_{i_m} \ldots [\![c^{i_1}_{j_1}]\!]_{i_1} A \in \mathcal{CS}$, then $[\![c^{i_{m+1}}_{j_{m+1}}]\!]_{i_{m+1}} [\![c^{i_m}_{j_m}]\!]_{i_m} \ldots [\![c^{i_1}_{j_1}]\!]_{i_1} A \in \mathcal{CS}$, for some constant $c^{i_{m+1}}_{j_{m+1}} \in Tm_{m+1}$.*

In the systems that contain axioms **A6**, we can replace the rule **R2** by the following simple one:

R3. *Axiom Necessitation rule:* $\vdash [\![c^i]\!]_i A$, where A is an axiom, c^i is a justification constant and $i \in G$.

In these systems a constant specification $\mathcal{CS}$ is simply a set of formulas of the form $[\![c^i]\!]_i A$, such that c^i is a justification constant, A is an axiom and $i \in G$. In this case for axiomatically appropriate, we do not need the upward closed property.

Let $\mathbf{JL^D}(\mathcal{CS})$ be the fragment of $\mathbf{JL^D}$ where the (Iterated) Axiom Necessitation rule only produces formulas from the given $\mathcal{CS}$. Thus $\mathbf{JL^D}(\emptyset)$ is the fragment of $\mathbf{JL^D}$ without (Iterated) Axiom Necessitation rule. By $\mathbf{JL^D} \vdash F$ we mean $\mathbf{JL^D}(\mathcal{CS}) \vdash F$ for some constant specification $\mathcal{CS}$.

Example 1. One can easily show that $\mathbf{JK_n^D}(\emptyset) \vdash [\![s]\!]_i(A \to B) \wedge [\![t]\!]_j A \to [\![s \cdot_{\mathcal{D}} t]\!]_{\mathcal{D}} B$. This is similar to the fact that $\mathbf{K_n^D} \vdash \mathcal{K}_i(A \to B) \wedge \mathcal{K}_j A \to \mathcal{D}B$.

Example 2. In $\mathbf{JK_n^D}$ the following rule is admissible:

$$\frac{A_1 \wedge \ldots \wedge A_n \to B}{[\![t_1]\!]_1 A_1 \wedge \ldots \wedge [\![t_n]\!]_n A_n \to [\![t]\!]_{\mathcal{D}} B}$$

for some term t in $Tm_{\mathcal{D}}$.

The proof of the following lemmas are similar to those in [10, 3].

Lemma 1 (Lifting Lemma). *For each $* \in \{1, \ldots, n, \mathcal{D}\}$, the following statements are provable:*

1. *If $[\![t_1]\!]_* A_1, \ldots, [\![t_m]\!]_* A_m, B_1, \ldots, B_l \vdash F$ in $\mathbf{JL^D}(\mathcal{CS})$, then*
$$[\![t_1]\!]_* A_1, \ldots, [\![t_m]\!]_* A_m, [\![x_1^*]\!]_* B_1, \ldots, [\![x_l^*]\!]_* B_l \vdash [\![p(\vec{t}, \vec{x})]\!]_* F \qquad (\dagger)$$
in $\mathbf{JL^D}(\mathcal{CS}')$, for some justification variables x_i^* (in Var^*), term $p(\vec{t}, \vec{x})$ in Tm_* and $\mathcal{CS}' \supseteq \mathcal{CS}$ (all $*$'s in $(\dagger)$ stand for the same agent).
2. *In part (1), if $\mathcal{CS}$ is axiomatically appropriate, then $(\dagger)$ is provable in $\mathbf{JL^D}(\mathcal{CS})$.*

Lemma 2 (Internalization Lemma). *For each $* \in \{1, \ldots, n, \mathcal{D}\}$, the following statements are provable:*

1. *If $\mathbf{JL^D}(\mathcal{CS}) \vdash F$, then $\mathbf{JL^D}(\mathcal{CS}') \vdash [\![p]\!]_* F$, for some term p in Tm_* and $\mathcal{CS}' \supseteq \mathcal{CS}$.*
2. *Suppose $\mathcal{CS}$ is axiomatically appropriate. If $\mathbf{JL^D}(\mathcal{CS}) \vdash F$, then $\mathbf{JL^D}(\mathcal{CS}) \vdash [\![p]\!]_* F$, for some term p in Tm_*.*

It is worth noting that the term p, that is constructed in the proof of lifting or internalization lemma, is different for each agent $*$.

4 Semantics

In this section, we consider $[\![\cdot]\!]_{\mathcal{D}}$ as an agent, rather than as explicit distributed knowledge, and give pseudo-Fitting models for all systems $\mathbf{JL^D}$. Fitting models first introduced by Fitting in [5] for **LP**.

Definition 5. *A pseudo-Fitting model $\mathcal{M}$ for $\mathbf{JK_n^D}$ is a tuple*

$$\mathcal{M} = (\mathcal{W}, \mathcal{R}_1, \ldots, \mathcal{R}_n, \mathcal{R}_{\mathcal{D}}, \mathcal{E}_1, \ldots, \mathcal{E}_n, \mathcal{E}_{\mathcal{D}}, \Vdash_p)$$

(or $\mathcal{M} = (\mathcal{W}, \mathcal{R}_*, \mathcal{E}_*, \Vdash_p)$ for short) where $(\mathcal{W}, \mathcal{R}_1, \ldots, \mathcal{R}_n, \Vdash_p)$ is a Kripke model, in which $\mathcal{R}_{\mathcal{D}}$ is also a binary accessibility relation between worlds. Admissible evidence functions $\mathcal{E}_*$ are mappings from the set of terms and formulas to the set of all worlds, i.e., $\mathcal{E}_*(t, A) \subseteq \mathcal{W}$, for any justification term t in Tm_* and formula A, and satisfying the following conditions. For all justification terms s and t and for all formulas A and B:

$\mathcal{E}1$**.** $\mathcal{E}_*(s, A) \cup \mathcal{E}_*(t, A) \subseteq \mathcal{E}_*(s +_* t, A)$,

$\mathcal{E}2$. $\mathcal{E}_*(s, A \to B) \cap \mathcal{E}_*(t, A) \subseteq \mathcal{E}_*(s \cdot_* t, B)$,
$\mathcal{E}3$. $\mathcal{E}_i(t, A) \subseteq \mathcal{E}_{\mathcal{D}}(t, A)$, *for each* $i \in G$.

If $n = 1$, evidence functions should also satisfy:

$\mathcal{E}4$. $\mathcal{E}_{\mathcal{D}}(t, A) \subseteq \mathcal{E}_1(t, A)$.

The forcing relation $\Vdash_p$ is a relation between pairs $(\mathcal{M}, w)$ and propositional letters, that can be extended to all formulas as follows:

1. *$\Vdash_p$ respects classical Boolean connectives,*
2. *$(\mathcal{M}, w) \Vdash_p [\![t]\!]_* A$ iff $w \in \mathcal{E}_*(t, A)$ and for every $v \in \mathcal{W}$ with $w\mathcal{R}_*v$, $(\mathcal{M}, v) \Vdash_p A$.*

We say that A is true in a model $\mathcal{M}$ ($\mathcal{M} \Vdash_p A$) if it is true at each world of the model. For a set X of formulas, $\mathcal{M} \Vdash_p X$ if $\mathcal{M} \Vdash_p F$ for all formulas F in X. Given a constant specification $\mathcal{CS}$, a model $\mathcal{M}$ respects $\mathcal{CS}$ (or meets $\mathcal{CS}$) if $\mathcal{M} \Vdash_p \mathcal{CS}$.

Pseudo–Fitting models for the other distributed justification logics have more restrictions on accessibility relations and evidence functions. For $\mathbf{JT}^{\mathbf{D}}_{\mathbf{n}}$ each $\mathcal{R}_*$ is reflexive. For $\mathbf{JS4}^{\mathbf{D}}_{\mathbf{n}}$ each $\mathcal{R}_*$ is reflexive and transitive and evidence functions should satisfy:

$\mathcal{E}5$. If $w \in \mathcal{E}_*(t, A)$ and $w\mathcal{R}_*v$, then $v \in \mathcal{E}_*(t, A)$,
$\mathcal{E}6$. $\mathcal{E}_*(t, A) \subseteq \mathcal{E}_*(!_* t, [\![t]\!]_* A)$,

For $\mathbf{JS5}^{\mathbf{D}}_{\mathbf{n}}$ each $\mathcal{R}_*$ is an equivalence relation and evidence functions should satisfy:

$\mathcal{E}7$. If $[\mathcal{E}_*(t, A)]^c \subseteq \mathcal{E}_*(?_* t, \neg[\![t]\!]_* A)$, where the superscript operation "c" on sets is the complement relative to the set of worlds $\mathcal{W}$,
$\mathcal{E}8$. If $w \in \mathcal{E}_*(t, A)$, then $(\mathcal{M}, w) \Vdash_p [\![t]\!]_* A$.

Next, we prove the completeness theorem for $\mathbf{JL}^{\mathbf{D}}$. Since the proof is similar to the completeness theorem of justification logics in [2, 5], we omit the details.

Theorem 2 (Completeness). *For a given constant specification $\mathcal{CS}$, distributed justification logics $\mathbf{JL}^{\mathbf{D}}(\mathcal{CS})$ are sound and complete with respect to their pseudo-Fitting models that respecting $\mathcal{CS}$.*

Proof. Soundness is straightforward, as usual, so we focus on completeness. For completeness we first construct a canonical model $\mathcal{M} = (\mathcal{W}, \mathcal{R}_*, \mathcal{E}_*, \Vdash_p)$ as follows:

- $\mathcal{W}$ is the set of all maximally consistent sets in $\mathbf{JL}^{\mathbf{D}}(\mathcal{CS})$,
- $\Gamma\mathcal{R}_*\Delta$ iff if $[\![t]\!]_* F \in \Gamma$ then $F \in \Delta$,
- $\mathcal{E}_*(t, F) = \{\Gamma \in \mathcal{W} \mid [\![t]\!]_* F \in \Gamma\}$
- for each propositional letter P: $(\mathcal{M}, \Gamma) \Vdash_p P$ iff $P \in \Gamma$.

Specially, for each distributed justification logic $\mathbf{JL}^{\mathbf{D}}$, the evidence function $\mathcal{E}_*$ in the canonical model $\mathcal{M}$ satisfies the proper $\mathcal{E}1 - \mathcal{E}8$ properties in the definition of pseudo–Fitting model. We only show the new property $\mathcal{E}3$ ($\mathcal{E}4$ is similar). Let $\Gamma \in \mathcal{E}_i(t, A)$. Then $[\![t]\!]_i A \in \Gamma$. Since $[\![t]\!]_i A \to [\![t]\!]_{\mathcal{D}} A \in \Gamma$, we have $[\![t]\!]_{\mathcal{D}} A \in \Gamma$, and therefore $\Gamma \in \mathcal{E}_{\mathcal{D}}(t, A)$.

One can prove the Truth Lemma: for all formulas F we have

$$F \in \Gamma \quad \text{iff} \quad (\mathcal{M}, \Gamma) \Vdash_p F.$$

The proof is by induction on the complexity of F and is similar to that for justification logics in [2]. Now suppose $\mathbf{JL^D}(\mathcal{CS}) \nvdash A$, then $\{\neg A\}$ is a $\mathbf{JL^D}(\mathcal{CS})$-consistent set. Extend it to a maximal consistent set Γ by standard Lindenbaum construction, then by truth lemma we have $(\mathcal{M}, \Gamma) \nVdash_p A$.

Note that in the canonical model $\mathcal{M}$ we have $\mathcal{R}_\mathcal{D} \subseteq \cap_{i=0}^n \mathcal{R}_i$, and for $n = 1$, $\mathcal{R}_\mathcal{D} = \mathcal{R}_1$. Moreover, $\cup_{i=0}^n \mathcal{E}_i(t, A) \subseteq \mathcal{E}_\mathcal{D}(t, A)$, for every term t and formula A. □

One of the key tools we will employ in the proof of realization theorem is the fully explanatory property of models, that first proved by Fitting in [5] for the logic of proofs. The definition of the fully explanatory property is the same as that for justification logics (see [2, 5]), except that in order to prove the realization theorem for $\mathbf{JL^D}$ we add a new item.

Definition 6. *A* $\mathbf{JL^D}$*-model* $\mathcal{M}$ *is a strong model if it has the* Fully Explanatory property*:*

1. *if for every* v *such that* $w\mathcal{R}_*v$ *we have* $(\mathcal{M}, v) \Vdash_p A$*, then for some term* $t \in Tm_*$ *we have* $(\mathcal{M}, w) \Vdash_p [\![t]\!]_* A$*, and*
2. *if for every* v *such that* $w\mathcal{R}_1 v, \ldots, w\mathcal{R}_n v$ *we have* $(\mathcal{M}, v) \Vdash_p A$*, then for some term* $t \in Tm_\mathcal{D}$ *we have* $(\mathcal{M}, w) \Vdash_p [\![t]\!]_\mathcal{D} A$*.*

In fact, for our purpose we only need to state the clause 1 of the fully explanatory property for agents i in G.

Theorem 3 (Strong Completeness). *For any axiomatically appropriate constant specification* $\mathcal{CS}$*,* $\mathbf{JL^D}(\mathcal{CS})$ *is sound and complete with respect to their strong models that respecting* $\mathcal{CS}$*.*

Proof. It suffices to prove that, for any axiomatically appropriate constant specification $\mathcal{CS}$, the canonical model of $\mathbf{JL^D}(\mathcal{CS})$ satisfies the fully explanatory property. We only prove the clause 2 of Definition 6 (the proof of clause 1 is similar to the case of justification logics, cf. [2, 5]). We prove the theorem by contraposition. Let $\mathcal{M} = (\mathcal{W}, \mathcal{R}_*, \mathcal{E}_*, \Vdash_p)$ be the canonical model of $\mathbf{JL^D}(\mathcal{CS})$, and $\Gamma \in \mathcal{W}$. Suppose $[\![t]\!]_\mathcal{D} A \notin \Gamma$ (or equivalently $(\mathcal{M}, \Gamma) \nVdash_p [\![t]\!]_\mathcal{D} A$) for each $t \in Tm_\mathcal{D}$. We will show that there is $\Delta \in \mathcal{W}$ with $\Gamma R_\mathcal{D} \Delta$ such that $(\mathcal{M}, \Delta) \nVdash_p A$. Consider the set

$$S = \{F \,|\, [\![t]\!]_\mathcal{D} F \in \Gamma, \text{for some}\, t \in Tm_\mathcal{D}\} \cup \{\neg A\}.$$

We prove that S is consistent. Otherwise, for some $[\![\, t_1 \,]\!]_\mathcal{D} X_1, \ldots, [\![\, t_m \,]\!]_\mathcal{D} X_m$ in Γ the following formula is provable in $\mathbf{JL^D}(\mathcal{CS})$

$$X_1 \to (X_2 \to \cdots \to (X_m \to A) \cdots).$$

Since the constant specification $\mathcal{CS}$ is axiomatically appropriate, by Lemma 2, there is a term s in $Tm_\mathcal{D}$ such that $\mathbf{JL^D}(\mathcal{CS})$ proves $[\![\, s \,]\!]_\mathcal{D}(X_1 \to (X_2 \to \cdots \to (X_m \to A) \cdots))$. By axiom **A2**, and the fact that $[\![\, t_i \,]\!]_\mathcal{D} X_i \in \Gamma$ for each $i = 1, \cdots, m$, we conclude that

$$[\![\, t_1 \,]\!]_\mathcal{D} X_1 \to ([\![\, t_2 \,]\!]_\mathcal{D} X_2 \to \cdots \to ([\![\, t_m \,]\!]_\mathcal{D} X_m \to [\![\, t \,]\!]_\mathcal{D} A) \cdots)$$

is provable, for $t = s \cdot_\mathcal{D} t_1 \cdot_\mathcal{D} \cdots \cdot_\mathcal{D} t_m$. Hence, $(\mathcal{M}, \Gamma) \Vdash_p [\![t]\!]_\mathcal{D} A$, which is a contradiction. Thus S is a consistent set. Now extend S to the maximal consistent set Δ. Therefore, by the truth lemma $(\mathcal{M}, \Delta) \nVdash_p A$. On the other hand, it is obvious that $\Gamma \mathcal{R}_\mathcal{D} \Delta$, and since $\mathcal{R}_\mathcal{D} \subseteq \cap_{i=0}^n \mathcal{R}_i$, for each $i \in G$ we have $\Gamma \mathcal{R}_i \Delta$. □

Note that the proof of strong completeness theorem is not constructive, in the sense that, we did not construct a term $t \in Tm_{\mathcal{D}}$ such that $(\mathcal{M}, \Gamma) \Vdash_p [\![t]\!]_{\mathcal{D}} A$.

5 Realization Theorem

We will use the strong completeness theorem of $\mathbf{JL^D}$ to present a uniform realization theorem for all distributed knowledge systems $\mathbf{JL^D}$. The author used the same method in [6] to give a uniform realization theorem for the most common modal logics. The method is similar to Artemov's realization algorithm used to show $\mathbf{LP}^\circ = \mathbf{S4}$ in [1]. However, there are some differences. Artemov used proof trees in a Gentzen sequent calculus of $\mathbf{S4}$, but we employ proof trees in labelled Gentzen sequent calculus in $\mathbf{L^D}$. Moreover, to realize instances of the rule $(R\Box)$ in the proof tree, Artemov used the lifting lemma, but to realize instances of the rules $(R\mathcal{K}_i)$ and $(R\mathcal{D})$ in a proof tree in $\mathbf{G3L^D}$ we employ the fully explanatory property of pseudo-Fitting models. Thus, unlike the Artemov's realization algorithm, our algorithm is not constructive. Before stating the realization theorem we need some definitions.

Definition 7. *For a* $\mathbf{JL^D}$*-formula* F*, the forgetful projection of* F*, denoted by* F°*, is defined inductively as follows:*

1. *For propositional letter* P*,* $P^\circ = P$*, and* $\bot^\circ = \bot$*,*
2. $(A \to B)^\circ = A^\circ \to B^\circ$*,*
3. $([\![t]\!]_i A)^\circ = \mathcal{K}_i A^\circ$*,*
4. $([\![t]\!]_{\mathcal{D}} A)^\circ = \mathcal{D} A^\circ$*.*

For a set X *of justification formulas we let* $X^\circ = \{F^\circ \,|\, F \in X\}$*.*

In the rest of this section we will prove the following results:

$$\begin{array}{ll} \mathbf{JK_n^D}^\circ = \mathbf{K_n^D}, & \mathbf{JT_n^D}^\circ = \mathbf{T_n^D}, \\ \mathbf{JS4_n^D}^\circ = \mathbf{S4_n^D}, & \mathbf{JS5_n^D}^\circ = \mathbf{S5_n^D}. \end{array} \tag{1}$$

Definition 8. *Let* A *be a formula in the language of* $\mathbf{L^D}$*. A realization of the formula* A *is a* $\mathbf{JL^D}$*-formula* A^r *such that* $(A^r)^\circ = A$*.*

More precisely, a realization A^r is obtained by substituting each modality $\mathcal{K}_i$ in A for a term in Tm_i, and each modality $\mathcal{D}$ in A for a term in $Tm_{\mathcal{D}}$. A realization is called *normal* if all negative occurrences of modalities are replaced by distinct variables.

Definition 9. *let* $\mathcal{M} = (\mathcal{W}, \mathcal{R}_*, \mathcal{E}_*, \Vdash_p)$ *be a pseudo-Fitting model for* $\mathbf{JL^D}$ *and* L *be the set of labels used in derivations. An interpretation* $[\cdot]$ *based on model* $\mathcal{M}$ *is a total function* $[\cdot] : L \to \mathcal{W}$*. Then we will define when an interpretation* $[\cdot]$ *validates a formula:*

- $[\cdot]$ *validates the labelled formula* $w \Vvdash A$*, if* $(\mathcal{M}, [w]) \Vdash_p A$*,*
- $[\cdot]$ *validates the relational atom* $wR_i v$*, if* $[w]\mathcal{R}_i[v]$*.*

Now consider a model $\mathcal{M}$ for $\mathbf{JL^D}$. A sequent $\Gamma \Rightarrow \Delta$ is valid for an interpretation $[\cdot]$ based on model $\mathcal{M}$, if whenever $[\cdot]$ validates all the formulas in Γ then it validates at least one formula in Δ. A sequent is valid in a model $\mathcal{M}$ if it is valid for every interpretation based on $\mathcal{M}$. A sequent is valid in $\mathbf{JL^D}$ (or it is $\mathbf{JL^D}$-valid), if it is valid in every $\mathbf{JL^D}$-model $\mathcal{M}$.

The following lemma is helpful in the rest of the paper. To distinguish between labels in the syntax of our sequent calculus and possible worlds in models, we will denote the possible worlds by letters: $k, l, m, \ldots$.

Lemma 3. *The $\mathbf{JL^D}$-formula A is valid in model $\mathcal{M}$ if and only if the sequent $\Rightarrow w \Vdash A$ is valid.*

Proof. ($\Rightarrow$) Suppose A is valid in the model $\mathcal{M} = (\mathcal{W}, \mathcal{R}_*, \mathcal{E}_*, \Vdash_p)$. Then for every interpretation $[\cdot]$ based on $\mathcal{M}$, and every label $w \in L$ we have $[w] \in \mathcal{W}$, and therefore $(\mathcal{M}, [w]) \Vdash_p A$. Thus, the interpretation $[\cdot]$ validates the sequent $\Rightarrow w \Vdash A$. Since the interpretation $[\cdot]$ was arbitrary, the sequent $\Rightarrow w \Vdash A$ is valid in $\mathcal{M}$.
($\Leftarrow$) Suppose the sequent $\Rightarrow w \Vdash A$ is valid in $\mathcal{M}$, i.e. for every interpretation $[\cdot]$ based on $\mathcal{M}$ we have $(\mathcal{M}, [w]) \Vdash_p A$. For an arbitrary world $k \in \mathcal{W}$, define the interpretation $[\cdot]'$ on $\mathcal{M}$ such that $[w]' = k$. Hence, $(\mathcal{M}, k) \Vdash_p A$, and A is valid in $\mathcal{M}$ as well. □

In the proof of realization theorem, we need to use only strong models of $\mathbf{JL^D}$ that respecting an axiomatically appropriate constant specification. For convenient consider the *Total Constant Specification*:

$\mathcal{TCS} = \{[\![c^{i_m}_{j_m}]\!]_{i_m} \ldots [\![c^{i_1}_{j_1}]\!]_{i_1} A \mid A$ is any axiom of $\mathbf{JL^D}$ and c_{j_l}'s are constants$\}$.

Clearly, $\mathcal{TCS}$ is axiomatically appropriate. In the following theorem, $\mathbf{L^D}$ and $\mathbf{JL^D}$ are respectively modal and justification logics that are described by equations (1).

Theorem 4 (Realization Theorem). $\mathbf{JL^D}^\circ = \mathbf{L^D}$

Proof. By induction on the derivation of formula F in $\mathbf{JL^D}$, one can easily prove that: if $\mathbf{JL^D} \vdash F$, then $\mathbf{L^D} \vdash F^o$, and therefore $\mathbf{JL^D}^\circ \subseteq \mathbf{L^D}$. Next suppose $\mathbf{L^D} \vdash F$. Then there exists a cut-free derivation of the sequent $\Rightarrow w \Vdash F$ in $\mathbf{G3L^D}$, for some label w. We will describe a procedure in which we can find a sequent $\Rightarrow w \Vdash F^r$ that is $\mathbf{JL^D}$-valid with respect to strong models that respecting $\mathcal{TCS}$, and therefore, by Lemma 3 and Theorem 3, we have $\mathbf{JL^D}(\mathcal{TCS}) \vdash F^r$.

Let $\Box \in \{\mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{D}\}$. First we construct disjoint family of $\Box$'s, with the usual definition of related $\Box$'s in a derivation (i.e. occurrences of $\Box$ in the related formulas in rules are related). Negative and positive occurrences of $\Box$ are also defined as usual. So in a cut-free derivation, all the occurrences of $\Box$ in a family has the same polarity. A family is called *essential* if it contains at least one instance of the $\Box$ introduced by the rules $(R\Box)$, i.e. by the rules $(R\mathcal{K}_i)$ or $(R\mathcal{D})$. Then we replace all occurrences of $\Box$ in each family by distinct fresh justification variables, such a replacement is called a *pre-realization*, as follows:

$$\mathcal{K}_i A \rightsquigarrow [\![x]\!]_i A$$

$\mathcal{D}A \rightsquigarrow [\![x]\!]_{\mathcal{D}}A$

where in each replacement x is a new variable. Now we have a tree of sequents in the language of $\mathbf{JL^D}$. Beginning from sequent axioms, we inductively prove that each sequent in the tree is $\mathbf{JL^D}$-valid with respect to strong models that respecting $\mathcal{TCS}$. The important point here is the correspondence between the rules for accessibility relations in $\mathbf{G3L^D}$, and the properties that accessibility relations in the models of $\mathbf{JL^D}$ have. The validity of sequent axioms are obvious. We shall verify the validity of rules by induction. We detail the proof in some cases. For example, consider an instance of the rule $(L\mathcal{K}_i)$ for $i \in G$. After pre-realizing the modalities $\mathcal{K}_i$ by a justification variable in Tm_i, say x, the rule looks like:

$$\frac{v \Vdash A, w \Vdash [\![x]\!]_i A, wR_i v, \Gamma \Rightarrow \Delta}{w \Vdash [\![x]\!]_i A, wR_i v, \Gamma \Rightarrow \Delta}(L\mathcal{K}_i)$$

Suppose $\mathcal{M}$ is a strong $\mathbf{JL^D}$-model respecting $\mathcal{TCS}$, and $[\cdot]$ is an arbitrary interpretation based on $\mathcal{M}$ that validates all formulas in Γ, $(\mathcal{M},[w]) \Vdash_p [\![x]\!]_i A$ and $[w]\mathcal{R}_i[v]$. We have to show that $[\cdot]$ validates at least one formula in Δ. By the induction hypothesis, the premise $v \Vdash A, w \Vdash [\![x]\!]_i A, wR_i v, \Gamma \Rightarrow \Delta$ is valid in every strong $\mathbf{JL^D}$-model respecting $\mathcal{TCS}$. From $(\mathcal{M},[w]) \Vdash_p [\![x]\!]_i A$ and $[w]\mathcal{R}_i[v]$, by the Definition 5, we conclude that $(\mathcal{M},[v]) \Vdash_p A$. Thus, $[\cdot]$ validates at least one formula in Δ as desire.

Other logical rules of $\mathbf{G3K_n^D}$ are treated similarly, except the rules $(R\Box)$. For instance, assume an instance of the rule $(R\mathcal{D})$ after replacing its essential $\mathcal{D}$ by variable x:

$$\frac{wR_1v, \ldots, wR_nv, \Gamma \Rightarrow \Delta, v \Vdash A}{\Gamma \Rightarrow \Delta, w \Vdash [\![x]\!]_{\mathcal{D}} A}$$

Suppose $\mathcal{M}$ is a strong $\mathbf{JL^D}$-model respecting $\mathcal{TCS}$, and $[\cdot]$ is an arbitrary interpretation based on $\mathcal{M}$ that validates all formulas in Γ. By the induction hypothesis, the premise $wR_1v, \ldots, wR_nv, \Gamma \Rightarrow \Delta, v \Vdash A$ is valid in every strong $\mathbf{JL^D}$-model respecting $\mathcal{TCS}$. Let k be an arbitrary element of $\mathcal{W}$ such that $[w]\mathcal{R}_1k, \ldots, [w]\mathcal{R}_nk$ and $[\cdot]'$ be the interpretation identical to $[\cdot]$ except possibly on v, where we put $[v]' = k$. Clearly $[\cdot]'$ validates all the formulas in the antecedent of the premise, so it validates a formula in Δ or it validates $v \Vdash A$, i.e. $(\mathcal{M},k) \Vdash_p A$. In the former case, since v is not in Δ, $[\cdot]$ validates a formula in Δ. In the latter, since k is arbitrary and by assumption $\mathcal{M}$ is fully explanatory, by clause 2 of Definition 6, there is a term t in $Tm_{\mathcal{D}}$ such that $(\mathcal{M},[w]) \Vdash_p [\![t]\!]_{\mathcal{D}} A$. Now we replace all the variables x in the tree by the term t.

Lastly, let us consider the rules for accessibility relation, for instance the rule (Ref_i) in $\mathbf{G3T_n^D}$ (and its extensions $\mathbf{G3S4_n^D}$ and $\mathbf{G3S5_n^D}$):

$$\frac{wR_iw, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}(Ref_i)$$

Suppose $\mathcal{M}$ is a strong $\mathbf{JT_n^D}$-model respecting $\mathcal{TCS}$, and $[\cdot]$ is an arbitrary interpretation based on $\mathcal{M}$ that validates all formulas in Γ. Thus the accessibility relation $\mathcal{R}_i$ in $\mathcal{M}$ is reflexive, and therefore for every $w \in L$ we have $[w]\mathcal{R}_i[w]$. By the induction hypothesis, the premise $wR_iw, \Gamma \Rightarrow \Delta$ is valid in every strong $\mathbf{JT_n^D}$-model respecting $\mathcal{TCS}$. Hence, $[\cdot]$ validates at least one formula in Δ as desire.

Eventually, we get a $\mathbf{JL^D}$-valid (with respect to strong models that respecting $\mathcal{TCS}$) sequent $\Rightarrow w \Vdash F^r$, in which F^r is a realization of the formula F. Thus, by Lemma 3, the formula F^r is $\mathbf{JL^D}$-valid with respect to strong models respecting $\mathcal{TCS}$, and therefore by strong completeness theorem we have, $\mathbf{JL^D}(\mathcal{TCS}) \vdash F^r$. The realization built in this proof is normal. $\Box$

6 Conclusions

This paper is the first study of the justified distributed knowledge systems. The advantage of this study is to incorporate the notion of evidence (or justification) into the distributed knowledge systems. For future work, it is natural to combine the justified distributed knowledge logic $\mathbf{JS4_n^D}$ with the explicit evidence system with common knowledge $\mathbf{LP_n^C}$ introduced in [3]. The method of proving realization theorem used in this paper, might be helpful to find justification counterpart of other notions of group knowledge, as well as common knowledge.

There remains also some questions: Is there Fitting models (pseudo-Fitting models which do not contain accessibility relation $\mathcal{R_D}$) for $\mathbf{JL^D}$? Are $\mathbf{JL^D}$ conservative over multi-agent justification systems $\mathbf{JL_n}$ (the systems $\mathbf{JL^D}$ without distributed knowledge operator)? Are there cut-free tableau or Gentzen systems for $\mathbf{JL^D}$?

Acknowledgements
I would like to thank the referees for their useful comments and suggestions.

References

1. Artemov, S.: Explicit provability and constructive semantics, Bulletin of Symbolic Logic, 7(1):1–36 (2001).
2. Artemov, S.: The logic of justification. The Review of Symbolic Logic 1(4), 477–513 (2008)
3. Bucheli, S., Kuznets, R., Studer, T.: Explicit Evidence Systems with Common Knowledge. E-print 1005.0484, arXiv.org, May 2010.
4. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning about knowledge. MIT Press (1995).
5. Fitting, M.: The logic of proofs, semantically, Annals of Pure and Applied Logic, 132(1)(2005) 1–25.
6. Ghari, M.: Labelled sequent calculus for justification logics, Manuscript 2010.
7. Hakli, R., Negri, S.: Proof theory for distributed knowledge. In Computational Logic in Multi-Agent Systems. 8th International Conference (CLIMA-VIII), LNCS 5056, pages 100-116. Springer-Verlag (2007)
8. Meyer, J.-J.Ch., van der Hoek, W.: Epistemic logic for AI and computer science. Cambridge University Press (1995).
9. Renne, B.: Evidence elimination in multi-agent justification logic. In Aviad Heifetz, editor, Theoretical Aspects of Rationality and Knowledge, Proceedings of the Twelfth Conference (TARK 2009), pages 227-236, Stanford University, California, July 68, 2009.
10. Yavorskaya (Sidon), T.: Interacting explicit evidence systems. Theory of Comput. Syst. 43, 272–293 (2008)

Epistemic Term-Modal Logic

Rasmus K. Rendsvig

Roskilde University

Abstract. In the present paper syntax, constant domain semantics with non-rigid constants and an axiom system for n agent systems with identity for a term-modal logic is presented. Examples of expressibility will be given, and venues of further research addressed.

In his 1637 *Discourse on Method*, French philosopher René Descartes famously wrote *Cogito ergo sum*: I think, therefore I am. Without dwelling in details of Descartes' writings, one may see the proposition as loosely based on existential instantiation – thinking is a property, and if something possess the property, it must, too, exist. But though a classic epistemic and existential proposition, *Cogito ergo sum* is not expressible in classic, first-order epistemic logic.

In fact, even in full-fledged second-order epistemic modal logic, there are certain such elements of first-order reasoning that cannot be expressed. These regard the reasoning agents, their role as knowers and their existence. For though our agents are able to reason both about objects and their properties and their own and other agents information about such matters, they are not able to reason about neither themselves nor other agents as objects. Had Descartes used classic first-order epistemic logic to formalize his arguments, he could never have concluded his own existence.

The reason for this is two-fold. From a semantic point of view, the agents of epistemic logic do not, strictly speaking, exist in the most basic sense, namely in the domains of the models for our various first-order modal logics. Even if we stipulate a subset of the domain to be agents and add agent predicates, this will be insufficient due to syntactic limitations: the property of being "a knower" is expressed by structurally crude modal operators. The agent-denoting indices of such operators are not terms of the first-order language, and hence these do not work as such either.

To exemplify, if we assume $P(a)$ true at some state in a model, it would not be a fair reading of $K_aP(a)$ that agent a knows of himself/agent a that he is P, for the two occurrences of 'a' is neither syntactically nor semantically connected. In fact, the formula blatantly misuses notation. This is the case as the set of operators are defined using an index set *distinct* from the terms of the first-order language – often $I = \{1, 2, ..., n\}$. Hence, we should write, for example, $K_1P(a)$ clearly marking the lack of connection between the knower, 1, and the know-of object, a.

As a result, there are certain aspects of regular first-order reasoning, like that of Descartes, that we are still not able to express in first-order modal logic.

To give an example, imagine student a having to obtain information φ regarding who is to speak after lunch at some summer school. Let us further assume that all members of the Program Committee knows φ and that this is known by a, i.e. $K_a \forall x\, (PC\,(x) \to K_x \varphi)$. Thus a could conclude that b should know φ if b is known to be an organizer. Yet, the universal implication known by a cannot be expressed in classic first-order epistemic logic – it is simply not well-formed.[1]

The lack of expressibility is due to that fact that the knowledge/belief/etc. operators "only" serve as operators – they do not, too, serve as predicates. And as predicates are what expresses properties in first-order modal logic, neither knowledge, belief or other intentional attitudes can be formalized as properties.

In the present paper syntax, constant domain semantics with non-rigid constants and an axiom system for n agent systems with identity that allow for this kind of reasoning will be presented. In order to capture the relevant aspect of agency, namely existence, agents are added explicitly to the domain and are each correlated with an accessibility relation and operators. The operators are indexed by terms of the first-order language in order to ensure that they operate as both operators and predicates. Examples of expressibility will be given and venues of further research addressed. The semantics proposed have been constructed so as to allow for the construction of *normal* term-modal logics, and a variant of the *canonical model theorem* has been proven in [10]. The term 'term-modal logic' was coined in [11], where completeness is shown for sequent calculi and tableaux systems for K, D, T, K4, D4 and S4 for semantics with monotone domains, rigid constants and function symbols and without identity. Earlier, in [9] a term-modal logic with constant domains and non-rigid constants is formulated for belief and soundness shown for KD45. Truth-value gaps are used to render $B_t\varphi$ false for non-agent denoting terms, resulting in non-standard truth-conditions for the operators, in turn resulting in the lack of validity of Dual and Knowledge Generalization. In [8] a first-order dynamic term-modal logic is constructed. Non-rigid constants are used along with a constant domain consisting only of agents, and quantification is done by wildcard assignment. In[6], Hintikka does at late passage use the operators as predicates and quantify over the indices, and the mention of the idea can be found as early as [12].

Changes to syntax

The changes to the syntax of ordinary first-order epistemic logic in order to gain the expressibility required to quantify over agents and denote these with constants are not difficult. Define a language $\mathcal{L}^{\mathbf{n}}$ for a n agent logic a countably infinite set of variables VAR, a set of constants, CON, and a set of relation symbols, REL, where both CON and REL are (possibly empty) countable.[2]

[1] The same kind of reasoning could be forced in propositional epistemic logic with an 'Everybody in G knows that' operator, E_G, as $E_G\varphi \to K_i\varphi$ is valid for all $i \in G$ – but the information whether i is in G or not is not information available to the agents as it is a meta-logical condition.

[2] The notation REL_n will be used to denote the subset of REL consisting of n-ary relation symbols.

The set TER of terms of $\mathcal{L}^{\mathbf{n}}$ are $VAR \cup CON$. The language is relativized to n agents by requiring the existence of a subset $AGT \subseteq VAR$ where $|AGT| \geq n$. The well-formed formulas of $\mathcal{L}^{\mathbf{n}}$ (henceforth formulas) are given by

$$\varphi ::= R^n (t_1, ..., t_n) \mid t_1 = t_2 \mid \neg\varphi \mid \varphi \wedge \psi \mid \forall x\varphi \mid K_t\varphi.$$

where $R^n \in REL_n$, $t, t_1, ..., t_n \in TER$ and $x \in VAR$.
The remaining logical connectives, the existential quantifier and the dual operator as well as definitions of free and bound variables and sentences are all as usual.

This use of the modal operators allow for them to serve both a role as operators, but also a role as classic predicates, hence providing us with the wished for added expressibility. To exemplify, where φ is a formula, both $\exists x K_x\varphi$ and $K_a\varphi \wedge (a = b) \rightarrow K_b\varphi$ are well-formed formulas proper.

Index set and classes of operators. In regular modal logics, the set of modal operators is, so to speak, under control. That is, it is simply defined so that for each $i \in I$ for some finite set of agents I, K_i is an operator. Given the above syntax the set of operators depend on the set of terms. Once we define the semantics we will only have a finite set of accessibility relations, and each operator is then correlated with a relation via the elements of the domain, thus creating classes of semantically equivalent operators.

There are two reasons to introduce an index set in order to partition the operators, even though it is not necessary to define the set of well-formed formulas. The first pertains to the adding of additional axioms: if there is no partition of the operators, there is not way to add axioms for a specific subset, i.e. a class of operators. We do want to be able to add class-specific axioms, though, as each class will contain exactly the operators for one specific agent. Hence adding axioms for a specific class will allow differing epistemic strength of the agents.

Secondly, a partition of the operators is very handy when proving completeness of axiom systems by canonical models, see [10]. The reason is that the classes of semantically equivalent operators are induced by the domain of the semantic structures, as defined below. But as the domain of the canonical models is normally defined through the accessibility relations (which is usually defined through inclusion of operators in maximal consistent sets), we need to "cut the loop", making sure we can define these classes by other than semantic means. To this end we introduce, for a language $\mathcal{L}^{\mathbf{n}}$, an index set $I = \{1, 2, ..., n\}$ and a surjection

$$i_{AGT} : I \longrightarrow \Phi_{AGT}$$

where Φ_{AGT} is a partition of AGT with each class being countably infinite.[3] This ensures that we will not run out of variables denoting each agent. We enumerate the class $i_{AGT}(k) \in \Phi_{AGT}$ by $\{x_{k1}, x_{k2}, ...\}$. On the basis of this partition of the agent denoting variables, we define a partition of the operators $\mathsf{K} = \{K_x\varphi : x \in VAR, \varphi \in \mathcal{L}^{\mathbf{n}}\}$ of $n + 1$ sets such that for $k = 1, ..., n$, $\mathsf{K}_k =$

[3] That is $\Phi_{AGT} = \{AGT_k : \cup_{k \in I} AGT = AGT \wedge (k \neq l \Rightarrow AGT_k \neq AGT_l) \wedge \forall k \in I : |AGT_k| = \aleph_0\}$

$\{K_{x_{km}}\varphi : x_{km} \in i_{AGT}(k)\}$ and $\mathsf{K}_A = \{K_x\varphi : \neg\exists k \in I : x \in i_{AGT}(k)\}$. We denote the partition $\{\mathsf{K}_1, ..., \mathsf{K}_n, \mathsf{K}_A\}$ by K/i, and where $K_x\varphi \in \mathsf{K}_i, i \in \{1, ..., n, A\}$, we will call K_x a K_i-modality. For each $k \in I$, the class of operators will be correlated by the semantics with one of n agent in the domain. The last class is related to all other objects.

Semantics

Frames. As we are now dealing with languages with more expressive power than standard first-order modal logic, the definition of frames becomes increasingly complex.[4] Thus items 4 and 5 below will not occur in standard definition of first-order frames. They are required for present purposes as we need to be sure this information is encoded at the most basic level of the semantic structures. Item 4 defines the set of agents included in the domain, and item 5 correlates each object of the domain with an accessibility relation.
Definition: n-frame. An **n**-*frame* for a language $\mathcal{L}^{\mathbf{n}}$ is a quintuple $\mathcal{F} = \langle W, R^{\mathbf{n}}, Dom, Agt^{\mathbf{n}}, \sim\rangle$ where

1. W is an non-empty is a set of states
2. $R^{\mathbf{n}}$ is a non-empty set of n relations on W (that is, $R^{\mathbf{n}} = \{R_1, ..., R_n\}$,and $\forall R_i \in R^{\mathbf{n}} : R_i \subseteq W \times W$)
3. Dom is a non-empty domain of quantification
4. $Agt^{\mathbf{n}}$ is a privileged subset of Dom consisting of n elements, called *agents*
5. $\sim: Dom \longrightarrow R^{\mathbf{n}} \cup \{W \times W\}$ is a function such that $\forall i \in Agt^{\mathbf{n}} :\sim (i) \in R^{\mathbf{n}}$ and $\forall j \in Dom \backslash Agt^{\mathbf{n}} :\sim (j) = W \times W$ ■

To ease notation in the following, we enumerate $Agt^{\mathbf{n}}$ by $\{a_1, ..., a_n\}$ and assume that $\sim (a_i) = R_i$. The elements of W will also be referred to as *worlds*, and below we will often omit explicit reference to **n** when speaking of **n**-frames.

Though it would be interesting to work with more general frames, with no restrictions on the domain and agent set, the present semantics are simpler, and hence easier to work with. Allowing for varying domains – and with them varying agent sets – would allow for interesting applications, like the death or firing of agents, but would complicate proving meta-theoretical results.

In 5. in the definition, each agent from $Agt^{\mathbf{n}}$ is related its accessibility relation and further, each non-agent object is related to the universal binary relation on W. The latter is partly an artificial choice as elements of the domain that are not agents should never know anything. From a mathematical point of view this is choice is the preferred one over mapping the class to $\emptyset$, as this would require a far more complex axiom system. In addition, our choice leads to operators from K_A to behave like global modalities, which adds extra expressibility to the logics and eases meta-theoretical results. Though, as formulas $K_t\varphi$ where t is a non-agent denoting term, and K_t thus a global modality, will only be true when

[4] At least the expressive power is different. It has not been proven that we have an conservative extension of the systems defined in for example [3] or [4], though this is suspected.

φ is a valid in a specific model, one can take it that this construction merely models what it is to be dumb as a door. Further, adding global modalities for all non-agent objects allows the rule of Knowledge Generalization to preserve truth and Dual to remain valid (see further below).

Interpretation, model and valuation. We now define an *interpretation* to be a function

$$\mathcal{I} : REL_n \times W \longrightarrow \mathcal{P}(Dom^n)$$

that to each n-ary relation symbol assigns some set of n-tuples $(d_1, ..., d_n)$, where $\{d_1, ..., d_n\} \subseteq Dom$ relative to each world $w \in W$. The special case of identity is treated by letting $\mathcal{I}(=, w) = \{(d, d) : d \in Dom\}$ for all w. Further, let

$$\mathcal{I} : CON \times W \longrightarrow Dom$$

and we denote the value of constant c in w with $\mathcal{I}(c, w)$ where $\mathcal{I}(c, w) \in Dom$.[5] A frame $\mathcal{F}$ augmented with an interpretation $\mathcal{I}$ is called a *model*, and is denoted $M = \langle \mathcal{F}, \mathcal{I} \rangle$, and a *valuation* is defined as a surjective (onto) function

$$v : VAR \longrightarrow Dom$$

where in particular

$$v : AGT \longrightarrow Agt^{\mathbf{n}}$$

too is surjective, with the further requirement that if $x, y \in i_{AGT}^{-1}(k)$ for some $k \in I$, then $v(x) = v(y)$ and if $x \in VAR \backslash AGT$ then $v(x) \notin Agt^{\mathbf{n}}$. Where v and v' are valuations in M, and $v(y) = v'(y)$ for all $y \in VAR$, except (possibly) x, v' is called an *x-variant* of v. We will use $t^{w,v}$ to denote the extension of the term t at world w under valuation v given the model specified by the context. That is, where $t \in CON, t^{w,v} = \mathcal{I}(t, w)$ and where $t \in VAR, t^{w,v} = v(t)$.

That constants are chosen to be defined as non-rigid is motivated by the interpretation of first-order epistemic logic found in [6] and [3]. Here, the formula $K_c a = b$ is read 'c knows that a and b are the same/denote the same object'. Giving the constants a rigid interpretation would result in all identity statements being valid, and hence known, in each particular model. Hence agents would always be able to tell all things apart perfectly – no two things would ever be indistinguishable to any agent, and it would not be possible to model scenarios where an agent has lost track of which coffee cup was his.

When using non-rigid constants, a scoping mechanism is often included in order to disambiguate – like the predicate abstraction of [4]. In the present, such is not required as the reading of possibly ambiguous formulas, as $\widehat{K}_t\varphi(a)$, have a natural reading in the epistemic interpretation where knowledge operators are taken to be primary over constants in the sense of [4, p.189]. Such a reading further fits with a sequential decomposition of formulas when using the following truth conditions and allows for meaningful constructions of *de re/de dicto* statements as presented in [3].

[5] It worth noticing that we here define constants as being *non-rigid*, i.e. constants are allowed to take different values in different worlds. This affects the validity of certain axioms, as will be mentioned later.

Truth conditions. Where M is a model, $w \in W, \mathcal{I}$ is an interpretation and v a valuation we denote truth of φ/φ being satisfied at w in M under v by $M, w \vDash_v \varphi$, and define truth/satisfaction recursively as follows:

$$
\begin{array}{ll}
M, w \vDash_v P(t_1, t_2, ..., t_n) & \text{iff } (t_1^{w,v}, t_2^{w,v}, ..., t_n^{w,v}) \in \mathcal{I}(w, P) \\
M, w \vDash_v t_1 = t_2 & \text{iff } (t_1^{w,v}, t_2^{w,v}) \in \mathcal{I}(=, w) \text{ iff } t_1^{w,v} = t_2^{w,v} \\
M, w \vDash_v \neg\varphi & \text{iff not } M, w \vDash_v \varphi \\
M, w \vDash_v \varphi \wedge \psi & \text{iff } M, w \vDash_v \varphi \text{ and } M, w \vDash_v \psi \\
M, w \vDash_v \forall x P(x) & \text{iff } M, w \vDash_{v'} P(x) \text{ for all } x\text{-variants of } v
\end{array}
$$

Recall that $\sim (t^{w,v}) \in R^{\mathbf{n}} \cup \{W \times W\}$. We define truth of modal statements thus:

$$
\begin{array}{l}
M, w \vDash_v K_t\varphi \text{ iff } \forall w' : (w, w') \in \sim (t^{w,v}),\ M, w' \vDash_v \varphi \\
M, w \vDash_v \hat{K}_t\varphi \text{ iff } \exists w' : (w, w') \in \sim (t^{w,v}) \text{ and } M, w' \vDash_v \varphi
\end{array}
$$

These definitions for operator semantics ensures the validity of both the rule of inference *Knowledge Generalization* and the axioms K and *Dual*, all constitutive of standard, normal modal logics.[6]

The Axiom System $\mathbf{K_n}$

In the following, we list the axioms proposed in [10] for a term-modal version of normal modal logic for n agents in a language $\mathcal{L}^{\mathbf{n}}$, which we will here denote $\mathbf{K_n}$. The axiom system is inspired by those for first-order modal logic used in [2,3,7]. The axiomatic system for $\mathbf{K_n}$ includes all substitution instances of validities of propositional logic[7] and the following axioms (axiom schemes), starting with first-order axioms:

- $\forall$: Where φ is any formula of $\mathcal{L}^{\mathbf{n}}$ and y is a variable not bound in φ: $\forall x\varphi \rightarrow \varphi(y/x)$
- Id: Where t is any term: $t = t$
- PS: For all variables x, y: $(x = y) \rightarrow (\varphi(x) \leftrightarrow \varphi(y))$
- $\exists$Id: For any constant c: $(c = c) \rightarrow \exists x (x = c)$

We must restrict both $\forall$ and PS to variables only since unrestricted versions of these two axioms result in an unsound system when the semantics are defined using non-rigid constants.[8] Further, we include the following modal axioms:

- K: $K_t(\varphi \rightarrow \psi) \rightarrow (K_t\varphi \rightarrow K_t\psi)$
- Dual: $K_t\varphi \leftrightarrow \neg\hat{K}_t\neg\varphi$

Both K and Dual are constitutive axioms of normal modal logics. Further, where $K_x\varphi \in \mathsf{K}_A$, we add the following axioms schemes in order for these modalities be global:

[6] Proofs are left as an exercise to the reader to look up in [10].

[7] That is, where φ is a validity of propositional logic with all propositional variable uniformly replaced by formulas of $\mathcal{L}^{\mathbf{n}}$, φ is an axiom of $\mathbf{K_n}$.

[8] See [3], p. 88-90, for a proof with respect to regular first-order modal logic. The proof given there carries over almost without change.

- T: $K_x\varphi \rightarrow \varphi$
- 4: $K_x\varphi \rightarrow K_xK_x\varphi$
- B: $\varphi \rightarrow K_x\hat{K}_x\varphi$
- $\hat{K}_y\varphi \rightarrow \hat{K}_x\varphi$

Here, T, 4 and B in conjunction results in all non-agent objects' accessibility relations being equivalence relations, and Inclusion ensures that all relevant states are related, why the relation(s) are universal. Finally, we include mixed axioms to control the interplay between quantifiers and operators:

- Where $t \neq x$, the Barcan Formula[9]: $\forall x K_t\varphi(x) \rightarrow K_t\forall x\varphi(x)$
- Knowledge of Non-identity[10]:$(x_1 \neq x_2) \rightarrow K_t(x_1 \neq x_2)$

The inference rules of $\mathbf{K_n}$ consists of Modus Ponens

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi}, \tag{MP}$$

and Universal Generalization well-known from first-order logic: where x does not occur free in φ,

$$\frac{\varphi \rightarrow \psi}{\varphi \rightarrow \forall x\psi} \tag{Gen}$$

and finally Knowledge Generalization constitutive of normal modal logics:

$$\frac{\varphi}{K_t\varphi} \tag{KG}$$

That KG continues to preserve truth under the semantics provided is due to the choice that all non-agent elements of the domain are correlated with total binary relation on W. Had we instead chosen that $K_t\varphi$ should be false for all non-agent denoting terms t and all formulas, as might have been a more just modeling as no non-agent objects ever know anything, the rule would obviously need side-conditions in order to preserve truth. One way to remedy this would be to add agent axioms, $A(x)$, for all the agent variables of AGT, and require that both $A(x)$ and φ be provable before one could conclude $K_x\varphi$. One could then remedy the semantics be requiring that $A(x)$ be true at w in conjunction with our ordinary requirements, but this would result in a conjunction being the main connective in the truth definition for such modal operators, which would then again result in the invalidity of the axiom Dual.

Assuming standard definition of $\mathbf{K_n}$-proofs (a la [1]), we can now define a *normal n agent term-modal logic* as any set of formulas Λ from $\mathcal{L}^{\mathbf{n}}$ that contains all $\mathbf{K_n}$ axioms and which is closed under $\mathbf{K_n}$ inference rules. If we further adopt regular definitions for validity and semantic consequence, say that a logic Λ is *(strongly) complete* with respect to S if, for any set of formulas $\Gamma \cup \{\varphi\}$, if $\Gamma \models_{\mathsf{S}} \varphi$, then $\Gamma \vdash_{\Lambda} \varphi$, where S is a class of frames (or models), the following result can be proven:

[9] The BF is not valid if we allow $t = x$. Fortunately, when BF is used in the proof of the existence lemma given in [10], this does not matter.

[10] The axiom is restricted to variables as an unrestricted version is invalid with non-rigid constants.

Theorem: Canonical Class Theorem. Any normal n agent term-modal logic Λ is (strongly) complete with respect to the class of canonical models for Λ. ■[11]

Adding Classic Epistemic Axioms

As the construction of the logics presented here have epistemic applications in mind, further axioms are required for the standard interpretation, namely those of classic epistemic logic – T, $K_i\varphi \rightarrow \varphi$, and 5, $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$, classically for any $i \in I$, which for validity requires that the accessibility relations of the agents are reflexive for T and euclidean for 5.[12]

When using term-modal operators, even when the accessibility is assumed to have the appropriate properties, this does not entail the validity of axioms with nested operators, such as 4 or 5 – that is, if these are added for all terms.

We can illustrate this with the very simple case of 4, $K_i\varphi \rightarrow K_iK_i\varphi$, which when using regular operators and the corresponding semantics, characterize transitivity. For assume for some model M, state w, valuation v, constant a and formula φ that $M, w \models_v K_a\varphi$. Hence, for all w' such that $(w, w') \in\sim (\mathcal{I}(a, w))$, $M, w' \models_v \varphi$. Now assume $M, w \models_v \neg K_aK_a\varphi$. Then for some w'' such that $(w, w'') \in\sim (\mathcal{I}(a, w))$, $M, w'' \models_v \neg K_a\varphi$, and for this we must have for some w''' such that $(w'', w''') \in\sim (\mathcal{I}(a, w''))$ and $M, w''' \models_v \neg\varphi$. Now we may ordinarily conclude by transitivity that as $(w, w''), (w'', w''') \in\sim_i$, so will (w, w''') and hence $M, w''' \models_v \varphi$ by assumption, which leads to a contradiction. This last step is unwarranted, though, under the present semantics as $\sim (\mathcal{I}(a, w))$ and $\sim (\mathcal{I}(a, w''))$ may in fact not be the same relation over W at all, as the interpretation of a may vary form world to world, why there is no guarantee that $\sim (\mathcal{I}(a, w)) =\sim (\mathcal{I}(a, w''))$.

If we on the other hand add such axioms restricted to variables occurring as indexes, these will be valid. That is, if we add $K_x\varphi \rightarrow K_xK_x\varphi$, this will be valid exactly on transitive frames, and $\neg K_x\varphi \rightarrow K_x\neg K_x\varphi$ will be valid exactly on euclidean frames. Hence adding such versions of T and 5 for all variables will result in a system reminiscent of classic epistemic logic. Here, each agent's accessibility relation will be an equivalence relation, and the accessibility relation of all non-agents will be the universal relation on the frame of the given model. As mentioned, such a system will not validate constant-indexed versions of for example 4.

It is interesting to note that the invalidity of this was thought of as reasonable in [6], where Hintikka writes: "we must therefore assume that the person referred to by a knows that he is referred to by it ... that it is true to say "a knows that he is a"...". For if we assume for some model M that $M, w \models_v \exists x K_a (x = a)$, i.e. that a knows who a is in the reading of [6] and [3], it follows that $M, w \models_v K_a\varphi \rightarrow K_aK_a\varphi$. Indeed, it is seen that this results exactly as the assumptions insures that $\mathcal{I}(a, w')$ remains constant for all w' in $\sim (\mathcal{I}(a, w))$, i.e. that the constant a designates rigidly over the states accessible for agent $\mathcal{I}(a, w)$ from w.

[11] For all definitions and the proof of this theorem, see [10].
[12] As T and 5 characterizes reflexive and euclidean frames, respectively, cf. [1].

Examples of Expressibility

Returning to the motivating example from the introduction, we are now in a position to express the student's predicament at the summer school, for $K_a\left(PC\left(x\right) \rightarrow K_x\varphi\right)$ will in fact be a well-formed formula. Hence the student of the example gain the ability to infer from strictly extensional features to second-order information in non-*ad hoc* manner.

Relating to the reasoning of Descartes, the knowledge operators now also function as predicates, hence allowing us to conclude existence from knowledge. Not only is $K_a\varphi \rightarrow \exists x K_x\varphi$ expressible in the present term-modal logic, it is also a validity.

Further, we gain the expressibility to define an "everybody knows that"-operator, E_G, using a unary predicate G to denote the group of agents of interest: we can simply define $E_G\varphi$ as $\forall x\left(G\left(x\right) \rightarrow K_x\varphi\right)$. This is sharp contrast to classic first-order epistemic logic, where $E_G\varphi$ is not well-formed for predicate G from the first-order language.

Moreover, $E_G\varphi$ is true just in the case where all agents in G knows φ – that is, it captures exactly the behavior of the E_G-operator of [3], though it cannot be nested as easily. This is due to the fact that predicates non-rigid, and operators for varying groups would result, cf. [8]. The behavior of rigid, nested operators can be captured if we let $\forall^n x$ denote the quantifier block $\forall x_n \forall x_{n-1} ... \forall x_1$, $G^n\left(x\right)$ denote the conjunction $G\left(x_n\right) \wedge G\left(x_{n-1}\right) \wedge ... \wedge G\left(x_1\right)$ and $K_x^n\varphi$ denote the formula $K_{x_n}K_{x_{n-1}}...K_{x_1}\varphi$.[13] We can then define a series of operators by setting $E_G^0\varphi := \varphi$ and $E_G^{n+1}\varphi := \forall x_{n+1}\forall^n x\left(G^n\left(x\right) \wedge G\left(x_{n+1}\right) \rightarrow K_{x_{n+1}}K_x^n\varphi\right)$. This will result in fixing G throug relevant worlds. We can then define the truth conditions for a 'common knowledge in G' operator, C_G, by $M, w \models_v C_G\varphi$ iff $M, w \models_v E_G^k\varphi$ for $k = 1, 2, ...$

Utilizing the newly-gained expressibility, we can further ensure that membership of G is common knowledge in G and not know by anyone else by requiring the truth of $C_G\forall y\left(G\left(y\right) \leftrightarrow K_{x_1}G\left(y\right)\right)$,[14] and adding to this define "secret club" common knowledge of φ by $C_G\forall y\left(K_y\varphi \leftrightarrow G\left(y\right)\right)$, knowledge that would be handy to be able to express when modeling convention-based choice in coordination games in the style of [5].

Conclusion and Further Perspectives

The system $\mathbf{K_n}$and corresponding semantics presented is behaves nicely with respect completeness, and it is easy to add axioms ensuring the regular properties of first-order epistemic logic, where the resulting systems have an interesting added expressibility. The systems allow for reasoning about knowledge as a property of existing agents, which in turn results in interesting interplay between terms and predicates from the first-order language and the modal operators.

It would be interesting to investigate whether C_G-operator in fact behaves like that of [3], and it is an open question what relation the proposed system

[13] We let both $\forall^0 x$ and $G^0\left(x\right)$ denote the empty string, and $K_n^0\varphi := \varphi$

[14] Where x_1 is bound by the universal quantifier of E_G^1.

stands in with other term-modal systems, like those of [8,11], and to what degree meta-theoretical results can be obtained for varying domains and agent sets for the currently specified systems. Further, to eliminate truth of $K_t\varphi$ where t is a non-agent denoting term, one might consider using a two-sorted language, as proposed in[8]. It would interesting to see what meta-theoretical results could be proven, as well as consider the philosophical foundations for systems in which agents can always tell agents non-agent objects.

References

1. Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic.* Cambridge University Press, 2001.
2. T. Brauner and S. Ghilardi. First-order modal logic. In F. Wolter P. Blackburn, J. van Bentham, editor, *Handbook of Modal Logic.* Elsevier Science, 2007.
3. Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge.* The MIT Press, 1995.
4. Melvin Fitting and Richard L. Mendelsohn. *First-Order Modal Logic.* Springer, 1999.
5. Pelle Guldborg Hansen. *Social Convention: In Defense of David Lewis.* PhD thesis, Roskilde University, 2010.
6. Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Cornell University Press. Reprinted in 2005, prepared by V.F. Hendricks and J. Symons, King's College Publications, 1962.
7. G.E. Hughes and M.J. Cresswell. *A New Introduction to Modal Logic.* Routledge, 1996.
8. Barteld Kooi. Dynamic term-modal logic. In F. Veltman J. van Benthem, S. Ju, editor, *A meeting of the minds, Proceedings of the Workshop on Logic, Rationality and Interaction, Beijing, 2007*, Texts in Computer Science 8, pages 173–185. Texts in Computer Science 8. pp. 173-185, College Publications, London, 2007.
9. A. Lomuscio and M. Colombetti. Qlb: a quantified logic for belief. In *Lecture Notes in AI, Vol. 1193: Proceedings of ATAL-96*, volume 1193, 1996.
10. Rasmus K. Rendsvig. *EA Logics and Normality.* Working paper, see http://sites.google.com/site/rendsvig/Home/uploads, 2009.
11. Lars Thalmann. *Term-Modal Logic and Quantifier-free Dynamic Assignment Logic.* PhD thesis, Uppsala University, 2000.
12. Georg Henrik von Wright. *An Essay in Modal Logic.* Studies in logic and the foundations of mathematics. North-Holland Publishing Company, 1951.

Universal Turing machines without using codification

Anderson de Araújo

State University of Campinas (UNICAMP)
PhD Program in Philosophy (Logic) - IFCH,
Group for Theoretical and Applied Logic - CLE,
P.O. Box 6133, 13081-970, Campinas - SP, Brazil.
anderaujo@gmail.com

Abstract. This article shows the existence of a universal Turing machine without using codification. That purpose is achieved by delineating a first-order axiomatic theory with axioms for all Turing machines and defining a model for this theory which is a universal Turing machine. The main conclusion is that codification is not necessary for proving the universality of a computing device, a conclusion that can provide a new approach to the universality in quantum Turing computability.

Keywords: Universal Turing machines, Quantum Turing computability, Codification.

1 Introduction

In the study of the notion of Turing computability, the existence of universal Turing machines is a central point [6]. In 1985, Deutsch [7] introduced *quantum Turing machines* as precise models of quantum computers and proposed a model of universal quantum Turing machines. Although many works have developed the idea of quantum Turing machines, still today there is no consensus about the existence of a universal quantum Turing machine, as we can verify in [18, 10]. Nowadays one of the main themes of research in theoretical computer science is quantum computing [14], it would be desirable, then, to understand how universality emerges in Turing computability as a whole.

Let Mac be the set of all Turing machines. An *encoding* is a function f from Mac to $\mathbb{N}$ which associates to each Turing machine $M \in Mac$ the (numerical) *code* $f(M)$. By *codification* we mean the process of encoding and decoding (the inverse of encoding). In general, codification has a central role in the construction of universal computing devices. In quantum computing that is not different and, roughly speaking, we can say that the problems with respect to universality in quantum Turing machines are associated to the possibility of combining simulation (which involves codification) and superposition of configurations in quantum computations[1].

[1] Cf. [10] for details.

Nonetheless, when Turing [16] showed the existence of a universal Turing machine, he did not use codification. Indeed, Turing gave a standard description for each machine that computes a particular function as a kind of codification for the universal computing machine and, by doing so, he defined a table of instructions for a universal Turing machine based on this standard description. It was Kleene [12] that used his primitive recursive predicate T, which describes the trees of computation of the partial recursive functions via codification, in order to proof the existence of a universal partial recursive function.

In 1950's years, Davis [4, 5] defined the concept of universal Turing machine in terms of Kleene's predicate and therefore he established Gödel numbering as part of the definition of a universal Turing machine. Following Davis' approach, in turn, today the use of numerical codification in the study of universal computing machines is treated as essential. In order to verify this, it is sufficient to consider the papers in the compendium [11] that contains a half-century survey of research about universal Turing machines or the survey of recent works on small universal Turing machines presented in [17].

In the present paper, however, we shall show the existence of universal Turing machines without using any kind of codification. Thus, we would like to contribute to the undestanding of universality in Turing computability, which, as we said above, has a central importance in the study of quantum computing. The structure of the paper is the following. In section 2, it is delineated a first-order axiomatic theory for Turing machines. According our logical approach, Turing machines are conceived as models for first-order axiomatic theories. In section 3, a universal Turing machine is defined as a first-order structure that is a model for the axiomatic theory that contains axioms for all Turing machines.

2 Turing machines and Turing structures

When Turing defined his model of computation in [16], he had the insight to give a characterization of the computations of his automaton in a first-order language. This is what enabled Turing to show the unsolvability of Hilbert's *Entscheidungsproblem* as a consequence of the uncomputability of the halting problem. Turing's intuition can be sharpened, and an appropriate first-order formal system for Turing machines as a whole can be developed.

Definition 21 *The* Turing language $\mathcal{L}$ *is the first-order language with signature* $\mathcal{S} = \{0, +1, -1, \prec, Q, S\}$*, whose elements are, respectively, the constant* zero*, the unary function symbols* successor *and* predecessor*, the binary relation symbol for* order relation*, the ternary relation symbols for* state relations *and* symbol relations*.*

We shall assume the definitions of all syntactic notions as in [8]. In particular, we shall denote the *provability relation* by $\vdash$. To make the text more readable, n and $-n$ will be, respectively, abbreviations for 0 followed by n applications of $+\dot{1}$ and $-\dot{1}$. In this work, we shall pressupose all basic notions about Turing computability in [15], but we conceive Turing machines like in [9].

2

Definition 22 *A* Turing machine *is a non-empty function M such that for some natural number n,*

$$Dom(M) \subseteq \{0, 1, \ldots, n\} \times \{0, 1, 2\},$$

$$Cod(M) \subseteq \{0, 1, \ldots, n\} \times \{0, 1, 2\} \times \{\triangleleft, \triangleright\},$$

where:

1. *The elements in the first factor of $Dom(M)$ and $Cod(M)$ are the* states *of M, 0 is the* initial state *and n the* final state*;*
2. *The elements in the second factor of $Dom(M)$ and $Cod(M)$ are the* symbols *of M, 0 is the numeral zero, 1 is the numeral one, and 2 is the empty symbol;*
3. *The elements of the third factor of $Cod(M)$ are the* moves *of M, $\triangleleft$ is the* move to the left*, $\diamond$ is the* neutral move*, and $\triangleright$ is the* move to the right*.*

Of course, in defintion 22 we are pressuposing a binary represention of the natural numbers and are considering only computations over natural numbers. As usual, we shall denote the empty symbol by $\sqcup$ instead of 2. Given this representation of Turing machines, we can prove the next result without using codification.

Proposition 21 *The set of all Turing machines Mac can be effectively well ordered.*

Proof. Since each Turing machine has a finite non-empty number of instructions, we can define a sequence $\langle S_i \rangle_{i \in \mathbb{N}}$ where each S_i is a set of Turing machines in Mac with $i + 1$ instructions. By definition, the set of states of any Turing machine has the form $\{0, 1, 2, \ldots, n\}$, then we can show, with a straightforward induction, that each Turing machine in S_i has at most $i + 2$ states. Moreover, since each instruction has five components $(state, symbol, symbol, move, state)$ and the number of states in a Turing machine is at most $i+2$, we also know that each S_i has $\frac{((i+2)^2 \cdot 8)!}{(i+1)! \cdot (((i+2)^2 \cdot 8) - (i+1))!}$ Turing machines - this number correspond to the simple combination of the possible instructions. Therefore, each S_i can be effectively enumerated and, by extension, we can defined a well order in Mac induced by $\langle S_i \rangle_{i \in \mathbb{N}}$. □

We emphasize that the proof of proposition 21 does not involve any kind of codification, because this fact will be crucial in our proof of the existence of a universal Turing machine. Now we associate to each computation of a particular Turing machine a $\mathcal{S}$-structure.

Definition 23 *The* Turing structure *associated to the computation $C_M^{s_0, \ldots, s_n}$ of Turing machine M is the first-order $\mathcal{S}$-structure $\mathfrak{A}^s_M = (\mathbb{N}, 0^{\mathfrak{A}^s_M}, +1^{\mathfrak{A}^s_M}, -1^{\mathfrak{A}^s_M}, \prec^{\mathfrak{A}^s_M}, Q^{\mathfrak{A}^s_M}, S^{\mathfrak{A}^s_M})$ where:*

1. *$\mathbb{N}$ is the set of natural numbers;*
2. *$0^{\mathfrak{A}^s_M}$ is the number zero;*

3

3. $+1^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *is the succesor function defined on* $\mathbb{N}$*;*
4. $-1^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *is the (bounded) predecessor function defined on* $\mathbb{N}$*;*
5. $\prec^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *is the strict order relation of* $\mathbb{N}$*;*
6. $Q^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *and* $S^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *are relations in* $\mathbb{N}^3$ *recursively defined on the last factor:*
   – $(0,0,0) \in Q^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *and* $(0, s_0, 0) \in S^{\mathfrak{A}^{\boldsymbol{s}}_M}, \ldots, (n, s_n, 0) \in S^{\mathfrak{A}^{\boldsymbol{s}}_M}$*;*
   – *If* $(q, e, t) \in Q^{\mathfrak{A}^{\boldsymbol{s}}_M}$*,* $(s, e, t) \in S^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *and* $M(q, s) = (p, r, 0)$*, then* $(p, c - 1^{\mathfrak{A}^{\boldsymbol{s}}_M}, t + 1^{\mathfrak{A}^{\boldsymbol{s}}_M}) \in Q^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *and* $(r, c, t + 1^{\mathfrak{A}^{\boldsymbol{s}}_M}) \in S^{\mathfrak{A}^{\boldsymbol{s}}_M}$*;*
   – *If* $(q, e, t) \in Q^{\mathfrak{A}^{\boldsymbol{s}}_M}$*,* $(s, e, t) \in S^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *and* $M(q, s) = (p, r, 1)$*, then* $(p, c + 1^{\mathfrak{A}^{\boldsymbol{s}}_M}, t + 1^{\mathfrak{A}^{\boldsymbol{s}}_M}) \in Q^{\mathfrak{A}^{\boldsymbol{s}}_M}$ *and* $(r, c, t + 1^{\mathfrak{A}^{\boldsymbol{s}}_M}) \in S^{\mathfrak{A}^{\boldsymbol{s}}_M}$*.*

When there is no risk of confusion, we will denote Turing structures by $(\mathbb{N}, 0, +1, -1, \prec, Q, S)$. Further, we shall also pressupose all semantic notion as in [8]. In particular, we shall denote the *satisfatibility relation* by $\vDash$. What is the relationship between Turing machines and Turing structures? The next result gives the answer that we need.

Definition 24 *Let* M *be a Turing machine. A formula* $\varphi(x, y_0, \ldots, y_n, y_{n+1}, \ldots, y_{n+m+1}, t, \boldsymbol{z})$ *of the Turing language* $\mathcal{L}$ correspond *to the configuration* $C^{\boldsymbol{s}}_M(t)$ *of computation* $C^{\boldsymbol{s}}_M$ *when*

$$\mathfrak{A}^{\boldsymbol{s}}_M \vDash \varphi(q, a_0, \ldots, a_n, b_0, \ldots, b_m, t, \boldsymbol{z}) \text{ if and only if}$$
$$C^{\boldsymbol{s}}_M(t) = (q, b_0, \langle a_0, \ldots, a_n\rangle, \langle b_0, \ldots, b_m\rangle).$$

We say that a configuration $C^{\boldsymbol{s}}_M(t)$ *is* correspondable *in* $\mathcal{L}$ *when there is a formula of* $\mathcal{L}$ *that correspond to* $C^{\boldsymbol{s}}_M(t)$*.*

Theorem 21 (Correspondence) *Let* M *be a Turing machine. Then, all configuration of a computation of* M *is correspondable in the Turing language.*

Proof. Let M be a Turing machine and $C^{s_0,\ldots,s_k}_M$ be a computation of M. The proof is by induction of the steps of $C^{\boldsymbol{s}}_M$.

In the step 0, $C^{\boldsymbol{s}}_M(0) = (0, s_0, \emptyset, \langle s_0, \ldots, s_k\rangle)$. Clearly, the formula $Q(0,0,0) \wedge S(s_0, 0, 0) \wedge \cdots \wedge S(s_k, k, 0)$ of $\mathcal{L}$ correspond to the configuration $C^{\boldsymbol{s}}_M(0)$.

Suppose that there is a formula $\varphi(q, a_0, \ldots, a_n, b_0, \ldots, b_m, t, \boldsymbol{c})$ that correspond to $C^{\boldsymbol{s}}_M(t) = (q, b_0, \langle a_0, \ldots, a_n\rangle, \langle b_0, \ldots, b_m\rangle)$. Either $M(q, b_0) = (p, s, \triangleleft)$ or $M(q, b_0) = (p', s', \triangleright)$ (possibly $p = p'$ and $s = s'$).

If $M(q, b_0) = (p, s, \triangleleft)$, then $C^{\boldsymbol{s}}_M(t + 1) = (p, s, \langle a_0, \ldots, a_{n-1}\rangle, \langle a_n, s, b_1, \ldots, b_m\rangle)$ and we can verify that $\varphi(q, a_0, \ldots, a_n, b_0, \ldots, b_m, t, \boldsymbol{c})[q, b_0, t, \boldsymbol{c}/p, s, t + 1, \boldsymbol{d}]$ correspond to $C^{\boldsymbol{s}}_M(t + 1)$. Indeed, $\mathfrak{A}^{\boldsymbol{s}}_M \vDash \varphi(q, a_0, \ldots, a_n, b_0, \ldots, b_m, t, \boldsymbol{c})[q, b_0, t, \boldsymbol{c}/p, s, t + 1, \boldsymbol{d}]$ if and only if $(p, a_0, \ldots, a_n, s, b_1, \ldots, b_m, t + 1, \boldsymbol{d})$ satisfies $\varphi(x, y_0, \ldots, y_n, y_{n+1}, \ldots, y_{n+m+1}, w, \boldsymbol{z})$ for an apropriate sequence $\boldsymbol{d}$. By induction hypothesis, $(q, b_0, \langle a_0, \ldots, a_n\rangle, \langle b_0, \ldots, b_m\rangle)$ is the configuration of $C^{\boldsymbol{s}}_M$ in the step t and $\mathfrak{A}^{\boldsymbol{s}}_M \vDash \varphi(q, a_0, \ldots, a_n, b_0, \ldots, b_m, t, \boldsymbol{c})$ for an apropriate sequence $\boldsymbol{c}$. Clearly, the formulas $Q(q, e_{n+1}, t)$ and $S(a_0, e_0, t), \ldots, S(a_n, e_n, t), S(b_0, e_{n+1}, t), \ldots, S(b_m, e_{n+m1}, t)$ must be subformulas of $\varphi(q, a_0, \ldots, a_n, b_0, \ldots, b_m, t, \boldsymbol{c})$, for they are the unique formulas of $\mathcal{L}$ that represent the states and symbols of the computation $C_M^{s_0,\dots,s_k}$. As $M(q,b_0) = (p,s,\lhd)$, it follows that $(p,e_{n+1}-1,t+1) \in Q^{\mathfrak{A}_M^{\boldsymbol{s}}}$ and $(s,e_{n+1},t+1) \in S^{\mathfrak{A}_M^{\boldsymbol{s}}}$, i.e, $Q(p,e_{n+1}+1,t)$ and $S(a_0,e_0,t),\dots,S(a_n,e_n,t),S(s,e_{n+1},t),\dots,S(b_m,e_{n+m+1},t)$ are true in $\mathfrak{A}_M^{\boldsymbol{s}}$. Hence, considering $\boldsymbol{d} = e_{n+1}+1,e_0,\dots,e_{n+m+1}$ obtained from $\boldsymbol{c} = e_{n+1},e_0,\dots,e_{n+m+1}$, we know that $\mathfrak{A}_M^{\boldsymbol{s}} \vDash \varphi(p,a_0,\dots,a_n,s,b_1,\dots,b_m,t+1,\boldsymbol{d})$ if and only if $C_M^{\boldsymbol{s}}(t+1) = (p,s,\langle a_0,\dots,a_{n-1}\rangle,\langle a_n,s,b_1,\dots,b_m\rangle)$.

If $M(q,b_0) = (p',s',\rhd)$, then $C_M^{\boldsymbol{s}}(t+1) = (p',s',\langle a_0,\dots,a_n,s'\rangle,\langle b_1,\dots,b_m\rangle)$ and with a similar argument we can show that the formula $\varphi(q,b_0,a_0,\dots,a_n,b_0,\dots,b_m,t,\boldsymbol{c})[q,b_0,t,\boldsymbol{c}/p',s',t+1,\boldsymbol{d}]$ correspond to the configuration $C_M^{\boldsymbol{s}}(t+1)$, but in this case we consider $\boldsymbol{d} = e_{n+1}-1,e_0,\dots,e_{n+m+1}$. $\square$

Since the formula $Q(p,e_{n+1}+1,t) \wedge S(a_0,e_0,t) \wedge \cdots \wedge S(a_n,e_n,t) \wedge S(s,e_{n+1},t) \wedge \cdots \wedge S(b_m,e_{n+m+1},t)$ correpond to the configuration $C_M^{\boldsymbol{s}}(t) = (q,b_0,\langle a_0,\dots,a_n\rangle,\langle b_0,\dots,b_m\rangle)$ for each t, as showed by the theorem of correspondence 21, when we talk about *the* formula that correspond to the given configuration, we are talking about such a formula.

The correspondence between computations of Turing machines and computation in Turing structures permit us to proof results about one from the another, because we can define the computations of a Turing machine in the Turing structure $\mathfrak{A}_M^{\boldsymbol{s}}$.

Definition 25 *A* computation in $\mathfrak{A}_M^{\boldsymbol{s}}$ *is a (finite or infinite) sequence* $C_{\mathfrak{A}_M^{\boldsymbol{s}}}$ *whose items* $C_{\mathfrak{A}_M^{\boldsymbol{s}}}(i)$ *are elements in* $\mathbb{N}^2 \times \mathbb{N}^m \times \mathbb{N}^k$ *such that* $C_{\mathfrak{A}_M^{\boldsymbol{s}}}(i) = (q,b_0,\langle a_0,\dots,a_m\rangle,\langle b_0,\dots,b_k\rangle)$ *if and only if* $\varphi(q,a_0,\dots,a_n,b_0,\dots,b_m,i,\boldsymbol{e})$, *where* φ *is the formula that correspond to the configuration* $C_M^{\boldsymbol{s}}(i)$ *of the computation* $C_M^{\boldsymbol{s}}$ *of Turing machine* M.

Hence, we can proof that some property is validy about Turing machines, showing that it holds for Turing structures, and vice-versa.

3 Turing theories and universal Turing machines

Given that in the last section we associate $\mathcal{S}$-structures to Turing machines, now we can also associated Turing theories to Turing machines.

Definition 31 *The* Turing theory *associated to the computation* $C_M^{s_0,\dots,s_n}$ *of Turing machine* M *is the classic first-order* $\mathcal{S}$*-theory* $\mathcal{T}_M^{\boldsymbol{s}}$ *specified by the axioms:*

A1 $\forall x \neg(x \prec x) \wedge \forall x \forall y \forall z(x \prec y \wedge y \prec z \to x \prec z) \wedge \forall y \forall z(x \prec y \vee x \approx y \vee y \prec x) \wedge \forall x \neg(x \prec 0) \wedge \forall x \forall y \exists z(x \prec z \wedge (x \prec y \to z \prec y \vee z \approx y)))$;

A2 $\forall x \neg(x+1 \approx 0) \wedge \forall x \forall y(x+1 \approx y+1 \to x \approx y)) \wedge 0-1 \approx 0 \wedge \forall x(x+1-1 \approx x) \wedge \forall x \forall y(x \prec y+1 \leftrightarrow x \prec y \vee x \approx y)$;

A3 $\forall x \forall y(S(x,e,t) \wedge S(y,e,t) \to x \approx y) \wedge \forall x \forall y(Q(x,e,t) \wedge Q(y,e,t) \to x \approx y) \wedge \forall x \forall y(Q(q,x,t) \wedge Q(q,y,t) \to x \approx y)$;

5

A4 $\forall x \forall y (Q(a, x, y) \wedge S(b, x, y) \rightarrow Q(c, x + 1, y + 1) \wedge S(d, x, y + 1) \wedge \forall v \forall z (z \prec x \vee x \prec z \rightarrow (S(v, z, y) \leftrightarrow S(v, z, y + 1))))$ *if* $(a, b, c, d, \triangleright)$ *is an instruction of* M*;*

A5 $\forall x \forall y (Q(a, x, y) \wedge S(b, x, y) \rightarrow Q(c, x - 1, y + 1) \wedge S(d, x, y + 1) \wedge \forall v \forall z (z \prec x \vee x \prec z \rightarrow (S(v, z, y) \leftrightarrow S(v, z, y + 1))))$ *if* $(a, b, c, d, \triangleleft)$ *is an instruction of* M*;*

A6 $Q(0, 0, 0) \wedge S(s_0, 0, 0) \wedge \cdots \wedge S(s_n, n, 0) \wedge \forall x (n \prec x \rightarrow S(2, x, 0))$.

A similar axiomatization for Turing machines was proposed before by Boolos et alli [2, p.126-132], which is in tur based on the original Turing's work [16], but it does not permit the kind of caracterization of Turing computability that we are pursuing. In our propose, axiom $A1$ expresses that the elements in M are ordered by a linear strict order with a first but no last element. $A2$ expresses the usual behavior of the successor and predecessor functions on the order type considered. $A3$ establishes the uniqueness of the symbols and states, with respect to space and time, guaranteeing that M is an implementation of a deterministic machine. Axioms $A4$, $A5$ and $A6$ represent the three kind of instructions of M affirming that they are local operations, and $A6$ represents the input of computation $C_M^{s_0,\ldots,s_n}$ of M. In this way, it is an easy exercise to prove the correcteness of $\mathcal{T}_M^{\boldsymbol{s}}$.

Theorem 31 (Soundness) *For all sentences* φ *in* For*, if* $\mathcal{T}_M^{\boldsymbol{s}} \vdash \varphi$*, then* $\mathfrak{A}_M^{\boldsymbol{s}} \models \varphi$*.*

A completeness result for $\mathcal{T}_M^{\boldsymbol{s}}$ is out of the question, because Turing [16] showed that a theory like $\mathcal{T}_M^{\boldsymbol{s}}$ can express the halting problem for its model $\mathfrak{A}_M^{\boldsymbol{s}}$ and, thus, by the unsolvability of the halting problem, $\mathcal{T}_M^{\boldsymbol{s}}$ cannot be a complete theory. Nonetheless, for the purposes of computability, we only need a restrict notion of completeness.

Definition 32 *Let* $\mathfrak{A}_M^{\boldsymbol{s}}$ *be a Turing structure. The theory* $\mathcal{T}_M^{\boldsymbol{s}}$ *is* computational complete *when, for all configuration of a computation in* $C_{\mathfrak{A}_M^{\boldsymbol{s}}}$ *and formula* φ *that correspond to* $\mathfrak{A}_M^{\boldsymbol{s}}$*,* $\mathfrak{A}_M^{\boldsymbol{s}} \models \varphi$ *if, and only if,* $\mathcal{T}_M^{\boldsymbol{s}} \vdash \varphi$*.*

To prove the computation completeness of a theory $\mathcal{T}_M^{\boldsymbol{s}}$, we need the following lemma, which is analogous to a result proved by Turing in his proof of the unsolvability of *Entscheidungsproblem.*

Lemma 31 *For all step* t*, if* φ *is the formula of Turing language that correspond to the configuration of* $C_{\mathfrak{A}_M^{s_0,\ldots,s_n}}$ *at* p*, then* $\mathcal{T}_M^{\boldsymbol{s}} \vdash \varphi$*.*

Proof. The proof is by induction on t. Assume the hypothesis, and suppose that $\mathscr{Q} = \{0, 1, 2, \ldots, m\}$ is set of states of $\mathfrak{A}_M^{\boldsymbol{s}}$. For $t = 0$, $C_{\mathfrak{A}_M^{\boldsymbol{s}}}(0) = (0, 0, 0, s_0, \emptyset, \langle s_0, \ldots, s_n \rangle)$ and, due to axiom $A6$, we know that the result is true.

Assume that at step t the result is true, and we shall show that it also is true for step $t + 1$. Let $C_{\mathfrak{A}_M^{\boldsymbol{s}}}(t) = (q, b_0, \langle a_0, \ldots, a_k \rangle, \langle b_0, \ldots, b_l \rangle)$. Or M stops at step $t+1$ or M does not stop. At first, we shall treat the case where M stops. In such a case, at step t, M either applied an instruction $\rho = (q, b_0, m, s, \triangleright)$ or $(q, b_0, m, r, \triangleleft)$ (possibily $r = s$). It is sufficient to consider the case in which M applied $(q, b_0, m, s, \triangleleft)$ because in the other case the proof is similar or easier than it. So, suppose that M applied $(q, b_0, m, s, \triangleleft)$. At step t, either $\langle a_0, \ldots, a_k\rangle = \oslash$ or $\langle a_0, \ldots, a_k\rangle \neq \oslash$.

Suppose that $\langle a_0, \ldots, a_k\rangle = \oslash$. Then $C_{\mathfrak{A}^{\boldsymbol{s}}_M}(t) = (q, b_0, \oslash, \langle b_0, \ldots, b_l\rangle)$ is indeed the configuration of $C_{\mathfrak{A}^{\boldsymbol{s}}_M}$ at step t. Then, by induction hypothesis, $\mathcal{T}^{\boldsymbol{s}}_M \vdash Q(q, c, t) \wedge S(b_0, c, t) \wedge S(b_1, c+1, t) \wedge \cdots \wedge S(b_l, c+l, t)$ for $c \geq 0$. Due to axiom $A5$, $\mathcal{T}^{\boldsymbol{s}}_M \vdash Q(q, c, t) \wedge S(b_0, c, t) \rightarrow Q(m, c-1, t+1) \wedge S(s, c, t+1) \wedge \forall v \forall z(z \prec x \vee x \prec z \rightarrow (S(v, z, y) \leftrightarrow S(v, z, y+1))))$. By classical logic, $\mathcal{T}^{\boldsymbol{s}}_M \vdash S(0, c-1, t) \vee \neg S(0, c-1, t)$. If $\mathcal{T}^{\boldsymbol{s}}_M \vdash \neg S(0, c-1, t)$, then, due to the soundness, $\mathfrak{A}^{\boldsymbol{s}}_M \vDash \neg S(0, c-1, t)$, but this happens if and only if $a_k \neq 0$ at step t, which contradicts our hypothesis. Thus, $\mathcal{T}^{\boldsymbol{s}}_M \vdash S(0, c-1, t)$ and, by first-order logic, $\mathcal{T}^{\boldsymbol{s}}_M \vdash Q(m, c-1, t) \wedge S(0, c-1, t+1) \wedge S(s, c, t+1) \wedge S(b_1, c+1, t+1) \wedge \cdots \wedge S(b_l, c+l, t+1)$, which it is the formula that correspond to the configuration of $C_{\mathfrak{A}^{\boldsymbol{s}}_M}$ at step $t+1$.

Now suppose that $\langle a_0, \ldots, a_k\rangle \neq \oslash$. Then $C_{\mathfrak{A}^{\boldsymbol{s}}_M}(t) = C_{\mathfrak{A}^{\boldsymbol{s}}_M}(t) = (q, b_0, \langle a_0, \ldots, a_k\rangle, \langle b_0, \ldots, b_l\rangle)$ for a cell $c > 0$. By induction hypothesis, $\mathcal{T}^{\boldsymbol{s}}_M \vdash Q(q, c+k+1, t) \wedge S(a_0, c, t) \wedge S(a_1, c+1, t) \wedge \cdots \wedge S(a_k, c+k, t) \wedge S(b_0, c+k+1, t+1) \wedge S(b_1, c+k+2, t+1) \wedge \cdots \wedge S(b_l, c+k+l+1, t+1)$ for $c > 0$. Due to axiom $A5$, $\mathcal{T}^{\boldsymbol{s}}_M \vdash Q(q, c+k+1, t) \wedge S(b_0, c+k+1, t) \rightarrow Q(m, c+k+1-1, t+1) \wedge S(s, c+k+1, t+1) \wedge \forall v \forall z(z \prec x \vee x \prec z \rightarrow (S(v, z, y) \leftrightarrow S(v, z, y+1))))$. Hence, by first-order logic, $\mathcal{T}^{\boldsymbol{s}}_M \vdash Q(m, c+k, t) \wedge S(a_0, c, t) \wedge S(a_1, c+1, t) \wedge \cdots \wedge S(a_k, c+k, t) \wedge S(s, c+k+1, t+1) \wedge S(b_1, c+k+2, t+1) \wedge \cdots \wedge S(b_l, c+k+l+1, t+1)$, which it is the formula that correspond to the configuration of $C_{\mathfrak{A}^{\boldsymbol{s}}_M}(t)$ in this case.

The proof for M does not stop is analogous to the case of M stops, we only need to change the final state m by the state p used in passage from step t to step $t+1$. □

Theorem 32 (Computational completeness) *Theory $\mathcal{T}^{\boldsymbol{s}}_M$ is computational complete.*

Proof. It is an immediate consequence of lemma 31 and the correcteness of $\mathcal{T}^{\boldsymbol{s}}_M$. □

This computational completeness shows that as far as Turing computability is concerned, we do not need any more than first-order logic, because we have an equivalence between Turing machines, Turing structures and Turing theories with respect to the computations of Turing machines. In this way, we can proof the existence of a universal Turing machine without using codification.

Take the relation symbols Q and S of Turing languagem $\mathcal{L}$ and indexed them by i according to the list $\langle M_i\rangle_{i\in\mathbb{N}}$ with all Turing machines of Mac defined from proposition 21. Thus, we have the Turing language $\overline{\mathcal{L}}$ with signature $\overline{\mathcal{S}} = \{0, +1, -1, \prec, \{Q_i\}_{i\in\mathbb{N}}, \{S\}_{i\in\mathbb{N}}\}$. Let $\mathcal{T}^{\boldsymbol{s}}_i$ be the theory associated to the computation $C^{s_0,\ldots,s_k}_M$ of i-th Turing machine in the list $\langle M_i\rangle_{i\in\mathbb{N}}$ with the respective i-signature. Now form the sequence $\langle \mathcal{T}^{\boldsymbol{s}}_i\rangle_{i\in\mathbb{N}}$ with all Turing theories.

Definition 33 *A* universal Turing theory *is a first-order theory* $\mathcal{T}_U^{\boldsymbol{s}} = \bigcup_i \{\mathcal{T}_i^{\boldsymbol{s}}\}_{i \in \mathbb{N}}$.

We say that a specific computing device is *universal* for a class of computing devices if it can compute all each particular computing devices in the class computes [3]. The next result shows that there is a Turing machine structure that is universal for the class Mac, that is, it exhibits a universal Turing machine conceived as a logical structure.

Theorem 33 *The theory* $\mathcal{T}_U^{\boldsymbol{s}}$ *has a model.*

Proof. Let Γ be a finite subset of $\mathcal{T}_U^{\boldsymbol{s}}$. Then Γ is a finite set of formulas with form as axioms $A1 - A6$ but indexed by natural numbers. By virtue of indexes, we know that there will be no conflict in the instructions and Γ is consistent. Either Γ has all axioms of one or more particular Turing machines or it has only part of the axioms of particular Turing machines. In any case Γ has a model: the models of the particular Turing machines of which Γ has the axioms. Therefore, by the compacteness theorem, we conclude that $\mathcal{T}_U^{\boldsymbol{s}}$ has a model. □

Moreover, we can constructivelly define a model for $\mathcal{T}_U^{\boldsymbol{s}}$.

Definition 34 *A* universal Turing structure *is a first-order structure* $\mathfrak{A}_U^{\boldsymbol{s}} = \bigcup_i \{\mathfrak{A}_i^{\boldsymbol{s}}\}_{i \in \mathbb{N}}$*, where each* $\mathfrak{A}_i^{\boldsymbol{s}}$ *is the structure associated to the i-machine in* $\langle M_i \rangle_{i \in \mathbb{N}}$.

Theorem 34 *The structure* $\mathfrak{A}_U^{\boldsymbol{s}}$ *is a model for* $\mathcal{T}_U^{\boldsymbol{s}}$ *and it is universal for* Mac.

Proof. By virtue of indexes in the relations symbols Q and S, we know that $\mathfrak{A}_U^{\boldsymbol{s}}$ is well defined (they avoid conflict in the instructions). This implies that all axioms of $\mathcal{T}_U^{\boldsymbol{s}}$ will be true in $\mathfrak{A}_U^{\boldsymbol{s}}$, and we can consistently define the computations of $\mathfrak{A}_U^{\boldsymbol{s}}$ as in definition 23. Due to the computational completeness, $\mathfrak{A}_U^{\boldsymbol{s}}$ is in fact universal for Mac. Therefore, $\mathfrak{A}_U^{\boldsymbol{s}}$ is a model for $\mathcal{T}_U^{\boldsymbol{s}}$. □

To conclude our proof of the existence of a universal Turing machine without using codification for the class Mac, we only need to appeal to the correspondece theorem.

4 Conclusion

When Lewis and Papadimitrou [13][p.247-250] define universal Turing machines using the standard method of codifying all Turing machines, they stress that the existence of a universal Turing machine is a consequence of the fact that "Turing machines are also software". Here, we have showed that the existence of universal computing devices can be conceived as a logical fact, and so we can say "Turing machines are also logical objects". Hence, the common view that codification is essential for a universal computing device is questioned here, and we can ask: what is the role of codification for universal computing devices?

The construction of the universal Turing structure in theorem 34 was made without any kind of codification, but this implied the infiniteness of the Turing structure - it has infinitely many relations Q and S. Therefore, it seems that codification would be just a way of guaranteeing the finiteness of the universal Turing machine. But, what about universal quantum Turing machines?

We believe that we can chance the underline logic of Turing theories by some quantum logic. This would imply the definition of quantum Turing structures as models of quantum Turing structures. In this way, using the same kind of method propoused in this paper, we would be able to analise the universality in quantum computing. An interesting kind of logical system to be consider in carring out such an approach is the exogenous quantum logic [1], because these logics model quantum systems embodying all that is stated in the postulates of quantum physics. Nonetheless, in order to do this, we need to extend exogenous quantum logic to the first-order case, and it seems a hard task to show a computational completeness result in this context.

Acknowledgments

This research is funded by São Paulo Research Foundation (FAPESP), Thematic Research Project (Grant 2004/14107-2), Scholarship Grant (2008/06205-5).

References

1. P. Mateus, A. Sernadas, Weakly complete axiomatization of exogenous quantum propositional logic. *Information and Computation* 204, 771-794, 2006.
2. G. Boolos, J. Burgess, R. Jeffrey. *Computablility and Logic.* 4th ed. Cambridge, Cambridge Univeristy Press, 2002.
3. M. Burgin. Universality, Reducibility, and Completeness. *Lecture Notes in Computer Science* 4664: 24-38, 2007.
4. M. D. Davis. A note on universal Turing machines. C. E. Shannon, J. McCarthy. *Automata Studies*, Annals of Mathematics Studies, Princeton University Press, pp.167-175, 1956.
5. M. D. Davis. The Definition of Universal Turing Machine. *Proceedings of the American Mathematical Society*, vol. 8, No. 6, pp. 1125-1126, 1957.
6. J-C. Delvenne. What is a universal computing machine. *Applied Mathematics and Computation* 215, 1368-1374, 2009.
7. . D. Deutsch, Quantum theory, the Church-Turing principle and the universal quantum computer. Proceedings of the Royal Society of London, Series A, *Mathematical and Physical Sciences* 400: 97-117, 1985.
8. H. Ebbinghaus, J. Flum, W. Thomas. *Mathematical Logic.* New York, Spring Verlag, 2005.
9. H.B. Enderton. Elements of recursion theory. In J. Barwise (ed.), *Handbook of mathematical logic.* Elsevier, Amsterdam, pp. 527-566, 1977.
10. W. Fouché, J. Heidema, G. Jones, P. H. Potgieter. Universality and programmability of quantum computers. *Theoretical Computer Science* 403: 121-129, 2008.
11. R. Herken (ed.). *The Universal Turing Machine - A Half-Century Survey.* Oxford, Oxford University Press, 1988.

12. S. C. Kleene. On Notation for Ordinal Numbers. *Journal of Symbolic Logic* 4 (3): 150-155, 1938.
13. H. R. Lewis, C. H. Papadimitriou. *Elements of the theory of computation.* 2^{nd} ed. Prentice-Hall, 1998.
14. M.A. Nielson, I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.
15. M. Sipser. *Introduction to the theory of computation.* 2nd ed. Massachusetts, Thomson Course Technology, 2006.
16. A. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 42(2): 230-265, 1936.
17. D. Woods, T. Neary. The complexity of small universal Turing machines: A survey. *Theoretical Computer Science* 410: 443-450, 2009.
18. Y. Shi, Remarks on universal quantum computer. *Physics Letters* A 293 (5-6): 277-282, 2002.

Posters

A Labelling Based Justification Status of Arguments

Yining Wu

Faculty of Sciences, Technology and Communication
University of Luxembourg, Luxembourg

Abstract. In this paper, we define a labelling based justification status of the arguments in an argumentation framework. For current purposes we assume the labellings to be based on complete semantics. We show how existing proof procedures can be reused in our labellings approach, both to determine the justification status of an argument and to enter a discussion with the user should he/she decide to dispute this status.

1 Introduction

The main concept in Dung's theory [1] is that of an argumentation framework, which is essentially a directed graph in which the nodes represent arguments and the arrows represent an attack relation.

Given such a graph, different sets of nodes can be accepted according to various argument based semantics such as grounded, preferred and stable semantics [1], semi-stable semantics [2] or ideal semantics [3]. Many of these semantics can be seen as restricted cases of complete semantics; an overview is provided in Figure 1. The facts that every stable extension is also a semi-stable extension and that every semi-stable extension is also a preferred extension has been proved in [2]. The facts that every preferred extension is also a complete extension and that the grounded extension is also a complete extension have been stated in [1]. The ideal extension is also a complete extension [3]. So complete extensions can be seen as the base for describing various other semantics in abstract argumentation.

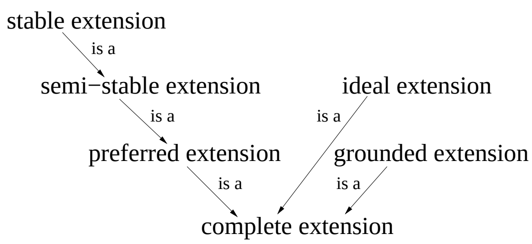


Fig. 1. An overview of the different semantics

A different way of defining argumentation semantics than the traditional extensions approach is the labellings approach [4, 5]. Where the extensions approach only identifies the set of arguments that are accepted, the labellings approach also identifies the set of arguments that are rejected and the set of arguments that are left undecided. The concept of argument labellings goes back to work of Pollock [6] and of Jakobovits and Vermeir [7]. However, for current purposes we will use the concept of complete labelling as defined by [4].

Essentially, a complete labelling can be seen as a subjective but reasonable position that an agent can take with respect to which arguments are accepted, rejected or undecided. In each such position the agent can use its own position to defend itself if questioned. It is possible to be disagree with a position, but at least the position is internally coherent. The set of all complete labellings thus stands for all possible and reasonable positions an agent can take.

In [4], it is stated that complete extensions and complete labellings are one-to-one related. In essence, complete extensions and complete labellings are different ways to describe the same concept.

In the current paper we will define the justification statuses of arguments based on the notion of a complete labelling. After this, some methods for determining this justification status will be given. This result provides an easy way to obtain all possible labels an argument could reasonably be assigned.

The remaining part of this paper is organized in the following way. We first state some preliminaries on argument semantics and argument labellings. Then we define the justification status of an argument and describe the methods for determining it. Finally we illustrate an implementation of the thus described theory and discuss the main results of the paper.

2 Argument Semantics and Argument Labellings

In this section, we briefly restate some preliminaries regarding argument semantics and argument-labellings. For simplicity, we only consider finite argumentation frameworks.

Definition 1. *An* argumentation framework *is a pair* (Ar, att) *where* Ar *is a finite set of arguments and* $att \subseteq Ar \times Ar$*.*

We say that argument A *attacks* argument B iff $(A, B) \in att$. An argumentation framework can be represented as a directed graph in which the arguments are represented as nodes and the attack relation is represented as arrows.

Definition 2 (defense / conflict-free). *Let* (Ar, att) *be an argumentation framework,* $A \in Ar$ *and* $\mathcal{A}rgs \subseteq Ar$*.* $\mathcal{A}rgs$ *is* conflict-free *iff* $\neg\exists A, B \in \mathcal{A}rgs : A$ *attacks* B*.* $\mathcal{A}rgs$ defends *argument* A *iff* $\forall B \in Ar : (B$ *attacks* $A \supset \exists C \in \mathcal{A}rgs : C$ *attacks* $B)$*. Let* $F(\mathcal{A}rgs) = \{A \mid A$ *is defended by* $\mathcal{A}rgs\}$*.*

We say that a set of arguments $\mathcal{A}rgs$ attacks an argument B iff there exists an $A \in \mathcal{A}rgs$ that attacks B. We write $\mathcal{A}rgs^+$ for the set of arguments that are attacked by $\mathcal{A}rgs$.

Definition 3 (acceptability semantics). *Let* (Ar, att) *be an argumentation framework. A conflict-free set* $\mathcal{A}rgs \subseteq Ar$ *is called an* admissible set *iff* $\mathcal{A}rgs = F(\mathcal{A}rgs)$*, and a* complete *extension iff* $\mathcal{A}rgs = F(\mathcal{A}rgs)$*.*

The concept of complete semantics was originally stated in terms of sets of arguments. It is equally well possible, however, to express this concept in terms of *argument labellings*. The approach of (argument) labellings has been used by Pollock [6] and by Jakobovits and Vermeir [7], and has more recently been extended by Caminada [8, 5], Vreeswijk [9] and Verheij [10]. In the current paper, we follow the approach of [8, 5] where a labelling assigns to each argument exactly one label, which can either be `in`, `out` or `undec`. The label `in` indicates that the argument is accepted, the label `out` indicates that the argument is rejected, and the label `undec` indicates that the status of the argument is undecided, meaning that one abstains from an explicit judgment whether the argument is `in` or `out`.

Definition 4 ([5]). *A* labelling *is a function* $\mathcal{L}ab : Ar \longrightarrow \{\texttt{in}, \texttt{out}, \texttt{undec}\}$*.*

We write $\texttt{in}(\mathcal{L}ab)$ for $\{A \mid \mathcal{L}ab(A) = \texttt{in}\}$, $\texttt{out}(\mathcal{L}ab)$ for $\{A \mid \mathcal{L}ab(A) = \texttt{out}\}$ and $\texttt{undec}(\mathcal{L}ab)$ for $\{A \mid \mathcal{L}ab(A) = \texttt{undec}\}$. Since a labelling can be interpreted as a partition of the set of arguments in the argumentation framework, we will sometimes write a labelling $\mathcal{L}ab$ as a triple $(\texttt{in}(\mathcal{L}ab), \texttt{out}(\mathcal{L}ab), \texttt{undec}(\mathcal{L}ab))$.

The idea of a *complete labelling* [8, 5] is that for a labelling to be reasonable, one should be able to give reasons for each argument one accepts (all attackers are rejected), for each argument one rejects (it has at least one attacker one accepts) and for each argument one abstains from expressing an explicit opinion about (there are insufficient grounds to accept it and insufficient grounds to reject it). This is made formal in the following definition.

Definition 5 ([5]). *Let* $\mathcal{L}ab$ *be a labelling of argumentation framework* (Ar, att)*. We say that* $\mathcal{L}ab$ *is a* complete labelling *iff for each and* $A \in Ar$ *it holds that:*

1. *If* $\mathcal{L}ab(A) = \texttt{in}$ *then* $\forall B \in Ar : (B\ att\ A \supset \mathcal{L}ab(B) = \texttt{out})$
2. *If* $\mathcal{L}ab(A) = \texttt{out}$ *then* $\exists B \in Ar : (B\ att\ A \wedge \mathcal{L}ab(B) = \texttt{in})$*.*
3. *If* $\mathcal{L}ab(A) = \texttt{undec}$ *then* $\neg\forall B \in Ar : (B\ att\ A \supset \mathcal{L}ab(B) = \texttt{out})$ *and* $\neg\exists B \in Ar : (B\ att\ A \wedge \mathcal{L}ab(B) = \texttt{in})$*.*

As stated in [8, 5], complete labellings coincide with complete extensions in the sense of [1].

Theorem 1 ([5]). *Let* $AF = (Ar, att)$ *be an argumentation framework.*

1. *If* $\mathcal{L}ab$ *is a complete labelling, then* $\texttt{Lab2Ext}(\mathcal{L}ab)$ *is a complete extension (where* $\texttt{Lab2Ext}(\mathcal{L}ab) = \texttt{in}(\mathcal{L}ab)$*)*
2. *If* $\mathcal{A}rgs$ *is a complete extension, then* $\texttt{Ext2Lab}(\mathcal{A}rgs)$ *is a complete labelling (where* $\texttt{Ext2Lab}(\mathcal{A}rgs) = (\mathcal{A}rgs, \mathcal{A}rgs^+, Ar\backslash(\mathcal{A}rgs \cup \mathcal{A}rgs^+))$*)*

Moreover, when restricted to complete labellings and complete extensions, the functions `Lab2Ext` *and* `Ext2Lab` *become bijections and each others inverses.*

Theorem 1 implies that complete labellings and complete extensions are one-to-one related. In essence, a complete extension can be seen as the `in`-labelled part of a complete labelling [8, 5].

Before we proceed, we state two propositions that are used in the remaining parts of this paper.

Proposition 1. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. A is on at least one complete extension iff it is in at least one admissible set.*

The validity of Proposition 1 can be seen as follows. Since every complete extension is also an admissible set, it follows that if A is in a complete extension, it is also in an admissible set. Furthermore, if A is in an admissible set, then from [1] it follows that A is also in a preferred extension, and every preferred extension is also a complete extension.

Proposition 2. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. A is in all complete extensions iff A is in the grounded extension.*

The validity of Proposition 2 can be seen as follows. Since the grounded extension is a complete extension, it follows that if an argument is in every complete extension, it is also in the grounded extension. Furthermore, since the grounded extension is the smallest complete extension, it follows that if an argument is in the grounded extension, it is also in every complete extension.

3 Justification Statuses of Arguments

In this section we first define the justification statuses of arguments. Then we provide procedures to determine them.

Definition 6. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. The justification status of A is a function $\mathcal{ML} : Ar \to 2^{\{\texttt{in},\texttt{out},\texttt{undec}\}}$ such that $\mathcal{ML}(A) = \{\mathcal{L}ab(A) \mid \mathcal{L}ab$ is a complete labelling of $AF\}$*

Given the above definition, one would expect there to be 8 (2^3) possible justification statuses, one for each subset of $\{\texttt{in}, \texttt{out}, \texttt{undec}\}$. However two of these subsets turn out not to be possible. First of all, it is not possible for a justification status to be $\emptyset$, because there always exists at least one complete labelling (the grounded labelling). Furthermore, it is also impossible for a justification status to be $\{\texttt{in}, \texttt{out}\}$, because when `in` and `out` are both included in the justification status, then `undec` should also be included, as is stated by the following theorem.

Theorem 2. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. If AF has two complete labellings $\mathcal{L}ab_1$ and $\mathcal{L}ab_2$ such that $\mathcal{L}ab_1(A) = \texttt{in}$ and $\mathcal{L}ab_2(A) = \texttt{out}$ then there exists a complete labelling $\mathcal{L}ab_3$ such that $\mathcal{L}ab_3(A) = \texttt{undec}$.*

Proof. Please refer to [11].

Since $\emptyset$ and $\{\texttt{in}, \texttt{out}\}$ are not possible as justification statuses, there are only 6 possible statuses left to be considered: $\{\texttt{in}\}$, $\{\texttt{out}\}$, $\{\texttt{undec}\}$, $\{\texttt{in}, \texttt{undec}\}$, $\{\texttt{out}, \texttt{undec}\}$ and $\{\texttt{in}, \texttt{out}, \texttt{undec}\}$. We now examine under which conditions these justification statuses occur.

First, we consider the circumstances under which the justification status is $\{\texttt{in}\}$.

Theorem 3. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. Then $\mathcal{ML}(A) = \{\texttt{in}\}$ iff A is in the grounded extension.*

Proof. Please refer to [11].

Next, we consider the circumstances under which the justification status is $\{\texttt{out}\}$.

Theorem 4. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. Then $\mathcal{ML}(A) = \{\texttt{out}\}$ iff A is attacked by the grounded extension.*

Proof. Please refer to [11].

Next, we consider the circumstances under which the justification status is $\{\texttt{undec}\}$.

Theorem 5. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. Then $\mathcal{ML}(A) = \{\texttt{undec}\}$ iff*

1. *A is not in any admissible set and*
2. *A is not attacked by any admissible set*

Proof. Please refer to [11].

Next, we consider the conditions under which the justification status is $\{\texttt{in}, \texttt{undec}\}$.

Theorem 6. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. Then $\mathcal{ML}(A) = \{\texttt{in}, \texttt{undec}\}$ iff*

1. *A is not in the grounded extension,*
2. *A is in an admissible set, and*
3. *A is not attacked by any admissible set.*

Proof. Please refer to [11].

Next, we consider the conditions under which the justification status is $\{\texttt{out}, \texttt{undec}\}$.

Theorem 7. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. Then $\mathcal{ML}(A) = \{\texttt{out}, \texttt{undec}\}$ iff*

1. *A is not in any admissible set,*
2. *A is attacked by an admissible set, and*

3. A is not attacked by the grounded extension.

Proof. Please refer to [11].

Next, we examine the conditions under which the justification status is $\{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$.

Theorem 8. *Let $AF = (Ar, att)$ be an argumentation framework and $A \in Ar$. Then $\mathcal{ML}(A) = \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$ iff*

1. A is in an admissible set
2. A is attacked by an admissible set

Proof. Please refer to [11].

From the above theorems, it follows that membership of an admissible set and membership of the grounded extension, of the argument itself and of its attackers, is sufficient to determine the argument's justification status. The overall procedure of doing so is shown in Figure 2.

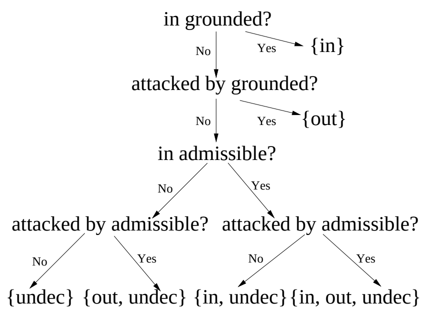


Fig. 2. determining the justification status of an argument

4 Discussion

In this paper, we have presented the justification statuses of arguments which indicate whether an argument has to be accepted, can be accepted, has to be rejected, can be rejected, etc. According to these statuses, we give the methods to justify the statuses of arguments by using discussion games.

We use this labelling based approach for computing the justification statuses of arguments because it yields more informative answers than the traditional extensions approaches.

Take the example in figure 3. Grounded semantics treats all arguments (A, B, C and D) the same (they are not labelled `in` in the grounded labelling).

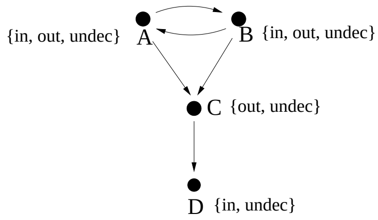


Fig. 3. An example

Credulous preferred semantics treats A, B and D the same (they are labelled `in` in at least one preferred labelling). Sceptical preferred semantics treats A, B and C the same (they are not labelled `in` in some preferred labellings). Also ideal semantics treats all arguments the same (they are not in the ideal extension).

However, our multi-labelling based approach for computing the justification status of an argument allows for a more fine grained distinction between arguments. According to the hierarchy of the justification statuses in figure 4, argument D is the strongest, argument C is the weakest, A and B are in between. Unlike sceptical preferred semantics, our multi-labelling approach does not make D completely justified although it does give it a relatively strong status.

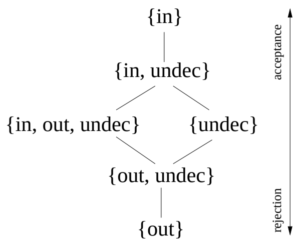


Fig. 4. The hierarchy of justification statuses

The approach of using sets of labels to determine the justification status of an argument is somewhat similar to the approach described in [12]. However, in [12] the authors do not specify a concrete semantics which to apply their approach to, and as a result of this, they do not provide any procedures regarding how to determine the justification status of an argument.

In our current implementation, we have used the discussion game of [13, 14] to determine membership of an admissible set, and the discussion game of [15, 16] to determine membership of the grounded extension. An alternative would be to use the algorithm of [9], which determines both of these memberships in a single pass. Since our notion of justification status depends only on membership of an admissible set and membership of the grounded extension, one is free to apply any kind of algorithm that can determine these.

Acknowledgments I would like to thank Martin Caminada without whose help this paper could not exist. I also wish to thank Mikolaj Podlaszewski who implemented the results of this paper [11].

References

[1] Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artificial Intelligence **77** (1995) 321–357

[2] Caminada, M.: Semi-stable semantics. In Dunne, P., Bench-Capon, T., eds.: Computational Models of Argument; Proceedings of COMMA 2006, IOS Press (2006) 121–130

[3] Dung, P.M., Mancarella, P., Toni, F.: Computing ideal sceptical argumentation. Artificial Intelligence **171**(10-15) (2007) 642–674

[4] Caminada, M.: On the issue of reinstatement in argumentation. Technical Report UU-CS-2006-023, Institute of Information and Computing Sciences, Utrecht University (2006)

[5] Martin Caminada, D.G.: A logical account of formal argumentation. Studia Logica (2009) 347–374

[6] Pollock, J.L.: Cognitive Carpentry. A Blueprint for How to Build a Person. MIT Press, Cambridge, MA (1995)

[7] Jakobovits, H., Vermeir, D.: Robust semantics for argumentation frameworks. Journal of logic and computation **9(2)** (1999) 215–261

[8] Caminada, M.: On the issue of reinstatement in argumentation. In Fischer, M., van der Hoek, W., Konev, B., Lisitsa, A., eds.: Logics in Artificial Intelligence; 10th European Conference, JELIA 2006, Springer (2006) 111–123 LNAI 4160.

[9] Vreeswijk, G.: An algorithm to compute minimally grounded and admissible defence sets in argument systems. In Dunne, P., Bench-Capon, T., eds.: Computational Models of Argument; Proceedings of COMMA 2006, IOS (2006) 109–120

[10] Verheij, B.: A labeling approach to the computation of credulous acceptance in argumentation. In Veloso, M.M., ed.: Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India. (2007) 623–628

[11] Wu, Y., Caminada, M., Podlaszewski, M.: A labelling based justification status of arguments. To appear in 13th Imternational workshop on non-monotonic reasoning, 2010

[12] Baroni, P., Giacomin, M.: Comparing argumentation semantics with respect to skepticism. In: Proc. ECSQARU 2007. (2007) 210–221

[13] Vreeswijk, G.A.W., Prakken, H.: Credulous and sceptical argument games for preferred semantics. In: Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA-00). Number 1919 in Springer Lecture Notes in AI, Berlin, Springer Verlag (2000) 239–253

[14] Caminada, M., Wu, Y.: An argument game for stable semantics. Logic Journal of the IGPL **17**(1) (2009) 77–90

[15] Prakken, H., Sartor, G.: Argument-based extended logic programming with defeasible priorities. Journal of Applied Non-Classical Logics **7** (1997) 25–75

[16] Modgil, S., Caminada, M.W.: Proof theories and algorithms for abstract argumentation frameworks. In Rahwan, I.E., Simari, G.E., eds.: Argumentation in Artificial Intelligence. (2009) 105129

Reasoning about Belief in Social Software using Modal Logic

Ronald de Haan

Utrecht University

Abstract. Social software is the interdisciplinary research program in which social procedures are analyzed and designed using formal, mathematical methods. The analysis of certain procedures requires explicit mention of belief. We develop a logic, based on propositional dynamic logic, that allows us to explicitly reason about belief in social software. We show how this logic can be used by analyzing the example in which a bank can go bankrupt only because of clients' beliefs.

1 Introduction

Social software is an umbrella term for an interdisciplinary research program that uses mathematical tools from computer science and game theory to analyze social procedures. Social procedures are analyzed as if they were computer algorithms. Several logics have been used in the research program (that started with [1]). In this article we will develop a logic to analyze belief in social procedures.

Modeling beliefs with modal logic has been done quite fruitfully using plausibility relations. Plausibility relations indicate what certain agents consider more plausible to be true, and certain notions of belief can be defined in terms of plausibility relations. Certain logics based on propositional dynamic logic (PDL) have been developed for this purpose in [2], [3] and [4].

In this paper we will consider a social situation in which belief plays an important, if not crucial, role. We will develop a PDL-based logic that allows us to analyze belief in social situations. We will use two different types of programs; one for actions that agents can perform, and one for agents' plausibility relations. We will demonstrate that this logic allows us to analyze belief in social situations by using the logic to analyze our example. Finally, we will elaborate on several properties of our logic.

2 Example

Consider the following situation. There is a bank with a number of clients. Every client i can all choose to perform one of two actions: have confidence (i.e., keep their money in the bank), denoted c, or get scared (i.e., take out their savings), denoted s. The bank will go bankrupt if a majority of clients performs s. It is commonly known that every client prefers performing c when the majority of clients does so. Likewise for s. Also, every client prefers a majority of people performing c over a majority performing s. Thus, every client would prefer that everyone perform c. However, if any client k believes that a majority believes that the majority is going to perform s, k would perform s as well. In order to analyze this situation formally, we need to have a means of talking about beliefs in such situations.

3 Framework

In order to analyze beliefs and actions formally, we introduce the following logic. This logic contains two modal operators with PDL programs: one for actions and one for plausibility relations. This approach is similar to the one in [5], the difference being that we use a PDL modality for beliefs.

Definition 1. *An action plausibility model $\boldsymbol{M}$ (or for short: model) for a set of agents Ag and a set of basic propositions $Prop$ is a tuple $\langle W, P, R, C, V\rangle$ where W is a non-empty set of worlds, P is a function that maps each agent i to a relation P_i (a plausibility relation), R is a finite set of relations on W (the action relations), C is a function that maps each world in W to an element in $Ag \cup \{\star\}$ (this coloring is intended as the 'control' of a world; $\star$ is used for terminal worlds), and V is a map from W to $\mathcal{P}(Prop)$ (a map that assigns to each world a $Prop$-valuation).*

P_i is the plausibility relation for i, where $w \to^{P_i} w'$ means that w' is at least as plausible as w. From this relation P_i we can extract a relation P_i^s, where $w \to^{P_i^s} w'$ means that w' is strictly more plausible than w. We don't have to specify P_i^s in our models, since P_i^s can be obtained from P_i by the following definition.

Definition 2. $P_i^s = \{(x,y) \mid x \to^{P_i^*} y \wedge \neg(y \to^{P_i^*} x)\}$*, where P_i^* is the reflexive, transitive closure of P_i.*

Definition 3. *Formulas φ and programs π (beliefs) and α (actions) are defined inductively as follows. Let q range over $Prop$, p over $\{P_j \mid j \in Ag\} \cup \{P_j^s \mid j \in Ag\}$, i over $Ag \cup \{\star\}$ and a over R.*

$$\varphi ::= q \mid c_i \mid \bot \mid \varphi_1 \vee \varphi_2 \mid \neg\varphi \mid [\pi]\varphi \mid [\alpha]\varphi$$

$$\pi ::= p \mid \pi_1;\pi_2 \mid \pi_1 \cup \pi_2 \mid \pi^* \mid \pi^{\smile} \mid ?\varphi$$

$$\alpha ::= a \mid \alpha_1;\alpha_2 \mid \alpha_1 \cup \alpha_2 \mid \alpha^* \mid ?\varphi$$

Abbreviations such as $\varphi_1 \wedge \varphi_2$, $\varphi_1 \to \varphi_2$, $\varphi_1 \leftrightarrow \varphi_2$, $\langle\pi\rangle\varphi$ and $\langle\alpha\rangle\varphi$ are defined as usual.

Definition 4. *Given an action plausibility model $\mathcal{M}$, worlds w, w' in this model and relations π and α, we define the truth of a formula φ inductively.*

1. $\mathcal{M}, w \models q$ *iff* $q \in V(w)$

2. $\mathcal{M}, w \models c_i$ *iff* $C(w) = i$
3. $\mathcal{M}, w \models \varphi_1 \vee \varphi_2$ *iff* $\mathcal{M}, w \models \varphi_1$ *or* $\mathcal{M}, w \models \varphi_2$
4. $\mathcal{M}, w \models \neg\varphi$ *iff* $\mathcal{M}, w \not\models \varphi$
5. $\mathcal{M}, w \models [\pi]\varphi$ *iff in all worlds* w' *such that* $w \rightarrow^{\pi} w'$, $\mathcal{M}, w' \models \varphi$
6. $\mathcal{M}, w \models [\alpha]\varphi$ *iff in all worlds* w' *such that* $w \rightarrow^{\alpha} w'$, $\mathcal{M}, w' \models \varphi$

Definition 5. *Given a basic relation* p *we can define binary relations* π *on* W *in a model* $\mathcal{M}$ *inductively.*

1. $\pi_1;\pi_2 = \{(x, y) \mid \exists z \in W,\ (x, z) \in \pi_1, (z, y) \in \pi_2\}$ *(concatenation)*
2. $\pi_1 \cup \pi_2 = \{(x, y) \mid (x, y) \in \pi_1 \vee (x, y) \in \pi_2\}$ *(union)*
3. $\pi^* = \bigcup_{i=0}^{\infty} \pi^i$ *(transitive, reflexive closure)*
4. $\pi^{\smile} = \{(x, y) \mid (y, x) \in \pi\}$ *(converse)*
5. $?\varphi = \{(x, x) \mid \mathcal{M}, x \models \varphi\}$ *(test)*

With π^0 we denote the relation $?\top$ and with π^{n+1} we denote the relation $\pi^n;\pi$.

Definition 6. *We can define binary relations* α *on* W *in a model* $\mathcal{M}$*, given a basic relation* a*. We define rules for concatenation, union, transitive reflexive closure and test analogously to the rules for relations* π*.*

We use the abbreviation t for the union of all relations in R. We let the abbreviation τ denote $[t]\bot$. This formula holds in all and only in terminal worlds.

3.1 Belief

We can use certain abbreviations for different kinds of epistemic and doxastic relations in our models. The abbreviation for knowledge is taken from [4].

For **knowledge**, we let $\sim_i$ abbreviate $(P_i \cup P_i^{\smile})^*$. According to this definition, every world that is more plausible or less plausible than this world is possible. Knowledge is thus represented by an equivalence relation.

For **weak belief**, we let $>_i^{max}$ abbreviate $P_i^*;?\neg\langle>_i\rangle\top$. This is a relation from a world to the set of maximally plausible worlds (that are reachable from that world). Something is believed weakly if it is true in every most plausible world.

3.2 Constraints

We lay several constraints on our models. These constraints formalize several intuitions about belief in social situations. We rule out situations (i.e., models) that conflict with these intuitions.

One such constraint is that all relations P_i are reflexive. Thus, all agents consider any world at least as plausible as itself. This constraint corresponds to the formula scheme $\varphi \rightarrow \langle P_i\rangle\varphi$, for all agents i.

Another constraint is that agents can distinguish worlds that they control from worlds that they do not control. Formally, for all $w, w' \in W$ if $C(w) \neq$

$C(w')$, then $w \not\to^{P_{C(w)}} w'$. This constraint corresponds to the formula scheme $c_i \to [\sim_i]c_i$, for all agents i.

Another such constraint is the constraint of awareness; i.e. agents are aware what actions they can perform. Formally, for all worlds $w, w' \in W$ and every relation $a \in R$ the following holds: if $C(w) = C(w')$, $w \to^{\sim_{C(w)}} w'$ and $\exists w'' \in W$ such that $w \to^a w''$, then $\exists w''' \in W$ such that $w' \to^a w'''$. Note that $\sim_i$ is an abbreviation. This constraints corresponds to the formula scheme $(c_i \wedge \langle a \rangle \top) \to [\sim_i]\langle a \rangle \top$, for all agents i and all actions a.

Fourthly, we add a constraint of nondeterminism. We have left open the possibility of nondeterministic actions. Certain actions can lead from one world to different other worlds. What we want to state now is that an agent cannot distinguish the different outcomes of a nondeterministic action he performs. Formally, $\forall w, w', w'' \in W$, $\forall a \in R$, if $w \to^a w'$ and $w \to^a w''$, then $w' \to^{\sim_{C(w)}} w''$. This constraint corresponds to the formula scheme $(c_i \wedge \langle a \rangle \varphi) \to [a]\langle P_i \cup (P_i)^\smile \rangle \varphi$, for all agents i and all actions a.

3.3 Axiomatization

This logic can be axiomatized fairly straightforwardly. Proving completeness of this axiomatization can be done with a standard Henkin construction, similarly to the completeness proof in [6]. The formulas given in 3.2 corresponding to the constraints can be used as axioms to enforce the given constraints.

4 Analysis

4.1 Algorithm

Consider the situation from section 2. The model of an instance of this situation with three players can be seen in figure 1. This model corresponds to the extensive form game of the same situation, with payoffs equal to the numbers next to the terminal worlds in figure 1. We will show (with a notion of backward-induction) that all agents will perform s_i, if using weak belief to decide what to do.

Let Act_w denote the set of possible actions that are to be performed at world $w \in W$. Let $T_{\mathcal{M}}$ denote the set of terminal worlds in $\mathcal{M}$: $\{x \mid x \in W,\ \mathcal{M}, x \models \tau\}$, abbreviated T if it is clear what model $\mathcal{M}$ is discussed. We assign a valuation to every terminal world for every player: $u\colon\ T \times Ag \to \mathbb{N}$.

We can define a minimal expected value of an action a in a world w_0 according to the beliefs of agent i that controls w_0 – denoted $E(w_0, a)$ – and the action in a world w that is optimal according to the beliefs of agent i that controls w – denoted $opt(w)$. These two notions are defined in terms of each other. The following algorithm determines $E(w_0, a)$.

- From world w_0 follow $>_i^{max}$ to world w_0'.
- Perform action a in world w_0, leading to world w_1.
- Then, iterate through the following loop until a terminal world is reached.

1. From world w_n follow $>^{max}_{C_{w_n}}$ to world w'_n.
2. Perform any optimal action b for player $C(w'_n)$ (an action $b \in opt(w'_n)$) in world w_n, leading to world w_{n+1}.

– When a terminal world w_t is reached, we obtain $E(w_0, a)$ by obtaining the value of $u(C(w_0), w_t)$.

In certain cases there are multiple possibilities. Sometimes there are more most plausible worlds. In such cases, calculate all possibilities, and take the minimum of all resulting values. Hence the name minimal expected value.

We define $opt(w) = \underset{x \in Act_w}{\operatorname{argmax}} E(w, x)$, where $w \in W - T$. We say an action a is belief-optimal for an agent i in a world w, denoted $\Pi(i, w, a)$, iff $a \in opt(w)$ and $i = C(w)$.

In this algorithm a certain kind of backward induction is used to determine what the optimal action is. For the inductive step the most plausible world is used to determine what action to perform next. However, we use the actual world to determine what the outcome will be. Thus, what actions agents perform is determined by their beliefs and the result of these actions is determined by what is actually the case.

According to these definitions, $\Pi(1, \epsilon, s_1)$, $\Pi(2, s_1, s_2)$ and $\Pi(3, s_1 s_2, s_3)$ all hold in the model Γ^b_S in figure 1. This shows that all agents would perform s_i if using (weak) belief to decide what to do.

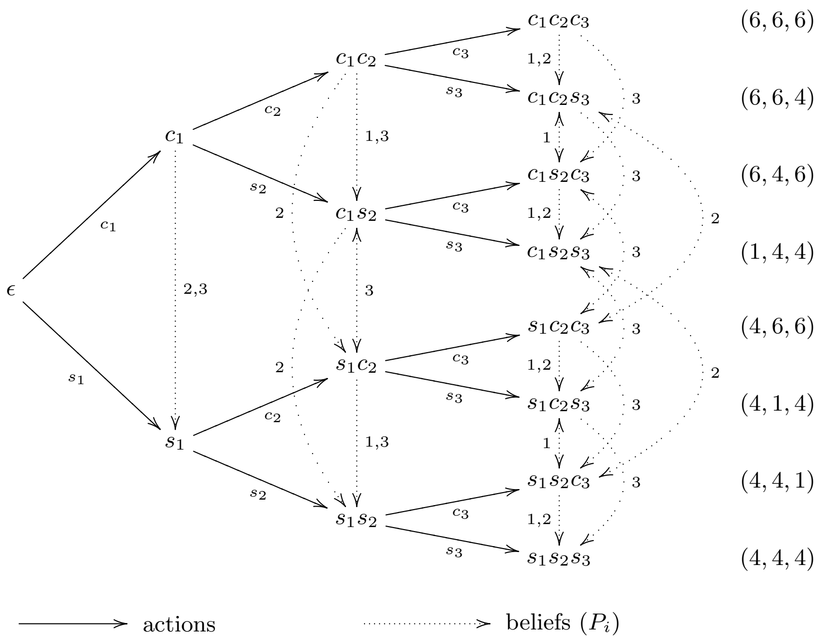
4.2 Logical Analysis

The approach in section 4.1 uses a notion of valuations of terminal worlds that goes beyond the logic we defined in section 3. Also, the algorithm is defined on models without using the logic. However, as we will see in this section, for the finite case we can perform the analysis of section 4.1 using formulas. This approach is similar to the approach in [7] defining backward induction in game logic. The main difference is that in our approach we express all necessary properties of the backward induction in terms of PDL relations and basic propositions, while in the approach in [7] notions such as prediction and utility are made explicit separately in the logic.

We begin with several auxiliary definitions before performing the analysis. In order to simulate valuations on terminal worlds we introduce (a bounded number of) extra atomic propositions v^i_n ($i \in Ag$, $0 \leq n < m$ for a certain $m \in \mathbb{N}$). Let N denote $\{n \in \mathbb{N} \mid 0 \leq n < m\}$. We give these propositions the following interpretation:

$$M, t \models v^i_n \text{ iff } M, t \models \tau \text{ and } u(i, t) \leq n$$

Note that τ is the formula that holds in all and only in terminal worlds. We thus dispose of the valuation u on terminal worlds, and place this information in the valuation of these atomic propositions v^i_n. Namely, it is the case that v^i_n holds in a terminal world iff agent i's valuation of that world is at least n. Also let v^i_m denote $\perp$.



$$C(\epsilon) = 1 \quad C(c_1) = C(s_1) = 2 \quad C(c_1c_2) = C(c_1s_2) = C(s_1c_2) = C(s_1s_2) = 3$$

$$C(c_1c_2c_3) = C(c_1c_2s_3) = C(c_1s_2c_3) = C(c_1s_2s_3) = C(s_1c_2c_3) =$$
$$C(s_1c_2s_3) = C(s_1s_2c_3) = C(s_1s_2s_3) = \star$$

Fig. 1. Model for the bankruptcy game with suspicion (Γ_S^b), including preference values on terminal worlds. Note that the reflexive plausibility relations, which are present in all worlds for all agents, are left out in the figure. The worlds that make b true (denoting 'the bank went bankrupt') are $c_1s_2s_3$, $s_1c_2s_3$, $s_1s_2c_3$ and $s_1s_2s_3$. The utility values are based on the agents' preferences given in the description of the example.

Definition 7. *We let $\bigcup$ denote the union of programs and $\sum$ denote the concatenation of programs.*

We can now define the following recursive pseudo-α-program representing one step in our algorithm. This is the inductive step that represents a belief-optimal move for the player that controls a certain world.

$$\begin{aligned} \mathit{opt} = (?\tau) \cup (?\neg\tau; \bigcup_{i \in Ag} \bigcup_{h \in N} \bigcup_{a_x \in R} (?c_i; ?[>^i_{max}][a^i_x; \mathit{opt}]v^i_h; \\ \sum_{a_y \in R, a_y \neq a_x} ?\neg\langle >^i_{max}\rangle\langle a^i_y; \mathit{opt}\rangle v^i_{h+1}; a^i_x; \mathit{opt})) \end{aligned} \tag{1}$$

In a terminal world the program is a plain reflexive relation. In a non-terminal world the program executes the optimal action. It does this as follows. In the program the union is taken over all agents $i \in Ag$, over all values $h \in N$ and all actions $a_x \in R$. For the agent that controls the current world (determined by $?c_i$) the program determines whether from the world that is most plausible for the agent that controls the current world there is an action for which the optimum (that is determined recursively) results in a certain value v^i_h such that there is no other action for which the optimum results in a higher value (v^i_{h+1}). This action is then executed, and the optimum is determined for the world that is reached by performing this action.

This recursive pseudo-α-program can be unfolded to a complete α-program for the finite case (finite cases are those in which Ag, R and N are finite). Let μ^m denote the unfolding of this pseudo-program for which the depth is bound to m and for which the innermost occurances of *opt* are replaced by $?\tau$. We can obtain a complete program μ^k that has the right interpretation for our purposes by choosing a large enough $k \in \mathbb{N}$. Let μ denote such a μ^k.

We can then define a formula $\mathit{opt}(i, a)$ (interpreted as a is the optimal move for agent i in world w) as follows.

$$\mathit{opt}(i, a) = c_i \wedge \langle a \rangle \top \wedge \bigvee_{h \in N} ([>^i_{max}][a; \mu]v^i_h \wedge \bigwedge_{b \in R, b \neq a} \neg\langle >^i_{max}\rangle\langle b; \mu\rangle v^i_{h+1}) \tag{2}$$

This formula is true in all non-terminal worlds controlled by i in which performing a results (when all agents perform their belief optimal actions) in a certain value h such that performing any other action $b \neq a$ does not result in a higher value.

Using this approach, we see that in model Γ^b_S (figure 1) the following holds: $\epsilon \models \mathit{opt}(1, s_1)$, $s_1 \models \mathit{opt}(2, s_2)$ and $s_1 s_2 \models \mathit{opt}(3, s_3)$. Thus, all players would choose s_i if using (weak) belief to decide what to do

Let model Γ^b_C be model Γ^b_S with reversed beliefs. Then, conversely, in model Γ^b_C the following holds: $\epsilon \models \mathit{opt}(1, c_1)$, $c_1 \models \mathit{opt}(2, c_2)$ and $c_1 c_2 \models \mathit{opt}(3, c_3)$. In this case all players would choose c_i if using (weak) belief to decide what to do.

Also, we can see that in model Γ^b_S, for instance, holds $\epsilon \models [>^1_{max}][(s_1 \cup c_1); \mu]b$, denoting that in world ϵ player 1 believes that (if all players act optimally) whatever action he chooses, the bank will go bankrupt. Conversely, in model Γ^b_C holds $\epsilon \models [>^1_{max}][(s_1 \cup c_1); \mu]\neg b$.

4.3 Advantages and Disadvantages

Our analysis of this situation has several advantages and several disadvantages. We will discuss some of these here.

One advantage of the use of our logic in the analysis of this situation is that we do not have to quantify agents' beliefs, unlike in many game-theoretical models. We only need to state the plausibility relations. This is often easier than quantifying an agent's beliefs.

A disadvantage of the use of our logic in the analysis of this example (or in general in the analysis of situations in which actions are effectively performed simultaneously) is that there is a certain asymmetry in the model with respect to the situation that the agents are in. We model the effectively simultaneous performance of agents' actions as a sequence of decisions in which the agents have no knowledge of what the other agents will do or have done. We only base the analysis of an agent's decision on the beliefs of the agents that are to choose after this agent, not on the beliefs of the agents that chose before this agent. It is, however, possible to work around this asymmetry in our modeling, i.e. we can ensure that the encoding of this information in the belief of this agent is consistent with the analysis based on the beliefs of agents that are to choose. Note that this objection does not apply to situations in which actions are indeed sequential.

5 Conclusion

We developed a logic that makes the explicit analysis of belief in social software possible. This logic is based on PDL. We showed how the example of a bank's going bankrupt due to its clients' beliefs can be analyzed using our logic. Finally, we discussed some advantages and disadvantages of this analysis.

References

1. Parikh, R.: Social software. Synthese **132**(3) (2002) 187–211
2. van Benthem, J., van Eijck, J., Kooi, B.P.: Logics of communication and change. Inf. Comput. **204**(11) (2006) 1620–1662
3. van Eijck, J.: Yet more modal logics of preference change and belief revision. In Apt, K., van Rooij, R., eds.: New Perspectives on Games and Interaction. Amsterdam University Press (2008) 81–104
4. van Eijck, J., Wang, Y.: Propositional dynamic logic as a logic of belief revision. In: WoLLIC '08: Proceedings of the 15th international workshop on Logic, Language, Information and Computation, Berlin, Heidelberg, Springer-Verlag (2008) 136–148
5. Schmidt, R.A., Tishkovsky, D.: On combinations of propositional dynamic logic and doxastic modal logics. J. of Logic, Lang. and Inf. **17**(1) (2008) 109–129
6. Kozen, D., Parikh, R.: An elementary proof of the completeness of PDL. Theoretical Computer Science **14** (1981) 113–118
7. Bonanno, G., Magill, M., Gaasback, K.V.: Branching time logic, perfect information games and backward induction. Working Papers 9813, University of California, Davis, Department of Economics (1998)

2

Logic and Language

Long Papers

A Kripkean Solution to Paradoxes of Denotation

Casper Storm Hansen*

University of Copenhagen
casper_storm_hansen@hotmail.com

Abstract. Kripke's solution to the Liar Paradox and other paradoxes of truth ([3]) is generalized to the paradoxes of denotation. Berry's Paradox and Hilbert and Bernays' Paradox are treated in detail.

1 Introduction

Priest has demonstrated ([4]) that all of the semantical paradoxes share a common structure and has argued, that the solution to this class of paradoxes should therefore also be shared. According to him, this is a reason to reject Kripke's famous solution to the paradoxes of truth ([3]), as it is indeed *only* a solution to these paradoxes and not to the paradoxes of denotation. In this paper I will show that this critique is misplaced. Kripke's solution can be generalized. For reasons of space I will just treat two of the paradoxes of denotation, namely Berry's and Hilbert and Bernays'.

Berry's Paradox ([6]) results from the definite description

> Berry's description: the least integer not describable in fewer than twenty syllables

which is a description of nineteen syllables. So the least integer not describable in fewer than twenty syllables is describable in only nineteen syllables.

Hilbert and Bernays' Paradox (originally presented in [1], natural language formulation in [5]) also results from a definite description, namely this:

> Hilbert and Bernays' description: the sum of 1 and the reference of Hilbert and Bernays' description

If we let n be the reference of Hilbert and Bernays' description, then it also refers to $n+1$. As the reference of a definite description is unique, it follows that $n = n + 1$.

I will assume familiarity with Kripke's paper.

* I would like to express my gratitude to Vincent Hendricks for his encouragement and assistance in the work that lead to this paper, and to the three anonymous reviewers for their valuable comments.

2 Informal Presentation of the Theory

In Kripke's theory sentences become true and false in a recursive process, where a sentence is given a truth value when there is, so to speak, enough information to do so. For instance a sentence of the form "sentence S is true" is made true after it has been decided that S is true, false after it has been decided that S is false, and is left undecided as long as S is. And a disjunction is made true at such time as one of the disjuncts is, since the information about the (eventual) truth value of the other disjunct is irrelevant.

To formulate Berry's description we need two linguistic resources that are not in the formal language of Kripke's paper: The ability to form definite descriptions and a binary predicate expressing that a given term refers to a given object. But when we equip the formal language with these resources, the principle in Kripke's theory can be transfered to these. We let a definite description refer to a given object, when it is determined that this is the unique object which satisfies the description. And if it is decided at some point in the iterative process that there are no objects or more than one object which satisfy the description, it is decided that the definite description fails to refer. And a sentence of the form "T refers to O" is made true, if at some point it is decided that the term T indeed does refer to the object O, and made false, if it is decided that T refers to something different from O or fails to refer.

In Kripke's theory the Liar Sentence "this sentence is false" is neither true nor false, it is "undefined". The reason is that it could only receive a truth value after it itself had received a truth value, so at no point in the iterative process does that happen. When the semantics of definite descriptions and the object-language reference predicate works as described, something similar is the case for Berry's description. Prior to the determination of the reference of Berry's description, the predicate "is an integer not describable in fewer than twenty syllables" is false of a lot of integers, for example 3 and 11 which are the referents of "the square root of 9" and "the number of letters in 'phobophobia'" respectively. But it is not true of any integers, for given any integer for which the predicate is not yet false, it is not yet ruled out that Berry's description might refer to that integer. Ergo the unique object satisfying Berry's description cannot be identified prior to this identification itself, so Berry's description is never assigned a referent and is hence undefined in the fixed point.

In formalizing Hilbert and Bernays' Paradox we will also use definite descriptions and the reference relation. But we need one more thing, namely functions. As is standard, the interpretation of a function symbol will be specified by the interpretation function, and the function symbol can take terms as its arguments. But the value of a function for given arguments may be undetermined for a while in the evaluation process, since it may be undetermined what the terms acting as arguments refer to. We will treat this similarly to the truth functions which constitute the semantics of the connectives and the quantifiers; when there is sufficient information, the function value will be determined. To take an example, consider $f(t_1, t_2, t_3)$ where f is a function symbol and t_1, t_2, and t_3 are terms, and suppose that at some stage in the evaluation process, the reference of t_1 and

t_2 but not t_3 has been determined. Then $f(t_1, t_2, t_3)$ will get a reference at this stage, iff the reference of t_3 does not matter, i.e. if $I(f)(r_1, r_2, d)$, where I is the interpretation function, and r_1 and r_2 are the referents of t_1 and t_2 respectively, has the same value for every value of d.

It is easy to see intuitively that also Hilbert and Bernays' description does not have a reference in the fixed point; a reference of the description cannot be determined prior to this determination itself.

As I plan on showing in a forthcoming longer paper, the Kripkean approach can be used to solve all the known paradoxes of denotation, such as for example the paradoxes of König and Richard. But here I will focus on the paradoxes of Berry and Hilbert and Bernays and present a formal language that has just the resources needed to formalize them.

3 Comments on the Formalization

In Kripke's theory the evaluations at the various levels consist of a set of true sentences and a set of false sentences. The extension of the theory here envisaged means that an evaluation must also contain a reference relation from the set of terms to the domain (supplemented with something to indicate that it has been decided that a given term fails to refer). But it is not necessary to complicate things by making an evaluation a triple. Instead we can take a cue from Frege ([2]) and identify a sentence *being* true/false with the sentence *referring* to Truth/Falsity. That way an evaluation can simply *be* a reference relation – one from the union of the set of sentences and the set of terms to the union of the domain and $\{\top, \bot, *\}$, where $\top$, $\bot$, and $*$ are symbols for Truth, Falsity, and failing to refer respectively.

We will use a standard first-order predicate language with function symbols supplemented with three things: A unary predicate T for "is true", a binary predicate R for "refers to", and a definite description operator: "$\imath v(\phi)$" is to be read as "the v such that ϕ".

4 Syntax

We now turn to the precise specification of the syntax (this section) and semantics (next section) of a formal language. For each $n \in \mathbb{N}$ let there be a countable set $\mathcal{P}_n$ of **ordinary n-ary predicates** and a countable set $\mathcal{F}_n$ of **n-ary function symbols**. In addition there are two **extra-ordinary predicates**, one unary, T, and one binary, R. We also have a set $\mathcal{C}$ of **constants** and a set of **variables**, both of cardinality $\aleph_0$.

The set of **well-formed formulas (wff's)** and the set of **terms** are defined recursively thus:

– Every constant and variable is a term.
– If P is an ordinary n-ary predicate and $t_1, \ldots, t_n$ are terms, then $P(t_1, \ldots, t_n)$ is a wff.

- If ϕ and ψ are wff's, then $\neg\phi$ and $(\phi \wedge \psi)$ are wff's.
- If ϕ is a wff and v a variable, then $\forall v\phi$ is a wff's.
- If t_1 and t_2 are terms, then $T(t_1)$ and $R(t_1, t_2)$ are wff's.
- If ϕ is a wff and v a variable, then $\imath v(\phi)$ is a term.
- If f is an n-ary function symbol and $t_1, \ldots, t_n$ are terms, then $f(t_1, \ldots, t_n)$ is a term.
- Nothing is a wff or term except by virtue of the above clauses.

The connective $\rightarrow$ is used as an abbreviation in the usual way.

Variables, constants, predicates (ordinary as well as extra-ordinary), function symbols, connectives, quantifiers, parenthesis, and commas are called **primitive symbols**.

When ϕ is a wff, v a variable, and c a constant, $\phi(v/c)$ is the wff which is identical with ϕ with the possible exception that all free occurrences of v are replaced with c.

A wff is a **sentence**, and a term is **closed**, if it does not contain any free variables. Let $\mathcal{S}$ and $\mathcal{CT}$ be the set of sentences and the set of closed terms respectively.

We will make use of a notion of **complexity** of a formula, but a precise definition can be dispensed with. Any reasonable definition will do.

5 Semantics

A **model** is defined as a pair $\mathfrak{M} = (D, I)$, where D, the **domain**, and I, the **interpretation function**, satisfy the following:

- D is a superset of $\mathcal{S} \cup \mathcal{CT} \cup \mathbb{N}$ such that
  - $* \notin D$, and
- I is a function defined on $\bigcup_{n \in \mathbb{N}}(\mathcal{P}_n \cup \mathcal{F}_n) \cup \mathcal{C}$ such that
  - for every $P \in \mathcal{P}_n$, $I(P) \subseteq D^n$,
  - for every $f \in \mathcal{F}_n$, $I(f)$ is a function from D^n to D,
  - for every $c \in \mathcal{C}$, $I(c) \in D$, and
  - $I[\mathcal{C}] = D$.

Let a model be fixed for the remainder of this paper. We now define an **evaluation** to be a relation $\mathcal{E}$ on $(\mathcal{S} \cup \mathcal{CT}) \times (D \cup \{\top, \bot, *\})$ such that elements of $\mathcal{S}$ are only related to elements of $\{\top, \bot\}$ and elements of $\mathcal{CT}$ are only related to elements of $D \cup \{*\}$. $\mathcal{E}$ is **consistent** if every sentence and closed term is related by $\mathcal{E}$ to at most one element. An evaluation $\mathcal{E}'$ **extends** $\mathcal{E}$ if $\mathcal{E} \subseteq \mathcal{E}'$.

The semantics is build up in levels as in Kripke's theory. We first specify how to "get from one level to the next": The **evaluation with respect to the evaluation** $\mathcal{E}$, $\mathrm{E}_{\mathcal{E}}$, is defined by recursion on the complexity of the formula[1]:

[1] The clauses make reference to $\mathrm{E}_{\mathcal{E}}$, but only with respect to less complex formulas than the one under consideration. By clause 6 and 7, a formula may "gain" its reference from a more complex formula, but here it is only the relation $\mathcal{E}$ that is used. In short, the reference of a formula only depends on the previous level and formulas of lower complexity. Hence, as stated, the definition is simply by recursion on the complexity of the formula.

1. If t is a constant then $t \mathrm{E}_{\mathcal{E}} I(t)$.
2. If s is of the form $P(t_1, \ldots, t_n)$ where P is an ordinary n-ary predicate and $t_1, \ldots, t_n$ are closed terms, then
   - $s \mathrm{E}_{\mathcal{E}} \top$ if there are $d_1, \ldots, d_n \in D$ satisfying $t_1 \mathrm{E}_{\mathcal{E}} d_1, \ldots, t_n \mathrm{E}_{\mathcal{E}} d_n$ such that $(d_1, \ldots, d_n) \in I(P)$, and
   - $s \mathrm{E}_{\mathcal{E}} \perp$ if there are $d_1, \ldots, d_n \in D$ satisfying $t_1 \mathrm{E}_{\mathcal{E}} d_1, \ldots, t_n \mathrm{E}_{\mathcal{E}} d_n$ such that $(d_1, \ldots, d_n) \notin I(P)$.
3. If s is of the form $\neg\phi$ where ϕ is a sentence, then
   - $s \mathrm{E}_{\mathcal{E}} \top$ if $\phi \mathrm{E}_{\mathcal{E}} \perp$, and
   - $s \mathrm{E}_{\mathcal{E}} \perp$ if $\phi \mathrm{E}_{\mathcal{E}} \top$.
4. If s is of the form $(\phi \wedge \psi)$ where ϕ and ψ are sentences, then
   - $s \mathrm{E}_{\mathcal{E}} \top$ if $\phi \mathrm{E}_{\mathcal{E}} \top$ and $\psi \mathrm{E}_{\mathcal{E}} \top$, and
   - $s \mathrm{E}_{\mathcal{E}} \perp$ if $\phi \mathrm{E}_{\mathcal{E}} \perp$ or $\psi \mathrm{E}_{\mathcal{E}} \perp$.
5. If s is of the form $\forall v\phi$ where v a variable and ϕ is a wff with at most v free, then
   - $s \mathrm{E}_{\mathcal{E}} \top$ if for all $c \in \mathcal{C}$, $\phi(v/c) \mathrm{E}_{\mathcal{E}} \top$, and
   - $s \mathrm{E}_{\mathcal{E}} \perp$ if there exists a $c \in \mathcal{C}$ such that $\phi(v/c) \mathrm{E}_{\mathcal{E}} \perp$.
6. If s is of the form $T(t)$ where t is a closed term, then
   - $s \mathrm{E}_{\mathcal{E}} \top$ if there is a $s' \in \mathcal{S}$ such that $t \mathrm{E}_{\mathcal{E}} s'$ and $s'\mathcal{E}\top$,
   - $s \mathrm{E}_{\mathcal{E}} \perp$ if there is a $s' \in \mathcal{S}$ such that $t \mathrm{E}_{\mathcal{E}} s'$ and $s'\mathcal{E}\perp$, and
   - $s \mathrm{E}_{\mathcal{E}} \perp$ if there is a $d \in D$ such that $t \mathrm{E}_{\mathcal{E}} d$, but no $s' \in \mathcal{S}$ such that $t \mathrm{E}_{\mathcal{E}} s'$.
7. If s is of the form $R(t_1, t_2)$ where t_1 and t_2 are closed terms, then
   - $s \mathrm{E}_{\mathcal{E}} \top$ if there is a $d \in D \cup \{*\}$ and a closed term t_1' such that $t_1 \mathrm{E}_{\mathcal{E}} t_1'$, $t_1'\mathcal{E}d$ and $t_2 \mathrm{E}_{\mathcal{E}} d$,
   - $s \mathrm{E}_{\mathcal{E}} \perp$ if there are $d_1, d_2 \in D \cup \{*\}$, such that $d_1 \neq d_2$, and a closed term t_1' such that $t_1 \mathrm{E}_{\mathcal{E}} t_1'$, $t_1'\mathcal{E}d_1$ and $t_2 \mathrm{E}_{\mathcal{E}} d_2$, and
   - $s \mathrm{E}_{\mathcal{E}} \perp$ if there is a $d' \in D \cup \{*\}$ such that $t_1 \mathrm{E}_{\mathcal{E}} d'$, but no closed term t_1' such that $t_1 \mathrm{E}_{\mathcal{E}} t_1'$.
8. If t is of the form $\imath v(\phi)$ where v is a variable and ϕ is a wff with at most v free, then
   - $t \mathrm{E}_{\mathcal{E}} d$ if d is an element of D such that for some $c \in \mathcal{C}$, $I(c) = d$ and $\phi(v/c) \mathrm{E}_{\mathcal{E}} \top$ and for all other elements d' of D, every $c' \in \mathcal{C}$, such that $I(c') = d'$, satisfies $\phi(v/c') \mathrm{E}_{\mathcal{E}} \perp$,
   - $t \mathrm{E}_{\mathcal{E}} *$ if there are two different elements d_1 and d_2 of D such that for some $c_1, c_2 \in \mathcal{C}$, $I(c_1) = d_1$, $I(c_2) = d_2$, $\phi(v/c_1) \mathrm{E}_{\mathcal{E}} \top$ and $\phi(v/c_2) \mathrm{E}_{\mathcal{E}} \top$, and
   - $t \mathrm{E}_{\mathcal{E}} *$ if for all elements d of D, there is a $c \in \mathcal{C}$ such that $I(c) = d$ and $\phi(v/c) \mathrm{E}_{\mathcal{E}} \perp$.
9. If t is of the form $f(t_1, \ldots, t_n)$ where f is a n-ary function symbol and $t_1, \ldots, t_n$ are closed terms, then $t \mathrm{E}_{\mathcal{E}} d$ if d is an element of D for which every n-tuple $(d_1, \ldots, d_n)$ such that for each $i \in \{1, \ldots, n\}$ either $t_i \mathrm{E}_{\mathcal{E}} d_i$ or t_i is not related to anything by $\mathrm{E}_{\mathcal{E}}$, satisfy $I(f)(d_1, \ldots, d_n) = d$.

Now we iterate the process by defining for all ordinals α the **evaluation with respect to the level** α, written E^α, by recursion:

$$\mathrm{E}^\alpha = \begin{cases} \emptyset & \text{if } \alpha = 0 \\ \mathrm{E}_{\mathrm{E}^{\alpha-1}} & \text{if } \alpha \text{ is a successor ordinal} \\ \bigcup_{\eta<\alpha} \mathrm{E}^\eta & \text{if } \alpha \text{ is a limit ordinal} \neq 0 \end{cases}$$

The following two lemmas show that the process is monotonic and does not result in any inconsistency:

Lemma 1. *For all ordinals α, β, if $\alpha < \beta$ then* $\mathrm{E}^\alpha \subseteq \mathrm{E}^\beta$.

Proof. By induction on the complexity of formulas it is seen that for each bullet in each of the nine clauses above, if the condition in that bullet is satisfied for some evaluation $\mathcal{E}$ it is also satisfied for every extension of $\mathcal{E}$. Ergo if $\mathcal{E} \subseteq \mathcal{E}'$ then $\mathrm{E}_\mathcal{E} \subseteq \mathrm{E}_{\mathcal{E}'}$. As it also holds that $\mathrm{E}^0 = \emptyset$ is a subset of every evaluation, the lemma follows. □

Lemma 2. *For every ordinal α,* E^α *is consistent.*

Proof. By outer induction on α and inner induction on the complexity of formulas, considering clause 1–9. □

For every ordinal α and every $x \in \mathcal{S} \cup \mathcal{CT}$ we define $[\![x]\!]^\alpha$ to be the unique y such that $x\,\mathrm{E}^\alpha\,y$, when there is a such. We say that x is **determined at level** α, if α is the first level where $[\![x]\!]^\alpha$ is defined.

We now come to the important fixed point theorem:

Theorem 1. *There is a unique consistent evaluation $\mathcal{E}$ such that for some ordinal α it holds that for all ordinals $\beta > \alpha$, $\mathrm{E}^\beta = \mathcal{E}$.*

Proof. As there are only countable many sentences and closed terms, the monotonic process must reach a fixed point. Consistency of the fixed point follows from lemma 2. □

Letting $\mathcal{E}$ and α be as in the theorem, we define the **evaluation**, E, as $\mathcal{E}$, and for all $x \in \mathcal{S} \cup \mathcal{CT}$ set $[\![x]\!]$ equal to $[\![x]\!]^\alpha$ when this is defined. The value of $[\![x]\!]$ is to be thought of as *the* reference of x.

6 Expressibility of the Reference Relation

Kripke's theory is famous for validating the Tarskian T-schema in the sense that, if (in the notation of this paper) s is a sentence and c is a constant such that $I(c) = s$, then $[\![s]\!] = \top$ if and only if $[\![T(c)]\!] = \top$. In other words: If a sentence is true, this can be expressed in the object language. In this theory a similar result holds for reference; if a closed term refers to a given object, then this can be expressed in the language itself. That is the content of the following theorem.

Theorem 2. *Let t be a closed term, d an element of D, and c_1 and c_2 constants such that $I(c_1) = t$ and $I(c_2) = d$. The following biimplication holds: $[\![t]\!] = d$ iff $[\![R(c_1, c_2)]\!] = \top$.*

Proof. From clause 1 it is seen that for all ordinals α we have $c_1 \mathrm{E}^\alpha t$ and $c_2 \mathrm{E}^\alpha d$. So it follows from bullet 1 of clause 7 that $[\![t]\!] = d$ iff $t \mathrm{E}^\beta d$ for some ordinal β iff $R(c_1, c_2) \mathrm{E}^{\beta+1} \top$ iff $[\![R(c_1, c_2)]\!] = \top$. □

7 Solution to Berry's Paradox

In formalizing the Berry Description we have to get around the fact that in the formal language, any natural number can be defined with a definite description of just one symbol, namely a constant. We can do this by defining "length of a term" not in the obvious way as the number of primitive symbols in the term, but slightly differently. Reflecting the fact that in natural languages there are only finitely many primitive symbols, let Φ be a function from the set of primitive symbols of our formal language to $\mathbb{N}$ which sends only a finite number of primitive symbols to each $n \in \mathbb{N}$. Then define the length of a term to be the sum of $\Phi(x)$ for every occurrence x of a primitive symbol in the term.

Now we can formalize the Berry Description. Let n, m, and x be variables and let N and L be unary predicates and $\geq$ a binary predicate, such that $I(N)$ is the set of natural numbers, and $I(\geq)$ is the relation "larger than or equal to" on the set of natural numbers. L is to be interpreted as "long", but we postpone the precise specification of $I(L)$, until we know just what "long" should mean to make our formalization "paradoxical".

We can formalize "x is a definite description of the natural number n" thus:

$$N(n) \wedge R(x, n)$$

So "The natural number n does not have a short definite description" can be formalized

$$N(n) \wedge \forall x(R(x, n) \to L(x)) \ ,$$

and "n is the least natural number that does not have a short definite description"

$$\begin{aligned}&(N(n) \wedge \forall x(R(x, n) \to L(x))) \wedge \\ &\qquad \forall m((N(m) \wedge \forall x(R(x, m) \to L(x))) \to \geq(m, n)) \ .\end{aligned}$$

Ergo, Berry's description in a version with length of formal expressions instead of number of syllables, "the least natural number that does not have a short definite description", can be formalized as (B):

$$\begin{aligned}&\iota n((N(n) \wedge \forall x(R(x, n) \to L(x))) \wedge \\ &\qquad \forall m((N(m) \wedge \forall x(R(x, m) \to L(x))) \to \geq(m, n))) \quad \text{(B)}\end{aligned}$$

Now we can set $I(L)$ to be the set of terms which are longer than the length of (B).

That (B) fails to refer, i.e that there is no $d \in D$ such that $[\![\text{(B)}]\!] = d$, is proved as follows: Assume ad absurdum that there is such a $d \in D$. Then it follows by clause 8 that for a constant c with $I(c) = d$, we have

$$[\![(N(c) \wedge \forall x(R(x,c) \to L(x))) \wedge \\ \forall m((N(m) \wedge \forall x(R(x,m) \to L(x))) \to \geq(m,c))]\!] = \top \ .$$

Using clause 4 twice it can be inferred that

$$[\![\forall x(R(x,c) \to L(x))]\!] = \top \ ,$$

and consequently by clause 5 that

$$[\![R(c',c) \to L(c')]\!] = \top \ ,$$

where c' is a constant such that $I(c') = (B)$. It is already determined at level 1, that $L(c')$ is false. This follows from the specification of $I(L)$. Ergo we must have $[\![R(c',c)]\!] = \bot$. So at some level bullet 2 or 3 of clause 7 is satisfied. But bullet 3 can not be, for c' refers to (B) and since the referent of a constant is unique, not to some object which is not a term. And bullet 2 can not be either, for then (B) would have to refer to something different from d, but by assumption this is not the case. This is a contradiction.

8 Solution to Hilbert and Bernays' Paradox

The Hilbert and Bernays description can be formalized

$$+(\bar{1}, \imath v(R(h,v))) \ , \tag{HB1}$$

where v is a variable, h is a constant such that $I(h) = \text{(HB1)}$, and $+$ is a binary function symbol such that $I(+)$ is the function that sends every pair of numbers to their sum and every other pair to 0. $\bar{1}$ is a numeral for 1.

$[\![\text{(HB1)}]\!]$ is undefined, as we will proceed to prove. As the sum of 1 and n is not the same for every natural number n, (HB1) will get a reference, only if

$$\imath v(R(h,v)) \tag{HB2}$$

gets a reference (clause 9). By clause 8 this happens only if there is a constant c such that

$$R(h,c) \tag{HB3}$$

is related to $\top$. We have $h \, \mathrm{E}_{\emptyset} \text{(HB1)}$ from which it follows by bullet 1 of clause 7 that this can only be the case if (HB1) gets a reference. We have come full circle, and can conclude that neither (HB1), (HB2), nor (HB3) become related to anything.

References

1. Bernays, P. and D. Hilbert: Grundlagen der Mathematik (zweiter Band). Verlag von Julius Springer (1939)
2. Frege, G.: Über Sinn und Bedeutung. Zeitschrift fur Philosophie und philosophische Kritik 100, 26–60 (1892)
3. Kripke, S.: Outline of a theory of truth. The Journal of Philosophy 72, 690–716 (1975)
4. Priest, G.: Beyond the limits of thought (second ed.). Oxford University Press (2002)
5. Priest, G.: The paradoxes of denotation. In T. Bolander, V. F. Hendricks and S. A. Pedersen (eds.): Self-reference. CSLI Publications (2006)
6. Russell, B.: Mathematical logic as based on the theory of types. American Journal of Mathematics 30, 222–262 (1908)

Donkey Readings and Delayed Quantification

Mike Solomon

New York University

Abstract. Kanazawa (1994) proposes that which reading a donkey sentence receives is a consequence of the logical properties of its quantificational determiner. However, this analysis fails to account for weak readings with quantifiers like *every*. The analysis of Brasoveanu (2007), which places responsibility in the indefinite determiner, accounts for these cases, but overgenerates. I propose an extension of Kanazawa's system with a mechanism for delaying witness selection, which accounts for weak readings with *every* in the attested cases. This analysis accounts for the novel observation that existential quantification in these constructions always takes lowest scope.

1 Introduction

Two kinds of interpretations of donkey sentences have been dinstinguished in the literature: the *strong reading* and the *weak reading*. These are schematically paraphrased below, for Q a quantificational determiner.

(1) Q farmer who owns a donkey beats it.
(2) *Strong reading:*
Q farmer who owns a donkey beats every donkey he owns.
(3) *Weak reading:*
Q farmer who owns a donkey beats a donkey he owns.

The most natural interpretation of the classic example (4) is the strong reading, which is paraphrased as (5).

(4) Every farmer who owns a donkey beats it.
(5) Every farmer who owns a donkey beats every donkey he owns.

The natural interpretation of (6), on the other hand, is given by the weak reading, paraphrased as (7).

(6) No farmer who owns a donkey beats it.
(7) No farmer who owns a donkey beats a donkey he owns.

What determines which reading a given donkey sentence receives? I first discuss the approaches to this question of Kanazawa (1994) and Brasoveanu (2007), before turning to my own proposal, an extension of Kanazawa's.

2 Kanazawa 1994

The only evident difference between (4) and (6) is the identity of the quantificational determiner: *every* in (4), *no* in (6). It would therefore be natural to attribute the reading each sentence receives to its determiner. Kanazawa (1994) notices that in general, a given determiner selects the same reading across donkey sentences: *every* and *not every* select the strong reading; *some*, *no*, *several*, and modified numerals like *at least two* and *at most five* select the weak reading.

Kanazawa proposes that the quantificational determiner is responsible for which reading a donkey sentence receives. Which reading the determiner selects depends on the determiner's monotonicity properties. Kanazawa's generalization is that determiners which are left and right monotone in opposite directions select the strong reading, while determiners which are monotone in the same direction select the weak reading. This generalization matches the above observation: *every* is $\downarrow$MON$\uparrow$, *not every* $\uparrow$MON$\downarrow$, *some* $\uparrow$MON$\uparrow$, and *no* $\downarrow$MON$\downarrow$.

How does the monotonicity of the determiner determine which reading it selects in a donkey sentence? The idea is that monotonicity reflects how we reason about quantifiers; if we reason about quantifiers in donkey sentences the same way we do in the ordinary case, the above behavior is what results. In particular, the monotonicity properties of a quantificational determiner reflect which patterns of "inference from submodels" it licenses.

For instance, *every*$(A)(B)$ is falsified by a single counterexample: an A that is not a B. Therefore if *every*$(A)(B)$ is false in a submodel of the model under discussion, it is false in the model. Now consider (4), and suppose there is a farmer, f, who owns a donkey that he does not beat, d. If we restrict our attention to the submodel in which f is the only farmer and d is the only donkey, then (4) is false, on any construal. Applying our knowledge of *every*, we conclude that (4) must be false in the original model. Therefore (4) requires that every farmer beat every donkey he owns, that is, has the strong reading. The preservation of falsity from submodel to model is the characteristic inference pattern of $\downarrow$MON$\uparrow$ determiners. Similar reasoning derives the observed readings for the other three classes of determiners.

Kanazawa formalizes his proposal in DPL (Groenendijk & Stokhof 1991). The crucial feature of DPL for an analysis of donkey anaphora is that it allows a quantifer to semantically bind a variable outside its syntactic scope. DPL as originally formulated only includes the two first-order quantifiers, and so must be extended with generalized quantifiers.

Kanazawa first extends DPL with static generalized quantifiers: for each generalized quantifier symbol Q, formulas of form $Q(\phi, \psi)$ are added to the language, with the obvious interpretation. These formulas are internally static: no binding is possible between ϕ and ψ. To account for donkey anaphora as dynamic binding, He then introduces dynamic generalized quantifiers to DPL. It is possible to define an internally dynamic generalized quantifier $\mathcal{Q}$ in terms of a static quantifier Q and the internally dynamic connectives of DPL in two ways, corresponding to the two readings of donkey sentences. The quantifier $\mathcal{Q}_W$ corresponds to the weak reading of a donkey sentence with quantifier Q; $\mathcal{Q}_S$ corresponds to the strong reading.

(8) $\mathcal{Q}_W x(\phi, \psi) \quad \leftrightarrow \quad Qx(\phi, \phi \,;\, \psi)$

(9) $\mathcal{Q}_S x(\phi, \psi) \quad \leftrightarrow \quad Qx(\phi, \phi \Rightarrow \psi)$

I follow the presentation of DPL in Kanazawa 1994, which introduces the symbols ';', '$\Rightarrow$', and '$\mathcal{E}$' for dynamic conjunction, dynamic implication, and the dynamic existential quantifier.

With these dynamic quantifiers, (4) and (6) then receive the translations in (10) and (11), which yield the desired truth conditions.

(10) Every farmer who owns a donkey beats it.
$\mathcal{EVERY}_S x(\text{farmer}(x) \,;\, \mathcal{E}y(\text{donkey}(y) \,;\, \text{own}(x, y)), \text{beat}(x, y))$

(11) No farmer who owns a donkey beats it.
$\mathcal{NO}_W x(\text{farmer}(x) \,;\, \mathcal{E}y(\text{donkey}(y) \,;\, \text{own}(x, y)), \text{beat}(x, y))$

For each static generalized quantifier, dynamic generalized quantifiers corresponding to each of the weak and the strong readings can be defined. Why is it that each static quantifier seems to select a single dynamic version? Kanazawa defines a dynamic version of monotonicity, and shows that for each doubly monotone static generalized quantifier, only one of the two corresponding dynamic versions is doubly dynamically monotone. These versions are the ones that are selected. The dynamic behavior of a quantifier then follows from the static quantifier's monotonicity properties, which reflect how we reason about the quantifier.

3 Issues with the determiner account

Unfortunately, Kanazawa's generalization is not without exception. It has been known since Schubert & Pelletier 1989 that a donkey sentence with *every* can receive a weak reading:

(12) Every man who had a quarter put it in the parking meter.
(13) Every guest who had a credit card used it to pay his bill.

Consider (12). This sentence is most naturally read to require that every man who had a quarter put at least one of his quarters in the parking meter, not the bizarre circumstance that every man who had a quarter put every single one of his quarters in the parking meter.

If two donkey sentences with the same quantificational determiner receive different readings, how can the determiner be responsible for selecting the reading? Kanazawa acknowledges this point, and concludes that *every* and *not every* do not decisively select the strong reading for a sentence, but only bias the sentence toward this interpretation. In the face of strong pragmatic pressure toward the weak reading, this semantic bias can be overridden.

Certainly pragmatics plays an important role in the interpretation of these sentences, but this response is unsastisfying. First, it seems to force a less attractive interpretation of Kanazawa's account, on which the dynamic behavior of a quantifier does not *follow* from its ordinary inferential behavior, but rather on which monotonicity is just one factor among potentially many that bears on the *choice* of which of two dynamic versions of a quantifier to select.

Second, Kanazawa holds that ↑MON↑ and ↓MON↓ determiners categorically select weak readings in a way that cannot be overridden by pragmatics. That is, no donkey sentence with *some* or *no* can ever receive a strong reading. But there is no explanation why this class of determiners should be resistant to pragmatic pressures where the other is not. On the other hand, this claim still provides evidence that there is a necessary semantic connection between the quantificational determiner in a donkey sentence and the reading it receives.

However, Brasoveanu (2007) argues that even this is not the case. He observes that when the donkey pronoun, falls within the scope of "nuclear scope negation", the sentence must receive the strong reading, no matter the determiner (this observation goes back to Lappin & Francez 1994):

(14) No guest who had a credit card failed to use it.
(15) Every man who had a quarter refused to put it in the meter.
(16) At least one man who had a quarter forgot to put it in the meter.

Consider (14). On the weak reading, the sentence means that no guest had a credit that he failed to use. But this is equivalent to saying that every guest used every credit card he had. This is absurd, and not what (14) says. Instead, (14) means that every guest who had a credit card used at least one of his credit cards; in other words, for no man who had a credit card was it the case that for every credit card he had, he failed to use it. That is, (14) receives the strong reading, despite its determiner being *no*, which Kanazawa predicts should obligatorily select the weak reading.

Therefore, Brasoveanu concludes, it cannot be the identity of the quantificational determiner that is responsible for selecting the reading of a donkey sentence.

4 Brasoveanu 2007

Brasoveanu argues that every quantificational determiner is compatible with both readings. One possible response to this would be to maintain that the quantifier is responsible for the reading, but that every static quantifier Q is simply ambiguous between $\mathcal{Q}_W$ and $\mathcal{Q}_S$. Brasoveanu rejects this analysis on the basis of "mixed" readings of sentenes with multiple instances of donkey anaphora:

(17) Every man who bought a book online and had a credit card used it to pay for it.

This sentence is most naturally read to make a claim about every book that every man buys, but not to require that any man use every one of his credit cards to pay for a single book. That is, the donkey anaphora introduced by *a book* is read strong, while that introduced by *a credit card* is read weak.

Whether it is read strong or weak is independent for each instance of donkey anaphora, and so cannot be attributed to the single quantifier. Brasoveanu proposes to attribute the availability of the two readings to an ambiguity or underspecification of the indefinite determiner. The weak version of the indefinite determiner yields the weak reading; the strong version yields the strong reading. In which of these two ways a given token of the ambiguous *a* is interpreted is a matter of pragmatics. The nature of the quantifier plays no semantic role in determining the reading of a donkey sentence.

On Brasoveanu's analysis, the strong and weak readings of a donkey sentence are semantically distinct, but the choice between them is ultimately pragmatic. This type of pragmatic account makes the prediction that while a donkey sentence might favor one of the two readings, the other reading should surface if the preferred reading is excluded by the discourse. But this prediction is not borne out (see also (51) in Kanazawa 1994 for a similar example):

(18) Farmer John beats some but not all of the donkeys he owns.
$^{\#}$In fact, every farmer who owns a donkey beats it.

The first sentence of (18) contradicts the strong reading of the second sentence. Hence we would expect it to receive the weak reading. Instead, it is simply infelicitous.

Where Kanazawa's theory undergenerates in failing to account for weak readings of donkey sentences with ↓MON↑ and ↑MON↓ determiners, Brasoveanu's analysis overgenerates, predicting readings that contradict Kanazawa's generalization where they do not exist. In fact, these exceptional readings are only available in a limited class of environments, illustrated in the examples concerning putting a quarter in a parking meter and using a credit card for payment. What seems to distinguish this class is that, for instance, (13) is not about credit cards in the same way that (4) is about donkeys. (4) is an answer to the question "How do farmers treat their donkeys?"; (13) is an answer to the question "How did the guests pay their bills?", and not "What did the guests do with their credit cards?" Weak readings with ↓MON↑ or ↑MON↓ determiners are only possible in these kinds of cases.

It is in these same kinds of cases that nuclear scope negation sentences with ↑MON↑ or ↓MON↓ determiners appear to receive strong readings. That is, it is only in these cases that a donkey sentence can receive a reading other than the one predicted by Kanazawa's monotonicity account. What exactly is this reading? Brasoveanu claims that (16) receives the strong reading. But this is not quite right. The sentence is interpreted to mean that at least one man who had a quarter intended to put one in the parking meter but failed to do so. This is not the weak reading of the sentence: it is incompatible with every man putting a quarter in the meter even if at least one man forgot to put some other particular quarter in the meter. But it is not the strong reading either. The strong reading requires that there be a man who had a quarter such that for each of his quarters, he forgot to put *that particular quarter* in the meter. This in turn requires that for each quarter he intended to put that particular quarter in the meter, which is to say, that he intended to put *all* of his quarters in the meter. The natural interpretation of (16) has no such requirement. The strong reading entails too many intentions.

5 Proposal: delayed quantification

What then is the correct analysis of sentences like (16), if it is neither the strong nor the weak reading? A clue is given by its paraphrase in (19):

(19) At least one man who had a quarter forgot to put a quarter in the meter.

On the natural interpretation of (19), the indefinite *a quarter* in the nuclear scope scopes below the intensional verb *forgot*. But on the standard implementation of the weak reading, the indefinite takes scope above the verb phrase, which we saw produces incorrect results. If *forget* were not an intensional verb, but simple negation, the $\exists > \neg$ scoping produced by the strong reading would be equivalent to the desired $\neg > \exists$ scoping, but with *forget* the two are not equivalent. The correct analysis of (16) on its reading paraphrased in (19) must then involve an existential scoping below *forgot* – at the position of the donkey indefinite, and therefore can be neither of the readings we have considered so far.

What do we know at this point? In general, donkey sentences with ↓MON↑ and ↑MON↓ determiners receive strong – universally-quantified – readings. In exceptional cases, these sentences receive existentially-quantified readings. Donkey sentences with ↑MON↑ and ↓MON↓ determiners never receive strong readings. In general, they receive existentially quantified readings – the weak readings of Kanazawa and Brasoveanu. In exceptional cases – the same class of exceptional cases – these sentences receive readings with *lower scoping* existential quantification.

Quantificational determiners in general select the readings predicted by Kanazawa's account. In a class of exceptional cases, donkey sentences receive different, existentially-quantified readings. Are the exceptional existential readings the same for both kinds of quantifiers? That is, when a donkey sentence with *every* receives a "weak" reading, does the existential quantification take lowest scope?

(20) Every man who had a nice suit refused to wear it to the town meeting.

This sentence can be interpreted in two ways. It can mean that for every man who had a nice suit, it was the case that for each of his nice suits it was suggested to him that he wear that particular suit, and for each of his nice suits he refused. This is the strong reading. It can also, perhaps more naturally, mean that every man who had a nice suit refused to wear a nice suit to the town meeting – it was suggested to him that he wear a nice suit, and he refused. It *cannot* mean that every man who had a nice suit rejected at least one nice suit, but that some men may have still worn nice suits to the town meeting. So the existential quantification must take lowest scope, beneath the negative intensional verb, just as in the ↑MON↑ and ↓MON↓ cases. Thus the exceptional existential reading is the same across all quantifiers. It consists in the donkey pronoun being interpreted as an existential taking narrowest scope.

I prosose to extend Kanazawa's analysis with an operator to delay witness selection. The operator, when applied to a dynamic formula, collapses the set of its output assignments into a single assignment, assigning possibly non-atomic values to variables. These non-atomic values are existentially quantified over at the point of evaluation of the lexical relation which takes the variable as an argument. This has the effect of the variable being interpreted as an existential that takes lowest scope.

Formally, first define a formal sum operation '+' on the domain of individuals, so that if a and b are individuals, $a+b$ is a non-atomic individual. Note that this operation should not be interpreted as a mereological sum (there is no collective intepretation). Now for assignment functions g and h, define $g+h$ to be the assignment function such that for all variables x, $(g+h)(x) = g(x)+h(x)$. Write ΣS for the sum of the elements of a set S, and write '$\leq$' for the part-of relation for this sum. We can now define the delay operator '$\sqcup$' as follows:

(21) $[\![\sqcup\phi]\!] = \{\langle g,h\rangle \mid h = \Sigma\{k \mid \langle g,k\rangle \in [\![\phi]\!]\}\}$

Thus $\sqcup\phi$ relates to each input assignment g the single assignment function that is the formal sum of all of the output assignments related to g by ϕ. For any formula ϕ, ϕ and $\sqcup\phi$ are truth-conditionally equivalent, but not usually dynamically equivalent. Now we need these non-atomic individuals to be interpreted as existential quantification when they are evaluated as arguments of lexical relations. To do so we modify the semantics of lexical relations:

(22) $[\![R(t_1,\ldots,t_n)]\!] = \{\langle g,h\rangle \mid h \leq g \wedge \langle [\![t_1]\!]_h,\ldots,[\![t_n]\!]_h\rangle \in I(R)\}$

where $[\![t]\!]_h$ is the interpretation of term t with respect to h, and $I(R)$ is the interpretation of the relation symbol R. Note that atomic formulas of this form are no longer tests: their evaluation has the effect of "splitting" non-atomic individuals into atomic ones split across assignments. So (23) remains illicit: once the pronoun fixes its referent, it must refer to that individual from then on.

(23) John had a quarter. He put it$_i$ in the parking meter. $^{\#}$It$_j$'s in his pocket.

Although $\sqcup$ appears to allow *it* to act like *one* in cases of donkey anaphora, it is important that it does not do so generally.

(24) $^{\#}$John has a credit card. Mary has it too.

The non-atomic individual introduced by *a credit card* is the sum of the credit cards that satisfy the entire sentence, not the sum of all credit cards. This restricts *it* to credit cards that *John has*, thereby rendering (24) anomalous.

Intuitively, the multiple output assignments of an externally dynamic formula ϕ represent the possible ways of making ϕ true. The formula $\sqcup\phi$ collapses all of these possibilities into a single possibility; it disregards the differences between the ways of making ϕ true. If ϕ is an existential, each output assignment encodes a possible witness. Then the output assignment of $\sqcup\phi$ encodes that ϕ has a witness – it introduces a discourse referent – but it ignores the particular identity of the witness. $\sqcup$ *delays* the selection of a witness until it is forced by the evaluation of a lexical relation, at which point existential quantification applies. It is in just the kinds of contexts where this is appropriate – where $\sqcup$ can be used felicitously – that exceptional weak readings are possible.

Since the output of a formula $\sqcup\phi$ is always a single assignment, the choice between the connectives ';' and '$\Rightarrow$' is neutralized: $\sqcup\phi \Rightarrow \psi \leftrightarrow \sqcup\phi \,;\, \psi$. In the case of donkey anaphora, applying $\sqcup$ to the restrictor renders weak and strong dynamic generalized quantifiers equivalent:

(25) $\mathcal{Q}_W x(\sqcup\mathcal{E}y\phi, \psi) \quad\leftrightarrow\quad Qx(\sqcup\mathcal{E}y\phi, \sqcup\mathcal{E}y\phi \,;\, \psi) \quad\leftrightarrow\quad Qx(\mathcal{E}y\phi, \mathcal{E}y\phi \,;\, \psi)$

(26) $\mathcal{Q}_S x(\sqcup\mathcal{E}y\phi, \psi) \quad\leftrightarrow\quad Qx(\sqcup\mathcal{E}y\phi, \sqcup\mathcal{E}y\phi \Rightarrow \psi) \quad\leftrightarrow\quad Qx(\mathcal{E}y\phi, \mathcal{E}y\phi \,;\, \psi)$

Hence we can always translate *every* as $\mathcal{EVERY}_S$, and get a weak reading by using $\sqcup$. The classic weak reading example then receives the following translation:

(27) Every man who had a quarter put it in the meter.
$\mathcal{EVERY}_S x(\text{man}(x) \,;\, \sqcup\mathcal{E}y(\text{quarter}(y) \,;\, \text{had}(x,y)), \text{put-in-meter}(x,y))$

Of course, if we do not apply $\sqcup$, we get the strong reading. If we use a weak quantifier, we get an existential reading either way. If the nuclear scope of the quantifier is atomic, we get the same existential reading as we do with $\sqcup$; if it is not, as in the nuclear scope negation cases, the equivalence is broken.

This analysis makes the prediction that when a donkey sentence with *every* receives a weak reading, the existential can and must take scope under *any* quantificational operator, not just an intensional verb. This is borne out:

(28) Every man who has a credit card uses it every time he shops online.

The only plausible reading of (28) is a weak one, and the only available reading is the one on which *a credit card* ends up taking lowest scope: scope under *every time*. (28) means that for every man who has a credit card, whenever he shops online, he uses a credit card; it does not have the stronger reading that requires that every man have a single card he uses every time he shops online.

6 Conclusion

Extending Kanazawa's (1994) analysis of donkey anaphora with the $\sqcup$ operator to delay existential witness selection allows it to account for exceptional weak readings with $\downarrow$MON$\uparrow$ and $\uparrow$MON$\downarrow$ quantificational determiners, overcoming its empirical shortcoming while avoiding undue overgeneration.

References

Brasoveanu, Adrian. 2007. *Structured nominal and modal reference.* Ph.D. thesis, Rutgers University.

Groenendijk, Jeroen & Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14(1): 39–100.

Kanazawa, Makoto. 1994. Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting. *Linguistics and Philosophy* 17(2): 109–158.

Lappin, Shalom & Nissim Francez. 1994. E-type pronouns, I-sums, and donkey anaphora. *Linguistics and Philosophy* 17(4): 391–428.

Schubert, Lenhart K. & Francis Jeffry Pelletier. 1989. Generically speaking. In Barbara H. Partee, Gennaro Chierchia & Raymond Turner (eds.) *Properties, Types and Meaning, Volume II: Semantic Issues*, 193–268. Dordrecht: Kluwer.

The syntax and semantics of evaluative degree modification

Hanna de Vries

Utrecht institute of Linguistics OTS

Abstract. It is a well-known but little-studied fact that evaluative adverbs - adverbs indicating the attitude of the speaker towards the information she is conveying - can modify degree (*incredibly tall*, *ridiculously expensive*...). This paper offers a syntactosemantic account of this phenomenon. Following Morzycki (2004), I propose that evaluative degree modification involves a covert operator (which I will call EVAL); however, my proposal differs from that of Morzycki in several crucial respects. Most importantly, I argue that evaluative degree constructions should not be analysed as embedded exclamatives. Furthermore, I show how their syntactic behaviour illuminates their semantic composition.

1 Introduction

Evaluatives, a large and open class of adverbs, can systematically modify gradable adjectives as well as complete sentences. The different positions are associated with a clear difference in meaning:

(1) a. Maxwell is $\left\{\begin{array}{l}\text{surprisingly}\\ \text{remarkably}\\ \text{shockingly}\end{array}\right\}$ tall.

b. $\left\{\begin{array}{l}\text{Surprisingly}\\ \text{Remarkably}\\ \text{Shockingly}\end{array}\right\}$, Maxwell is tall.

The sentences in (1a) do not entail those in (1b): if we were expecting Maxwell to be tall, but just not *that* tall, we could utter (1a) but not (1b).

The semantics of the (b)-sentences seems uncomplicated: the adverb simply modifies the proposition expressed by *Maxwell is tall.* But what exactly do the adverbs modify in an evaluative degree construction (henceforth EDC) like *Maxwell is remarkably tall*? Do they similarly modify propositions, and if so, what do these propositions express? Where does the semantic difference between (1a) and (1b) come from?

I will adopt the following (fairly uncontroversial) assumptions about degree and degree phrases.

Degree constructions, like *Vernon is six feet tall* or *Vernon is taller than Maxwell*, involve a) a measurable property G; b) an individual x who gets 'measured', i.e. to whom G is applied, yielding a degree d; c) some other degree of said property d', and d) a comparison between d and d'. For example, a sentence like *Vernon is six feet tall* can be paraphrased as

‘There is a degree d such that Vernon is d-tall and d equals six feet’. Similarly, *Vernon is taller than Maxwell* has to consist of a degree predicate relating Vernon to his height, and a comparison between that height and some other value (provided, in this particular case, by Maxwell’s height). We capture this by assuming the following type and denotation for gradable adjectives like *tall*:

(2) $[\![tall]\!]_{\langle d,\langle e,t\rangle\rangle} = \lambda d\lambda x[\text{HEIGHT}(d)(x)]$

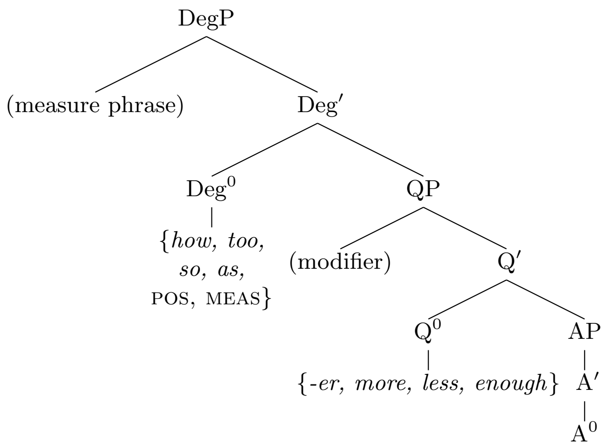
The comparison function, then, is provided by degree morphology: *-er*, for example, indicates a greater-than relationship between d and d'. Finally, the value of d' is provided by elements like *six feet* or *than Maxwell*.

While the above ingredients all need to be present in the semantics, they may be absent from overt syntax. The so-called ‘positive form’ (*Vernon is tall*) intuitively involves Vernon’s height being favourably compared to some contextually defined standard degree, but neither the comparison nor the standard are overt. Similarly, *Vernon is six feet tall* lacks an overtly stated equality relationship between Vernon’s height and the degree of six feet, yet it is obviously there in the semantics. We solve this by assuming covert degree morphology - POS for the positive form (where s_G is the contextually defined standard of G) and MEAS for sentences involving measure phrases.[1]

(3) a. $[\![\text{POS}]\!] = \lambda G\lambda x\exists d[G(d)(x) \wedge d \geq s_G]$
b. $[\![\text{MEAS}]\!] = \lambda G\lambda x\lambda d'\exists d[G(d)(x) \wedge d = d']$

I furthermore assume that the syntax of degree constructions is best described using a Degree Phrase, DegP (Abney 1987, Corver 1991, 1997), which looks as follows:

(4)



If Q^0 is empty or contains *-er*, head movement from A^0 to Q^0 takes place (Corver 1997). Modifiers like *very* and *extremely* are located in SpecQP, and measure phrases in SpecDegP.

[1] POS was already proposed in Cresswell (1976) and has since been commonly assumed; its measure phrase-introducing cousin first appeared in Kennedy (1997) and was baptised MEAS in Svenonius & Kennedy (2006).

2 EDCs are not embedded exclamatives

One of the most intuitive ways to paraphrase an EDC like *Vernon is remarkably tall* is something like 'the degree to which Vernon is tall is remarkable' (cf. Cresswell 1976, Katz 2005). However, in one of the few, if not the only, existing accounts of EDCs[2], Morzycki (2004) explicitly rejects this paraphrase as a correct representation of the semantics of EDCs. Consider a situation in which Vernon is in fact remarkably short - surely, we would be able to claim that 'the degree to which Vernon is tall is remarkable'. However, we would not call the remarkably short Vernon *remarkably tall.* In another scenario envisioned by Morzycki, Vernon was born at precisely 5:09 in the morning, on the fifth day of the ninth month of 1959 - and to our amazement, his height happens to be exactly five feet and nine inches. This is remarkable indeed, and yet, again, we would not be able to claim that Vernon is *remarkably tall.* This leads Morzycki to analyse EDCs as embedded exlamatives, which he takes to denote sets of true propositions, just like questions. For *Vernon is remarkably tall* to be true, one of the propositions in the set must be remarkable (see (5b)). For current purposes, this amounts to the denotation in (5c), in which reference is made to sets of degrees rather than sets of propositions.

(5) a. $[\![$*How tall Vernon is!*$]\!] = \{p : p$ is true and there is a degree of height d such that p is the proposition that Vernon is d-tall$\}$

b. $[\![$*Vernon is remarkably tall*$]\!] = [\![$*It is remarkable [how tall Vernon is!]*$]\!] = \exists p[p \in \{$'V. is 6 feet 1 inch tall', 'V. is 6 feet 2 inches tall', 'V. is 6 feet 3 inches tall', ... 'V. is n feet m inches tall'$\} \wedge$ REMARKABLE$(p)]$

c. $=$ REMARKABLE$(\hat{\ }\exists d[d \in \{$'6 feet 1 inch', '6 feet 2 inches', '6 feet 3 inches', ... 'n feet m inches'$\} \wedge$ Vernon is d-tall$])$

Following Zanuttini & Portner (2003), Morzycki argues that a crucial property of exclamatives is *domain widening.* To see the effect of this, consider the different implications about Maxwell's eating habits in (6a-b):

(6) a. Maxwell eats everything.

b. What things Maxwell eats!

Arguably, the domain of *everything* in (6a) is restricted by the context such that we do not expect it to include "lightbulbs, his relatives, or presidential elections" - or, in general, anything but ordinary food. For (6a) to be true, it is not necessary that Herman's eating habits include things like live locusts for breakfast; it merely suggests that Herman is a particularly easy dinner guest. In contrast, (6b) does suggest that the domain of things eaten by Herman also includes the extraordinary, like live locusts or raw serrano chillies. This is the effect of domain widening. Morzycki's semantics for EDCs, which takes into account both domain widening and factivity (essentially, the entailment of the positive form) is given in (7):

(7) $[\![\text{EDC}]\!] = R(\hat{\ }\exists d \exists C'[C' \supset C \wedge d \in C' - C \wedge G(d)(x) \wedge d \geq s_G])$ (for some gradable adjective G, evaluative adverb R, domain C and individual x)

[2] Another one is Katz (2005), although he only discusses the semantic, not the syntactic, side of the matter

In the rest of this section, I will argue that this is wrong.

Most importantly, there is something unjustifiably redundant about the domain widening part of the denotation in (7). It guarantees that the degree to which x is G is somehow so 'extreme' that it falls outside of the range of degrees we would naturally consider. But that is just another way of saying that what is going on is 'remarkable', or 'surprising', or 'unbelievable'. As an illustration, take a sentence like *Maxwell is remarkably tall.* Paraphrasing Morzycki's denotation, the semantics of this would boil down to something like 'It is remarkable that Maxwell's degree of tallness is such that it is somehow unexpected'.
This is, in fact, a general problem of analysing EDCs as embedded exclamatives. An exclamative (*How tall Maxwell is!*) can itself be paraphrased as something roughly like 'Maxwell is unexpectedly tall'. This suggests that the 'unexpectedness' of Maxwell being d-tall is the case even before the contribution of *remarkably* to the semantics.
We can test empirically whether this is true: we would expect the sense of unexpectedness or extremeness caused by domain widening to be there, regardless of the meaning of the modifier. This expectation is not borne out, however. The evaluative adverbs in (8) themselves do not express anything 'extreme', and indeed, the sentences in (8) do not seem to suggest unexpectedness or extremeness in any way.

(8) Maxwell is $\left\{\begin{array}{l}\text{disappointingly}\\ \text{arousingly}\\ \text{satisfyingly}\end{array}\right\}$ tall.

In short: the apparent domain widening effect of certain EDCs, like *Maxwell is remarkably tall*, seems to be a consequence of the semantics of the particular adverb, rather than a property of this kind of construction in general. If domain widening is a crucial part of the semantics of exclamatives, it follows that EDCs cannot involve embedded exclamatives.
I propose that the intuitive paraphrase we saw earlier, which was rejected by Morzycki, is in fact the right one. The nonexistent interpretations involving freakish heights are ruled out by an independent reason: the monotonicity of gradable predicates in the following sense (Heim 2000):

(9) A function f of type $\langle d, \langle e, t\rangle\rangle$ is monotone iff
$\forall x \forall d \forall d'[f(d)(x) = 1 \ \& \ d' < d \rightarrow f(d')(x) = 1]$

In words: If x has a certain property to a degree d, it also has this property to all lower degrees d'.
How does it follow from this that *Maxwell is remarkably tall* cannot be an appropriate description of a sitaution in which Maxwell's height equals his birthday? The crucial factor here is that *remarkable* is also monotone - downward monotone, to be precise. A downward monotone operator O, when applied to some proposition p, reverses p's entailments: if $p \models p'$, then $O(p') \models O(p)$. (It is easy to verify that this indeed holds for *remarkable* and other evaluatives.) Similarly, as the monotonicity of *tall* implies that $\text{TALL}(d)(x) \models \text{TALL}(d')(x)$ where $d' \leq d$,

(10) $\text{REMARKABLE}(\hat{}\,\text{TALL}(d')(x)) \models \text{REMARKABLE}(\hat{}\,\text{TALL}(d \geq d')(x))$

In other words, the monotonicity of both *remarkable/remarkably* and *tall* ensures that if it is remarkable that x is d'-tall, x being $d \geq d'$-tall must also be remarkable. Clearly, this cannot be true in a situation in which Vernon is remarkably short or has a height corresponding to his birthday.[3]

Thus, we have arrived at a semantics for EDCs that is both empirically more accurate and quite a bit simpler than Morzycki's. Unlike Morzycki, we do not need to assume that evaluatives are ambiguous between an $\langle\langle\langle s,t\rangle,t\rangle,t\rangle$ and an $\langle\langle s,t\rangle,t\rangle$ type. Moreover, our semantics defines the relationship between the different semantics associated with different adverb positions in an elegant, intuitive way that mirrors their syntactic difference, namely in terms of scope:

(11) a. Maxwell is remarkably tall:
$\exists d[\text{REMARKABLE}(\hat{}\,\text{TALL}(d)(\text{m})) \wedge \text{TALL}(d)(\text{m})) \wedge d \geq s_{tall}]$
b. Remarkably, Maxwell is tall:
$\text{REMARKABLE}(\hat{}\,\exists d[\text{TALL}(d)(\text{m}) \wedge d \geq s_{tall}]) \wedge \exists d[\text{TALL}(d)(\text{m}) \wedge d \geq s_{tall}]$

3 Syntactic movement, *much*-support and agreement

We now turn to the syntax of EDCs in order to see how the above denotation is arrived at. First, I propose to take the syntacticians seriously and assume that evaluative adverbs, like other degree modifiers, are located in SpecQP. This is supported, among other things, by the fact that evaluatives can gradually lose their meaning and flexibility and turn into 'proper' degree modifiers; for example, Dutch *ontzettend* has mostly lost its original meaning of 'shocking, horrifying' and is nowadays used almost exclusively as a degree modifier.

Now, consider the following examples of *so*-pronominalisation in English (from Corver 1997):

(12) a. John is *fond of Mary.* Bill seems [less *so*].
b. John is *fond of Mary.* *Maybe he is [too *so*].

When the whole AP is replaced by the pro-form *so*, there is no A^0 to raise to Q^0. Corver notes that this results in ungrammaticality when Q^0 is empty (12b). To make the Deg^0 + *so* combination grammatical, we need to insert the syntactic dummy *much* into Q^0 ('*much*-support'):

(13) John is *fond of Mary.* Maybe he is [too **much** *so*].

Now, consider the data in (14):

(14) Vernon is tall, even $\left\{\begin{array}{l}\text{remarkably}\\ \text{surprisingly}\\ \text{eerily}\\ \ldots\end{array}\right\}$ (*much) so.

The fact that *much*-insertion is ungrammatical here can only be explained by assuming that Q^0 is not empty - it must be occupied by a covert element. I propose that this covert element, which I will call EVAL, is a null degree morpheme that applies to an evaluative and a gradable adjective to yield a lambda term with exactly the same semantic type as the adjective itself, to which POS (in Deg^0) is then applied in the usual

[3] The reader is referred to Nouwen (2005) for the whole argument in detail.

way.

The analysis presented above defines the difference between evaluative modifiers and 'true' degree modifiers (e.g. *very, pretty*) in terms of their ability to modify degree directly - the latter can, while the former need the mediation of EVAL. In English, this claim is supported by the presence or absence of *much*-support in the case of *so*-pronominalisation (*very* and *pretty* do need *much*-support). Dutch does not have *so*-pronominalisation, but it does offer some interesting independent evidence in the form of gender agreement between adjective and modifier.

(15) a. Een belachelijk(*e) dure fiets
'a ridiculously(-INFL) expensive bike'
b. Een ontzettend(?e) mooie fiets
'an extremely(-INFL) beautiful bike'
c. Een ?heel/hele mooie fiets
'a very/very-INFL beautiful bike'

Belachelijke 'ridiculous' in (15a) cannot receive a degree-modifying interpretation; only the non-inflected form can. The use of the inflected form *onzettende* 'extremely' as a degree modifier, however, is relatively common; finally, degree-modifying *hele* 'very' has an overwhelming tendency to agree with the adjective.[4] Assuming that agreement reflects a Spec-Head relationship, the difference follows naturally from our assumptions: EDCs do not involve a Spec-Head relationship between the modifier and the adjective, as the presence of EVAL in Q^0 prevents the adjective from raising there. In contrast, *heel/hele* 'very', as a proper degree modifier, does not need the mediation of an element like EVAL in Q^0, so the adjective can raise to this position, ending up in a Spec-Head relationship with the degree modifier. Finally, the mixed behaviour of *ontzettend* is exactly what we would expect of an evaluative that is diachronically turning into a 'real' degree modifier.

The account presented here is similar in spirit to that of Morzycki (2004), who also deals with EDCs in terms of covert morphology; however, the syntactic and semantic details are quite different, as Morzycki locates evaluatives in SpecDegP and collapses the semantic contributions of POS and (his rather different version of) EVAL into one covert morpheme located in Deg^0. Neither of these choices, however, is compatible with the syntactic data presented here.

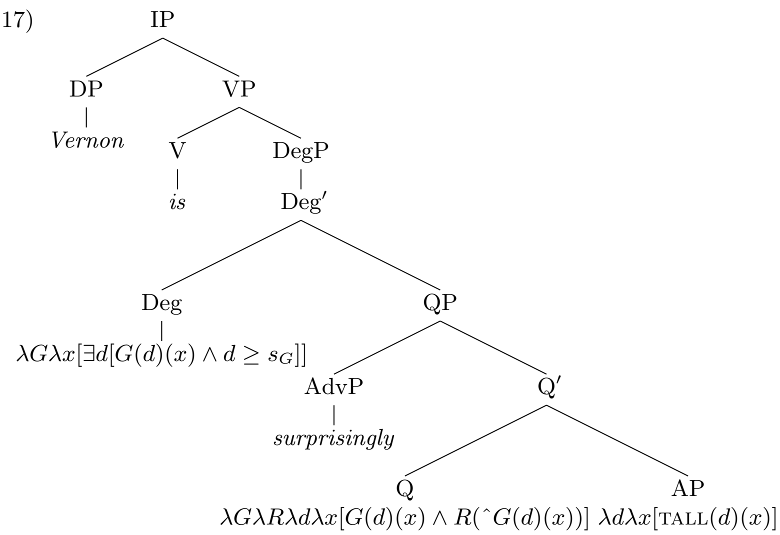
4 Assembling the pieces

The denotation I assume for EVAL is the following:

(16) $[\![\text{EVAL}]\!] = \lambda G \lambda R \lambda d \lambda x[G(d)(x) \wedge R(\hat{}G(d)(x))]$

Here, G is a gradable adjective and R an evaluative adverb. The semantics of a sentence like *Vernon is surprisingly tall*, then, is built up as follows (heads are applied to their specifiers and complements):

[4] Googling *belachelijke* {*dure/mooie*} returned 1,872 hits, against 17,110 hits for *belachelijk* {*dure/mooie*}. *Ontzettende* vs *ontzettend*: 29,172 vs. 31,160. *Hele* vs. *heel*: 900,200 vs. 265,900. Other evaluatives pattern with *belachelijk*.

(17)


(18) $[\![$*Vernon is* POS EVAL *surprisingly tall*$]\!]$

$= [\![\text{POS}]\!]([\![\text{EVAL}]\!]([\![\textit{tall}]\!])([\![\textit{surprisingly}]\!]))([\![\textit{Vernon}]\!])$

$= [\![\text{POS}]\!](\lambda G \lambda R \lambda d \lambda x[A(d)(x) \wedge R(\hat{}G(d)(x))](\lambda d \lambda x[\text{TALL}(d)(x)])$
$([\![\textit{surprisingly}]\!]))([\![\textit{Vernon}]\!])$

$= [\![\text{POS}]\!](\lambda R \lambda d \lambda x[G(d)(x) \wedge R(\hat{}\text{TALL}(d)(x)](\text{SURPRISING}))([\![\textit{Vernon}]\!])$

$= \lambda G \lambda x[\exists d[G(d)(x) \wedge d \geq s_G]](\lambda d \lambda x[\text{TALL}(d)(x) \wedge \text{SURPRISING}(\hat{}\text{TALL}$
$(d)(x))])([\![\textit{Vernon}]\!])$

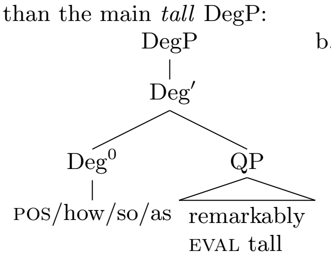
$= \lambda x[\exists d[\text{TALL}(d)(x) \wedge \text{SURPRISING}(\hat{}\text{TALL}(d)(x)) \wedge d \geq s_{tall}]](\text{v})$

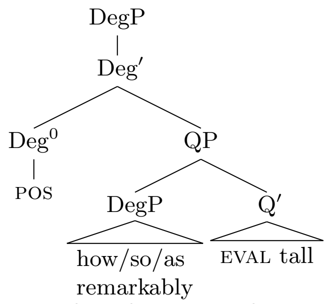
$= \exists d[\text{TALL}(d)(x) \wedge \text{SURPRISING}(\hat{}\text{TALL}(d)(\text{v})) \wedge d \geq s_{tall}]]$

Importantly, the above analysis leaves room for EDCs to be headed by other degree morphemes than POS, thus correctly predicting the existence of constructions like (19a-c):

(19)
a. How remarkably tall Maxwell is!
b. Maxwell is so remarkably tall that all tourists want to take a picture with him.
c. Maxwell is just as remarkably tall as Vernon.

One might object that there is a different way to analyse these sentences, namely (20b), in which *how, so*, and *as* are heading the *remarkably* DegP rather than the main *tall* DegP:

(20) a.


b.


However, there are good reasons to assume that this is not the case. First, note that we can replace each instance of *remarkably* in (19) with the non-gradable *very* (for which the structure in (20b), obviously, is no option). This shows that there is nothing wrong with the combination of an overt degree head and a modifier per se, so as long as we cannot prove that evaluative modifiers behave differently from *very*, we have no reason to assume that they do.

Secondly, we can test empirically whether (20a) or (20b) is right, at least in the case of the degree head *as*. It is well-known (cf. Kennedy 1997) that comparative and equative constructions are semantically anomalous if the adjectives that are compared are not measured along the same dimension ((21a) vs. (21b)). We can use this fact to determine which element is compared to which in a sentence like (20c). If *as* heads the embedded *remarkably* DegP, as in (20b), we would expect the (a)nomaly of the comparison to depend on the dimension of *remarkably*. In contrast, if *as* heads the main DegP, as in (20a), we would expect the dimension of the adjective to be decisive. And in fact, the latter expectation is borne out:

(21) a. Maxwell is just as tall as Vernon is wide.
b. #Maxwell is as tall as he is arrogant.

(22) a. Maxwell is as remarkably tall as Vernon is remarkably wide.
b. #Maxwell is as remarkably tall as Vernon is remarkably arrogant.

The difference between (22a) and (22b) would be inexplicable if the first instance of *remarkably* were compared with the second one, or with *wide/arrogant*; the only way to explain why (22b) is anomalous while (22a) is not, is to say that (22b) compares *tall* and *arrogant*, which do not have identical dimensions, whereas (22a) compares *wide* and *tall*, which do. This means that *tall*, and not *remarkably*, is the complement of *as*.

Concludingly, while there is no decisive way to prove for all possible overt degree heads that they can co-occur with an evaluative modifier in SpecQP, the facts we do have (*as*, at least, can; plus, evaluative modifiers pattern with *very* in every other respect) are suggestive enough. I conclude that the predictions of the analysis presented in this paper are, indeed, correct.

One exception, however, is MEAS. There is no semantic reason to rule out a combination of MEAS and EVAL; yet, as (23) shows, evaluative modifiers are incompatible with measure phrases (a property they share with ordinary degree modifiers like *very*):

(23) *Vernon is seven feet very/remarkably tall.

However, this structure can be ruled out on independent syntactic grounds: it is argued in Corver (1997, 2009) that measure phrases originate in SpecQP, which explains why they are in complementary distribution with modifiers.[5]

[5] As one of the reviewers of this paper pointed out, it is possible to get approximately the intended semantics ('There is a degree d such that $d = 7''0$ and Maxwell is d-tall and it is remarkable that Maxwell is d-tall') by using a slightly different construction: *Maxwell is a remarkable seven feet tall.* This suggests that the *semantics* of evaluatives and measure phrases are by no means incompatible. Constructions like these are a test case for the semantics presented in this paper, but I will leave this issue for future study.

5 Conclusions

In this paper, I have argued that an analysis of EDCs as embedded exclamatives runs into several conceptual and empirical problems, and subsequently, that the much more elegant alternative is in fact perfectly valid if we assume that gradable adjectives and evaluatives are monotone. Furthermore, I have proposed a syntax for EDCs based on evidence involving *much*-support and Dutch gender agreement; this syntax allows EDCs to be headed by, in principle, any degree head, which draws a nice parallel between EDCs and other degree constructions.
It also allows words like *very* and *pretty* to be treated syntactically like degree modifiers (occupying SpecQP), while still explaining why they occasionally behave differently from evaluatives that occupy the same position: *very*, *pretty* and other 'true' degree modifiers can directly modify the adjective without needing the intervention of something like EVAL.

References

Abney, Steven (1987). *The English Noun Phrase in its Sentential Aspect.* PhD dissertation, MIT, Cambridge, MA.

Corver, Norbert (1991). 'Evidence for DegP'. *Proceedings of NELS 21*, Umass, Amherst.

Corver, Norbert (1997). 'Much-support as a last resort'. *Linguistic Inquiry* 28:1, 119-164.

Corver, Norbert (2009). 'Getting the (syntactic) measure of Measure Phrases'. *The Linguistic Review* 26:1, 67-134.

Cresswell, Max (1976). 'The semantics of degree'. In: Partee, Barbara (ed.). *Montague Grammar*, 261-292. New York: Academic Press.

Heim, Irene (2000). 'Degree operators and scope'. In: *Proceedings of SALT X*, Ithaca, NY: CLC Publications.

Katz, Graham (2005). 'Attitudes towards degrees'. In: Maier, Emar, Corien Bary & Janneke Huitink (eds). *Proceedings of Sinn und Bedeutung 9*.

Kennedy, Christopher & Louise McNally (2005). 'Scale structure, degree modification, and the semantics of gradable predicates'. *Language* 81:2, 345-381.

Kennedy, Christopher (1997). *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison.* PhD dissertation, University of California, Santa Cruz.

Morzycki, Marcin (2004). 'Evaluative adverbial modification in the adjectival projection'. Ms, Université du Québec à Montréal.

Nouwen, Rick (2005). 'Monotone amazement'. In: Dekker, Paul & Michael Franke (eds). *Proceedings of the Fifteenth Amsterdam Colloquium*, University of Amsterdam.

Svenonius, Peter & Christopher Kennedy (2006). 'Northern Norwegian degree questions and the syntax of measurement'. In: Mara Frascarelli (ed.). *Phases of Interpretation*, 133-161. Berlin: Mouton de Gruyter.

Zanuttini, Raffaella and Paul Portner (2003). 'Exclamative clauses: At the syntax-semantics interface'. *Language* 79(1):39.

Epistemic Modals are (Almost Certainly) Probability Operators

Daniel Lassiter

New York University

Abstract. The epistemic modals *possible, probable, likely*, and *certain* demand a semantics which treats them both as modal operators and as gradable adjectives. I show that the standard theory of modality in linguistic semantics – due largely to the work of Angelika Kratzer – makes incorrect predictions about the interaction of *likely* and disjunction. However, enriching Kratzer's theory by directly incorporating probabilistic information allows us to give a compositional degree-based treatment of gradable modals, avoid the problems with disjunction, and explain numerous facts about the interaction of gradable modals with degree modifiers and negation.

1 Introduction

1.1 Gradable Modals

Most discussion of the semantics of English modals has focused on the meanings of modal auxiliaries such as *must*, *should*, and *can*, and on infinitival modals like *ought to*. However, there is a substantial number of adjectival modals in English as well, and these have received somewhat less attention in the literature. Importantly, like adjectives in general, many adjectival modals are readily **gradable**, just like the large and well-studied class of gradable adjectives. Some examples of the similarities are given in (1)-(4).

(1) Degree Modification
   a. Bill is very angry.
   b. It is very likely that Jorge will win the race.
   c. The glass is almost full.
   d. It is almost certain that Jorge will win the race.

(2) Comparison
   a. Bill is angrier than Sue.
   b. It is more likely that Jorge will win the race than it is that Sue will win.

(3) Degree questions
   a. How angry is Bill?
   b. How likely is it that Jorge will win the race?

(4) Explicit Degree Quantification
   a. The glass is 95% full.
   b. It is 95% certain that Jorge will win the race.

I will refer to modal expressions with these characteristics as **gradable modals**. The epistemic subtype, exemplified by *likely*, *possible*, and *certain*, will be referred to as **gradable epistemic modals**, hereafter GEMs.

1.2 Two Problems

We begin with two observations. First, as (1)-(3) illustrate, many expressions of modality in English are gradable adjectives, and can be syntactically quite complex. As Portner (2009) notes, it would be desirable to have an account of GEMs that is compatible with the best available syntactic and compositional semantic theory of adjectives like *angry* and *full*. Surprisingly, however, no such theory exists.

The second observation is the following. Imagine it is the beginning of baseball season. You have a friend who is a sports fanatic and is uncannily good at predicting who will succeed in the postseason from pre-season play. He tells you that the Bluejays are the best team in the Major Leagues this year, and that they are at least as likely to win the World Series as any other individual team this year. As it happens, you encounter a bookie who is willing to take bets on the Bluejays winning the World Series at 2:1 odds – that is, if you bet on the Bluejays and you are right, you will win twice what you bet. Should you risk a large sum of money on this bet? Not necessarily – after all, there are twenty-nine other major league baseball teams, and there is a non-negligible chance that one of them will outperform your friend's expectations and you will lose your money. That is, as a good gambler, you should know that even though the Bluejays are at least as likely to win as any other team, it may not be the case that the Bluejays are at least as likely to win as they are not to win.

It turns out that within the standard theory, it is impossible to make sense of the final sentence in the previous paragraph. The dominant theory predicts instead that the intuitively stronger claim in (5b) is actually an *entailment* of the intuitively weaker claim in (5a).

(5) a. The Bluejays are the best team in the league: for any team you like, the Bluejays are at least as likely to win the Series as that team is.

b. The Bluejays are at least as likely to win the Series as they are not to win.

Our starting point, then is this: we need a theory of the semantics of *likely* and other gradable modals that (a) is compatible with a good theory of the semantics of gradable adjectives, and (b) does not wrongly predict that the (5a) entails (5b).

2 Kratzer's Semantics for Gradable Modals

The standard theory of modality in linguistic semantics was developed in a series of papers by Angelika Kratzer (see especially Kratzer (1981) and Kratzer (1991)). On this approach, notions such as likelihood, obligation, etc. as derived from a more basic notion, that of **comparative possibility**. She introduces a binary relation $\succcurlyeq$ which holds of two worlds u,v just in case u is "more possible" than v in a way to be made precise.[1] $\succcurlyeq$ is interpreted relative to a modal base **f** and an ordering source **g**. **f** is a function which, given a world, returns a set of propositions that are relevant to the evaluation of the modal expression. In the case of epistemic modality, the modal base is the set of propositions known to the speaker (or whoever the contextually appropriate person(s) are).

The ordering source **g** is a function which, applied to a world w, returns a set of propositions which induces an ordering over the modal base. In the case of deontic

[1] Kratzer (1981) actually uses $\preccurlyeq$ where I will use $\succcurlyeq$. This choice makes sense within the historical setting of her theory, but it is confusing in the current context, since it seems to suggest 'less than or equal to', while the orderings we are interested in with respect to gradable modals correspond intuitively to an ordering in terms of 'greater than or equal to'. When we compare these notions to the orderings induced by gradable adjectives, the current notation will be preferable.

modals, for example, the ordering source is some contextually relevant set of laws, orders, norms, etc., and worlds are ranked by how close they come to satisfying all of the propositions in $\mathbf{g}(w)$. The ordering is determined by the rule in (6):

(6) For all worlds $u, z \in W$:
$u \succcurlyeq_{\mathbf{g}(w)} z$ if and only if $\{p : p \in \mathbf{g}(w) \text{ and } z \in p\} \subseteq \{p : p \in \mathbf{g}(w) \text{ and } u \in p\}$

That is, u is at least as good a world as z iff u satisfies every law (norm, etc.) that z does. u is strictly better than z iff u satisfies every law that z does, and z does not satisfy every law that u does. u and z are ranked equally ($u \approx z$) iff u and z satisfy the same laws.

Kratzer (1981) suggests that, in the epistemic domain, the ordering source is composed of propositions representing the "normal state of affairs": "Worlds in which the normal course of events is realized are a complete bore, there are no adventures or surprises". That is, in the special case in which the propositions in $\mathbf{g}$(w) are consistent and $\bigcap \mathbf{g}$(w) is thus non-empty, p is more likely than q just in case p is more similar to the maximally normal state(s) of affairs given by $\bigcap \mathbf{g}$(w) than q is.

Formally, Kratzer gives the semantics of $\succcurlyeq_{\mathbf{g}(w)}$ as follows:

(7) *Comparative Possibility*:
A proposition p is more possible than a proposition q in a world w in view of a modal base $\mathbf{f}$ and an ordering source $\mathbf{g}$ if, and only if, the following conditions are satisfied:

a. For all $u \in \bigcap \mathbf{f}(w)$:
If $u \in q$, then there is a world $v \in \bigcap \mathbf{f}(w)$ such that $v \succcurlyeq_{\mathbf{g}(w)} u$ and $v \in p$.

b. There is a world $u \in \bigcap \mathbf{f}(w)$ such that:
$u \in p$ and there is no world $v \in \bigcap \mathbf{f}(w)$ such that $v \in q$ and $v \succcurlyeq_{\mathbf{g}(w)} u$.

In words: p is more possible than q iff, for every q-world, there is some p-world that is at least as normal (relative to $\mathbf{g}(w)$), and there is at least one p-world for which no q-world is as normal. I will sometimes abbreviate "p is more possible than q" by $p \succ^{s} q$. Note that, unlike $\succ$, these are relations on propositions.

Similarly, we can define the relation "p is at least as possible as q" ($p \succcurlyeq^{s} q$)as follows:

(8) For all $u \in \bigcap \mathbf{f}(w)$:
If $u \in q$, then there is a world $v \in \bigcap \mathbf{f}(w)$ such that $v \succcurlyeq_{\mathbf{g}(w)} u$ and $v \in p$.

This is, of course, just the first clause of (7).

Kratzer's theory has many advantages: in particular, it gives plausible truth-conditions for sentences involving complex expressions of graded modality such as *ϕ is probable* and *ψ is a slight possibility.* However, this theory makes incorrect predictions for an important class: examples involving disjunction.

3 The Problem of Disjunction

Disjunction creates a deep problem for Kratzer's theory which, I will argue, cannot be resolved without enrichment of the underlying logic to include numerical probabilities. The problem is this: if *or* has its classical denotation, then Kratzer's theory predicts that the following inference should be valid.

(9) a. ϕ is at least as likely as ψ.
b. ϕ is at least as likely as χ.
c. $\therefore$ ϕ is at least as likely as $(\psi \vee \chi)$.

Proof of (9). (For readability, I suppress reference to $\mathbf{f}(w)$ and $\mathbf{g}(w)$.) (9a) holds just in case, for every ψ-world u there is a ϕ-world v such that $v \succcurlyeq u$. Similarly, by (9b), for every χ-world u' there is a ϕ-world v' such that $v' \succcurlyeq u'$. Let w be an arbitrary world in $\psi \vee \chi$. Suppose $w \in \psi$. Then there is a ϕ-world w' such that $w' \succcurlyeq w$, namely v. Similarly, if $w \in \chi$, then there is a ϕ-world w' such that $w' \succcurlyeq w$, namely v'. Since w was arbitrary, for all $w \in \psi \vee \chi$, there is a $w' \in \phi$ such that $w' \succcurlyeq w$. Thus, by the definition in (8), (9c) holds.

Informally, (9) is valid because of the following feature of Kratzer's definition of comparative possibility. Suppose, for simplicity, that for each of ϕ, ψ, and χ there is some (possibly singleton) set of worlds that outrank all other worlds in that proposition, and the top-ranked worlds in ϕ, ψ, and χ are all connected. The ordering of propositions according to (9) is uniquely determined by the relative positions of these top-ranked worlds. The premises (9a) and (9b) entail that the top-ranked worlds in ϕ outrank the top-ranked worlds in ψ, and also the top-ranked worlds in χ. Fig. 1 depicts the situation.

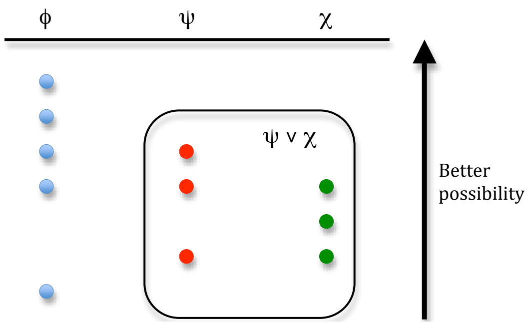


Fig. 1.

Because the top-ranked worlds in ϕ outrank the top-ranked worlds in ψ and in χ, they also outrank the top-ranked worlds in $\psi \vee \chi$. Since comparative possibility is only sensitive to the relative positions of the top-ranked worlds, it follows that ϕ is more likely than $\psi \vee \chi$ – and therefore, trivially, that ϕ is at least as likely $\psi \vee \chi$.

The problem is that the inference pattern in (9) is intuitively invalid, particularly when this schema is applied repeatedly. Recall that, in the scenario described above, you have it on good authority that the best team in Major League Baseball is the Toronto Bluejays. That is, your prescient friend says that

(10) "For any team in the Major Leagues, the Bluejays are at least as likely to win the Series as that team is."

This is obviously a weaker claim than (11):

(11) "The Bluejays are at least as likely to win as they are not to win."

But on Kratzer's theory, (10) plus a few simple facts about baseball entail (11). To see this, let $\{\mathbf{team}_1, \ldots, \mathbf{team}_{29}\}$ be the other 29 Major League baseball teams. (12) is a reasonable rendition of the truth-conditions of (10):

(12) $\forall x \in \{\mathbf{team}_1, \ldots, \mathbf{team}_{29}\}$:
It is at least as likely that the Bluejays will win as it is that x will win.

Let p be the proposition *The Bluejays win*, and let q_n be the proposition *Team$_n$ wins*. Since only Major League teams can compete, (12) is equivalent to (13):

(13) $(p \succcurlyeq^s q_1) \wedge (p \succcurlyeq^s q_2) \wedge \ldots \wedge (p \succ^s q_{29})$.

Feeding (13) into the inference schema (9), we see that (14) follows as well.

(14) $p \succcurlyeq^s (q_1 \vee q_2 \vee ... \vee q_{29})$.

Since we know that one of the thirty teams must win, the only way that the Bluejays can lose is if someone else wins – that is, *The Bluejays do not win* entails (14). Putting this all together, we see the problem: (10) entails (14), and (14), along with some simple assumptions about baseball, entails (11). But, intuitively, (11) is clearly **not** an entailment of (10): in fact, it is a much stronger claim.

It is a general feature of Kratzer's system that a disjunction ϕ, no matter how large, can never be more likely than another proposition ψ unless one of the disjuncts of ϕ is itself more likely than ψ. Is this a plausible prediction about valid inferences from sentences involving *likely*? In the case at hand, the answer is clearly "no". Somehow, a disjunction of lower-ranked possibilities can, as it were, "gang up" to overpower a higher-ranked possibility. This suggests that we need a theory of comparative likelihood which employs not just comparative measures, but **quantitative measures**. Numerical probability, I will show, provides precisely what we need.

4 Gradable Epistemic Modals as Probability Operators

My proposal is the following[2]:

(15) Gradable epistemic modals are **probability operators**: they denote measure functions, i.e. functions from propositions to elements of [0,1], subject to two restrictions:
(1) $prob(W) = 1$, and
(2) $prob(A \cup B) = prob(A) + prob(B)$ whenever $A \cap B = \varnothing$.

(For related discussion, see the forthcoming Yalcin (2010), from whom I borrow the term "probability operators".)

The first attraction of numerical probability is that it avoids the undesirable interaction with disjunction that we have noted for Kratzer's theory. First consider the problematic inference schema (9). This argument is clearly invalid on the present proposal. For instance, let $prob(\phi) = .3$, $prob(\psi) = .2$, $prob(\chi) = .2$, and $prob(\psi \wedge \chi) = 0$, so that the premises (9a) and (9b) are true. If this is right, then $prob(q \vee r) = .4$, and so (9c) does not hold. So Kratzer's problems with disjunction do not arise.

The remainder of the paper explores another strong motivation for the present proposal: the distribution of degree modifiers with GEMs is exactly what the present theory predicts, if we give GEMs the meanings in (16).

(16) a. $[\![\phi \text{ } is \text{ } certain]\!] = 1$ iff $prob(\phi) = 1$) (modulo pragmatic slack).
b. $[\![\phi \text{ } is \text{ } possible]\!] = 1$ iff $prob(\phi) > 0$ (modulo pragmatic slack).
c. $[\![\phi \text{ } is \text{ } probable/likely]\!] = 1$ iff $prob(\phi) > \mathbf{s}_{prob}$ (The contextually given standard for *relative adjectives* in Kennedy's theory, cf. below.).

5 Gradability

Kennedy & McNally (2005) and Kennedy (2007) argue that gradable adjectives denote measure functions, just as I have proposed for GEMs. However, most gradable adjective denote functions from individuals (rather than propositions) to points on

[2] (7) is a complete axiomatization of probability theory in the case of finite W; the complications that arise in the case of infinite W are not relevant here. Note that the usual properties of numerical probability, e.g. that $prob(\phi \vee \psi) = prob(\phi) + prob(\psi) - prob(\psi \wedge \chi)$, follow directly from (15).

a scale whose nature is (at least partially) determined by the lexical semantics of the adjective in question. So, for example, *tall* denotes a function from an object to its height. Formally, a scale is given by a triple $\langle D, \prec, \delta \rangle$, where D is a set of degrees, $\prec$ is a total ordering of D, and δ is the *dimension* of the adjective (e.g., in the case of *tall*, the dimension of height). The meaning of *tall* is as in (17):

(17) $[\![tall]\!] = \lambda x.\iota d.\ x\text{'s } height = d$

This is a function of type $\langle e, d \rangle$ – not the right type for a predicate, of course. Kennedy argues that adjectives always co-occur with a degree modifier which converts them to predicative type. When an adjective has an explicit degree modifer, the adjective is an argument of the modifier, as in (18).

(18) a. $[\![closed]\!] = \lambda x.\iota d.$ x's degree of closure $= d$
b. $[\![completely]\!] = \lambda A.\lambda x.\ A(x) = \max(D_A)$
c. $[\![completely\ closed]\!] = \lambda x.\iota d.\ x$'s degree of closure $= \max(D_{closed})$

Hence *The door is completely closed* means that the degree of closure of the door is the maximal degree in D_{closed}. This predicts immediately that we cannot use *completely* with an adjective associated with a scale with no maximal element.

Kennedy argues further that adjectives in the positive form are converted to predicative type by a silent degree morpheme **pos** which associates an adjective A with a contextually appropriate standard of comparison $\mathbf{s}_A$.

(19) $[\![\mathbf{pos}]\!] = \lambda A.\lambda x.A(x) \succ \mathbf{s}_A$

One of the most important parameters of variation among gradable adjectives, according to Kennedy & McNally, is the structure of D, i.e. the elements which enter into an ordering. They distinguish four possible types of scales – **totally open**, **totally closed**, **upper bounded**, and **lower bounded** – and argues that each is instantiated in the gradable adjectives and is linguistically significant.

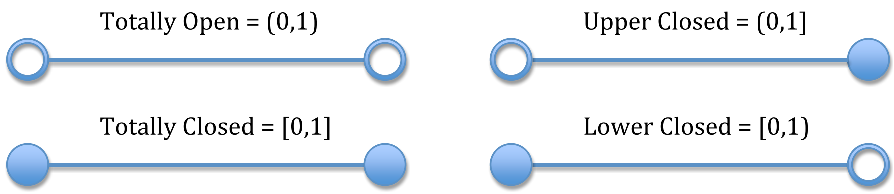


Fig. 2. Scale Types.

Kennedy discusses various tests for scale structure, several of which we will employ to discover the scale structure of gradable epistemic modals. The first test is degree modification: if an adjective can be modified by *completely* with a "maximum" interpretation, it denotes a function whose range is an upper closed scale, i.e. a **maximum standard** adjective. One example is *full*.

(20) The room is completely full.

Kennedy shows that the default interpretation of *full* and similar adjectives involves a maximum standard. This explains why (21) is strange – another test for scale structure. (Call this the "A but could be A-er" test.)

(21) #The room is full, but it could be fuller.

Note that this test does not tell us whether the scale in question is merely upper closed or is fully closed – it simply indicates that the scale has a maximum element.

Kennedy claims that, if an adjective can be modified by *slightly*, it denotes on a scale with a minimum point, i.e. a lower closed or fully closed scale. He dubs these "minimum-standard" adjectives because they typically require only that the degree to which an object possesses the property be greater than zero. An example is *bent*.

(22) The rod is slightly bent.

Minimum-standard adjectives do not accept degree modification with *completely*.

(23) #The rod is completely bent.

Minimum-standard adjectives also pass the "A but could be A-er" test.

(24) The rod is bent, but it could be more bent.

Minimum- and maximum-standard adjectives share the property that their negation entails that the corresponding reversed-polarity adjective holds:

(25) a. The rod is not bent. ⊨ The rod is straight.
b. The rod is not straight. ⊨ The rod is bent.

The third major class is the relative-standard adjectives, which are associated with contextually determined standards ($\mathbf{s}_A$ in the above). Relative-standard adjectives like *tall* indicate that the object in question has some significant degree of the property in question. Relative adjectives are odd with both *completely* and *slightly*.

(26) a. #Mary is completely tall.
b. #Mary is slightly tall.

In addition, relative adjectives do not maximize with *completely*, even if they denote on an upper closed scale. For instance[3]:

(27) a. The pizza is completely inexpensive. ⊭ The pizza is free. (Kennedy 2007)
b. On the moon, you would be completely lightweight. ⊭ You would be weightless.

Relative adjectives do not generate the same entailments as maximum and minimum adjectives in the tests we have seen, e.g. the "A but could be A-er" test:

(28) Mary is tall, but she could be taller.

Likewise, the negation of a relative adjective does not entail its negative counterpart:

(29) Mary is not tall. ⊭ Mary is short.

Finally, *proportional modifiers* – modifiers like *half*, *30%*, and *mostly* – occur only with adjectives which denote on a fully closed scale.

(30) a. # My cousin is 70%/half tall/short.
b. # This road is 70%/half dangerous/safe.

[3] Portner (2009) discusses this property of relative adjectives and identifies it as an objection to the analysis of GEMs that I am arguing for. It is true that this is a theoretical problem, but it is a general fact about relative adjectives, not a problem for *likely* and *probable* in particular.

As a reviewer notes, an alternative analysis is that *possible*, *likely*, and *probable* fail to maximize with *completely* because they are lexically specified to use only a part of the scale of probability, such as the open range $(0, 1)$. This is in fact precisely what Kennedy (2007) suggests for *inexpensive*. However, it will not work for *likely* and *probable*, which are acceptable with certain proportional modifiers (most frequently n%, see (32a)), indicating that their scale has a top element. In addition, I have reservations about the explanatory potential of such an approach, both for the gradable epistemic modals and for *inexpensive* and *lightweight*: see Lassiter (2010) for detailed discussion.

c. The glass is 70%/half full/empty.

These facts are as they are because proportional modifiers comparing the distance of a point to both the maximum and the minimum points on the relevant scale. If either of these points does not exist, then the modifier will be unacceptable.

6 Scales for Gradable Epistemic Modals

The proposal in (16) makes the following predictions:

(31) a. Gradable epistemic modals denote on a closed scale (i.e., [0,1]).
b. *Certain* is a maximum standard adjective.
c. *Possible* is a minimum standard adjective.
d. *Probable* and *likely* are relative adjectives.

Applying the tests developed in the previous section to the gradable epistemic modals, we see that these predictions are fully borne out. First, proportional modifiers are acceptable, indicating a closed scale.

(32) a. It is 70% likely that Jorge will win.
b. It is 99% certain that the Mets will lose.

Second, *certain* patterns with the maximum standard adjective *full* on the tests we have discussed.

(33) a. It is completely/#slightly certain that the Jets will win this year.
b. #It is certain that the Jets will win, but it could be more certain.
c. It is not certain that the Jets will win. ⊨ It is uncertain they will win.

Third, *possible* patterns in all relevant respects with the minimum-standard adjectives such as *bent*.

(34) a. #It is completely possible that the Jets will win this year.
b. It is slightly possible that the Jets will win.
c. It is possible that the Jets will win, but it could be more possible.
d. It is not possible that the Jets will win. ⊨ It is impossible they will win.

Finally, *likely* and *probable* pattern with the relative adjectives on all of the tests we have discussed.

(35) a. #It is completely likely/probable that the Jets will win this year.
b. #It is slightly likely/probable that the Jets will win.
c. It is likely that the Jets will win, but it could be more likely.

These facts lend considerable support to the hypothesis in (16) and its corrolary in (31): GEMs denote functions from propositions to portions of the closed scale [0,1].

7 Conclusions and Future Directions

My proposal, though a substantial revision of the standard theory, is not quite as radical as it may appear. Numerical probability is a strictly richer system than comparative possibility, and so anything that can be represented in Kratzer's theory can in principle be represented using probabilities as well.[4]

[4] I have not shown this, but the proof should be apparent to those familiar with measurement theory (Krantz, et al. 1971): Comparative possibility defines an ordinal scale, one of the weakest scale types, while numerical probability is much richer, even more than a ratio scale (a ratio scale has a bottom element, but [0,1] has a top element as well).

One important direction for future work is the applicability of a scale-based approach to deontic and dynamic modals. As Portner (2009) notes, one of the main attractions of Kratzer's theory is that auxiliary modals such as *must* that can be epistemic or deontic are not ambiguous, but differ only in the value of a contextual parameter. It may be possible to use a theory closely related to the one proposed here at least for deontic modals, as Levinson (2003) does for desire verbs. If so, we could avoid postulating ambiguities by individually varying the parameters that define scales, e.g. the set of degrees or the "dimension" (height, weight, likelihood, etc.). Much remains to be done, but I believe that this direction is promising.

To sum up, we identified two desiderata for a theory of gradable modals: it should be compatible with a good theory of the semantics of gradable adjectives, and it should predict the right inferences for complex likelihood judgments. If gradable epistemic modals are probability operators, we can satisfy both of these requirements. The present theory is built directly on the theory of gradable adjectives in Kennedy & McNally (2005) and Kennedy (2007); and, at least with respect to the problematic inference-types discussed, it makes intuitively correct predictions. In these respects it compares favorably with the standard theory of gradable modality due to Kratzer. Furthermore, treating GEMs as denoting measure functions whose range is the closed scale [0,1] explains a striking range of similarities between GEMs and non-modal adjectives in the range of degree modifiers they take and entailments when negated.[5]

References

C. Kennedy (2007). 'Vagueness and grammar: The semantics of relative and absolute gradable adjectives'. *Linguistics and Philosophy* **30**(1):1–45.

C. Kennedy & L. McNally (2005). 'Scale structure, degree modification, and the semantics of gradable predicates'. *Language* **81**(2):345–381.

D. Krantz, et al. (1971). *Foundations of Measurement.* Academic Press.

A. Kratzer (1981). 'The notional category of modality'. In H.-J. Eikmeyer & H. Rieser (eds.), *Words, worlds, and contexts: new approaches in word semantics*, pp. 38–74. de Gruyter.

A. Kratzer (1991). 'Modality'. In A. von Stechow & D. Wunderlich (eds.), *Semantics: An International Handbook of Contemporary Research.* de Gruyter.

D. Lassiter (2010). 'Gradable Epistemic Modals, Probability, and Scale Structure'. To appear in *Proceedings of Semantics and Linguistic Theory XX* .

D. Levinson (2003). 'Probabilistic Model-theoretic Semantics for *want.*'. In *Proceedings of Semantics and Linguistic Theory XIII.*

P. Portner (2009). *Modality.* Oxford University Press.

S. Yalcin (2010). 'Probability Operators'. To appear in *Philosophy Compass* .

[5] This paper owes a great deal to Seth Yalcin, who gave me the idea to think about modals in the context of the theory of gradability. The idea for this paper emerged from discussion with him, and since his departure from NYU we have reached similar conclusions, mostly working independently, as can be seen by comparing this paper to the forthcoming (Yalcin 2010). Thanks also to Chris Barker, who has been extremely generous with time and comments, to Anna Szabolcsi, Philippe Schlenker, and Larry Horn, and to Julien Dutant and three anonymous ESSLLI reviewers, each of whom provided very helpful written comments.

Tableaux for the Lambek-Grishin Calculus

Arno Bastenhof

Utrecht University

Categorial type logics, pioneered by Lambek ([8]), seek a proof-theoretic understanding of natural language syntax by identifying categories with formulas and derivations with proofs. We typically observe an intuitionistic bias: a structural configuration of hypotheses (a constituent) derives a single conclusion (the category assigned to it). Acting upon suggestions of Grishin ([3]) to dualize the logical vocabulary, Moortgat proposed the *Lambek-Grishin calculus* (**LG**, [11]) with the aim of restoring symmetry between hypotheses and conclusions.

We propose a theory of *labeled* modal tableaux ([14]) for **LG**, inspired by the interpretation of its connectives as binary modal operators in the relational semantics of [6]. After a brief recapitulation of **LG**'s models in §1, we define our tableaux in §2 and ensure soundness and completeness in §3. Linguistic applications are considered in §4, where grammars based on **LG** are shown to be context-free through use of an interpolation lemma. This result complements [10], where **LG** augmented by mixed associativity and -commutativity was shown to exceed LTAG in expressive power.

1 Ternary frames and Lambek calculi

We discuss ternary frame semantics for **NL** and its symmetric generalization. More in-depth discussions of the presented material is found in [5] and [6].

Ternary frames $\mathscr{F}$ are pairs $\langle W, R\rangle$ with W an inhabited set of *resources*, and $R \subseteq W^3$ a *(ternary) accessibility relation.* Propositional variables (atoms) $p, q, r, \ldots$ are identified with subsets of W. Formally, a *model* $\mathscr{M} = \langle \mathscr{F}, V\rangle$ extends $\mathscr{F}$ with an *(atomic) valuation* V mapping atoms to $\mathscr{P}(W)$. Connectives for (multiplicative) conjunction and implication, constructing derived formulas $A, B, C, \ldots$, arise as *binary modal operators* by extending V as in

$$\begin{aligned} V(A \otimes B) &:= \{x \mid (\exists y, z)(Rxyz \text{ and } y \in V(A) \text{ and } z \in V(B))\} \quad \textit{(fusion)} \\ V(C/B) &:= \{y \mid (\forall x, z)((Rxyz \text{ and } z \in V(B)) \Rightarrow x \in V(C))\} \quad \textit{(right impl.)} \\ V(A\backslash C) &:= \{z \mid (\forall x, y)((Rxyz \text{ and } y \in V(A)) \Rightarrow x \in V(C))\} \quad \textit{(left impl.)} \end{aligned}$$

Kurtonina ([5]) explores linguistic applications. W contains syntactic constituents and $Rxyz$ reads as binary merger: x results from merging y with z. Thus, one would adopt atoms np (its image under V the collection of noun phrases), s (sentences) and n (common nouns), with subcategorization encoded by implications: $np\backslash s$ categorizes intransitive verbs, $(np\backslash s)/np$ transitive verbs, etc.

Proofs (or *algebraic derivations*) of inequalities $A \leq B$ are intended to establish $V(A) \subseteq V(B)$ for arbitrary $\langle \mathscr{F}, V\rangle$. On the linguistic reading, these are the *language universals*: any language categorizing an expression by A (e.g.,

$np \otimes (np\backslash s)$, the merger of a noun phrase and an intransitive verb) must also categorize it by B (s, as follows from the rules below). Next to the preorder axioms *(Refl, Trans)* on $\leq$, the set $\{\otimes, /, \backslash\}$ is *residuated* (r), with *parent* $\otimes$ and (left and right) *residuals* $\backslash, /$ (the double line indicates *inter*derivability)

$$\frac{}{A \leq A}\ \textit{Refl} \qquad \frac{A \leq B \quad B \leq C}{A \leq C}\ \textit{Trans} \qquad \frac{A \otimes B \leq C}{B \leq A\backslash C}\ r \qquad \frac{A \otimes B \leq C}{A \leq C/B}\ r$$

validity w.r.t. arbitrary models being easily verified. The distinguished status of fusion leads us to write the corresponding accessibility relation as $R_\otimes$ from now on. We arrive at what is known as the *non-associative Lambek calculus* (**NL**). Note that associativity and commutativity of $\otimes$ are not generally valid in a model, but rather depend on special frame constraints (cf. [5]):

Inequality	**Frame constraint** $(\forall a, b, c, y, x)$
$A \otimes (B \otimes C) \leq (A \otimes B) \otimes C$	$(R_\otimes xay$ and $R_\otimes ybc) \Rightarrow (\exists t)(R_\otimes xtc$ and $R_\otimes tab)$
$(A \otimes B) \otimes C \leq A \otimes (B \otimes C)$	$(R_\otimes xtc$ and $R_\otimes tab) \Rightarrow (\exists x)(R_\otimes xay$ and $R_\otimes ybc)$
$A \otimes B \leq B \otimes A$	$R_\otimes xab \Rightarrow R_\otimes xba$

Grishin ([3]) first suggested extending **NL** by a family of *co*residuated connectives $\{\oplus, \oslash, \obackslash\}$ with parent $\oplus$ (*fission*) and left- and right coresiduals $\obackslash, \oslash$ (*subtractions*), mirroring $\{\otimes, /, \backslash\}$ in $\leq$:

$$\frac{C \leq A \oplus B}{C \oslash B \leq A}\ cr \qquad \frac{C \leq A \oplus B}{A \obackslash C \leq B}\ cr$$

Moortgat names this the *Lambek-Grishin calculus* (**LG**) in [11]. In contrast with *classical* **NL** ([2]), **LG** does not internalize its duality with linear negation. Thus, we cannot simply interpret fission and subtraction as the De Morgan duals of fusion and implication. Instead, we have to consider frames $\mathscr{F} = \langle W, R_\otimes, R_\oplus\rangle$ with a second accessibility relation $R_\oplus \subseteq W^3$:

$$\begin{array}{l} x \in V(A \oplus B) \Leftrightarrow (\forall y, z)(R_\oplus xyz \Rightarrow (y \in V(A) \text{ or } z \in V(B))) \\ y \in V(C \oslash B) \Leftrightarrow (\exists x, z)(R_\oplus xyz \text{ and } z \in V(B) \text{ and } x \in V(C)) \\ z \in V(A \obackslash C) \Leftrightarrow (\exists x, y)(R_\oplus xyz \text{ and } y \in V(A) \text{ and } x \in V(C)) \end{array}$$

We conclude by mentioning previous work on the proof theory of **LG**, motivating our own tableau approach. First, a negative result: while Lambek ([9]) gave a sequent calculus for **NL**, extending it to **LG** by mirroring the inference rules sacrifices Cut admissibility.[1] Moortgat ([11]) instead defines a *display calculus* for **LG**, based on the observation that (algebraic) transitivity is admissible in the presence of (co)residuation and monotonicity:

$$\frac{A \leq B \quad C \leq D}{\begin{array}{c} A \otimes C \leq B \otimes D \\ A \oplus C \leq B \oplus D \end{array}} \quad \frac{A \leq B \quad C \leq D}{\begin{array}{c} A/D \leq B/C \\ D\backslash A \leq C\backslash B \end{array}} \quad \frac{A \leq B \quad C \leq D}{\begin{array}{c} A \oslash D \leq B \oslash C \\ D \obackslash A \leq C \obackslash B \end{array}}$$

[1] Bernardi and Moortgat give a(n unpublished) counterexample with the two-formula sequent $A \otimes (C \oslash ((A\backslash B) \obackslash C)) \vdash B$.

Our own approach to **LG** theorem proving is rather in the tradition of labeled modal tableaux, mixing the language of formulas with that of the models interpreting them. Equivalently, the old "turn your derivations upside-down" trick renders it as a *labeled* sequent calculus, representing by a single labeled sequent those display sequents of [11] that are interderivable by (co)residuation. Moreover, as Lemma 34 shows, Cut-admissibility is recovered.

2 A labeled tableau calculus for LG

Fix a denumerable collection of variables $x, y, z, \ldots$, to be thought of as a set W of resources. By a *signed formula* we understand a formula suffixed by $\cdot^\bullet$ or $\cdot^\circ$. We also speak of *input* formulas $A^\bullet$ and *output* formulas A°. A *labeled* signed formula pairs a signed formula with a variable. Intuitively, a pair $x : A^\bullet$ asserts A to be true at point x, whereas $y : B^\circ$ asserts B to be false at point y. We sometimes use meta-variables ϕ, ψ, ω, using the suffix $\cdot^\perp$ for switching signs: $\phi^\perp$ denotes $x : A^\circ$ if $\phi = x : A^\bullet$ and $x : A^\bullet$ if $\phi = x : A^\circ$.

Tableau rules operate on *boxes* $\boxed{\Theta|\Gamma}$, understood linguistically as encoding syntactic descriptions: phrase structure is specified by means of an *unrooted tree* Θ, with Γ defining a cyclic order on the words attached to its leaves. More specifically, Γ denotes a finite list of signed formulas (categorizing words) labeled by variables found at the leaves of Θ, such that 'provability' of a box $\boxed{\Theta|\Gamma}$ will be closed under cyclic permutations of Γ. We describe trees Θ by multisets of conditions $R_\otimes xyz$, $R_\oplus xyz$: each variable in (a condition of) Θ has its own node, any condition $R_\oplus xyz$ or $R_\otimes xyz$ in Θ introduces a fresh node with edges (precisely) to x, y, z, and any variable occurs at most twice.[2]

Trees Θ	x	$R_\otimes xyz$	$R_\oplus xyz$	$(\Theta' \cup \Theta'') - (N(\Theta') \cup N(\Theta''))$
Conditions	-	-	-	$N(\Theta') \cap N(\Theta'') = \{x\}$ $x \in C(\Theta'), x \in H(\Theta'')$
Nodes $N(\Theta)$	$\{x\}$	$\{x, y, z\}$	$\{x, y, z\}$	$N(\Theta') \cup N(\Theta'')$
Hypotheses $H(\Theta)$	$\{x\}$	$\{y, z\}$	$\{x\}$	$(H(\Theta') \cup H(\Theta''))/\{x\}$
Conclusions $C(\Theta)$	$\{x\}$	$\{x\}$	$\{y, z\}$	$(C(\Theta') \cup C(\Theta''))/\{x\}$

Thus, for any such 'tree' Θ, $\boxed{\Theta|\Gamma}$ is a box in case the hypotheses of Θ label input formulas of Γ, whereas its conclusions label output formulas. Note that our use of multiset difference in the definition of complex trees implies $x \in \Theta$ only if Θ is a singleton. The purpose of such trees $\{x\}$ is to guarantee well-definedness for the concepts $N(\Theta)$, $H(\Theta)$ and $C(\Theta)$ w.r.t. two-formula boxes $\boxed{x|x : A^\bullet, x : B^\circ}$. We shall often abbreviate $(\Theta \cup \Theta') - (N(\Theta) \cup N(\Theta'))$ by Θ, Θ' (in particular: $\Theta, \{x\} = \Theta$), and similarly write Γ, Δ for list concatenation.

Labeled signed formulas are classified into types α, β according to Smullyan's unified notation:

[2] Such structures previously appeared in the literature on (non-associative) proof nets as *tensor trees* in [12] and as *tree signatures* in [7], the latter building forth on [2]. In [5], a similar encoding of rooted trees by means of the accessibility relation $R_\otimes xyz$ was proposed for **NL**.

α	$\alpha_1(y)$	$\alpha_2(z)$	$R_\alpha yz$	β	$\beta_1(y)$	$\beta_2(z)$	$R_\beta yz$
$x:(A/B)^\circ$	$y:B^\bullet$	$z:A^\circ$	$R_\otimes zxy$	$x:(A/B)^\bullet$	$y:B^\circ$	$z:A^\bullet$	$R_\otimes zxy$
$x:(B\backslash A)^\circ$	$y:A^\circ$	$z:B^\bullet$	$R_\otimes zyx$	$x:(B\backslash A)^\bullet$	$y:A^\bullet$	$z:B^\circ$	$R_\otimes zyx$
$x:(A\otimes B)^\bullet$	$y:A^\bullet$	$z:B^\bullet$	$R_\otimes xyz$	$x:(A\otimes B)^\circ$	$y:A^\circ$	$z:B^\circ$	$R_\otimes xyz$
$x:(A\oslash B)^\bullet$	$y:A^\bullet$	$z:B^\circ$	$R_\oplus zxy$	$x:(A\oslash B)^\circ$	$y:A^\circ$	$z:B^\bullet$	$R_\oplus zxy$
$x:(B\obslash A)^\bullet$	$y:B^\circ$	$z:A^\bullet$	$R_\oplus zyx$	$x:(B\obslash A)^\circ$	$y:B^\bullet$	$z:A^\circ$	$R_\oplus zyx$
$x:(A\oplus B)^\circ$	$y:B^\circ$	$z:A^\circ$	$R_\oplus xyz$	$x:(A\oplus B)^\bullet$	$y:B^\bullet$	$z:A^\bullet$	$R_\oplus xyz$

Tableaux may then be expanded by either one of the following rules, of which the second is said to *branch*:[3]

$$\frac{\boxed{\Theta \,|\, \Gamma,\alpha,\Delta}}{\boxed{R_\alpha yz,\Theta \,|\, \Gamma,\alpha_1(y),\alpha_2(z),\Delta}}\ \alpha \qquad \frac{\boxed{R_\beta yz,\Theta,\Theta' \,|\, \Gamma,\Delta,\beta,\Gamma',\Delta'}}{\boxed{\Theta \,|\, \Gamma,\beta_1(y),\Gamma'} \quad \boxed{\Theta' \,|\, \Delta,\beta_2(z),\Delta'}}\ \beta$$

Here, in (α), y,z are to be fresh in the current branch, whereas for (β), either $\Gamma=\varnothing$ or $\Delta'=\varnothing$, and

$$\begin{array}{l} N(\Theta)\bigcap N(\Theta')=\varnothing,\\ N(R_\beta yz)\bigcap N(\Theta)=\{z\},\\ N(R_\beta yz)\bigcap N(\Theta')=\{y\}\end{array} \quad \text{and} \quad \begin{array}{|l|l|l|}\hline & \in H(R_\beta yz), & \in C(R_\beta yz),\\ \hline \text{If } y & \text{then } y\in C(\Theta') & \text{then } y\in H(\Theta')\\ \text{If } z & \text{then } z\in H(\Theta) & \text{then } z\in C(\Theta)\\ \hline\end{array}$$

A tableau branch ending in a box $\boxed{x \,|\, x:p^\bullet, x:p^\circ}$ or $\boxed{x \,|\, x:p^\circ, x:p^\bullet}$ is *closed*, and a tableau is closed if all its branches are. We also say $\boxed{\Theta \,|\, \Gamma}$ closes if it has a closed tableau. A tableau for a two-formula *sequent* $A\vdash B$ is a tableau of $\boxed{x \,|\, x:A^\bullet, x:B^\circ}$, called a *proof* of $A\vdash B$ if it closes. An easy induction establishes

Lemma 21. The property of having a closed tableau (of a box) is preserved under renaming of variables.

Say a tableau of $\boxed{\Theta \,|\, \Gamma}$ *closes via* $\phi\in\Gamma$ if the first expansion immediately targets ϕ, and let the *degree* of a formula A denote the number of connectives in A.

Lemma 22. *Cyclic permutation* is admissible: if we have a closed tableau $\mathscr{T}$ of $\boxed{\Theta \,|\, \Gamma,\Delta}$, then $\boxed{\Theta \,|\, \Delta,\Gamma}$ also closes.

Proof. By induction on the (combined) degree of (the formulas in) Γ,Δ.

1. $\boxed{\Theta \,|\, \Gamma,\Delta}=\boxed{x \,|\, x:p^\bullet, x:p^\circ}$ or $\boxed{\Theta \,|\, \Gamma,\Delta}=\boxed{x \,|\, x:p^\circ, x:p^\bullet}$: immediate.
2. $\mathscr{T}$ closes via some $\alpha\in\Gamma,\Delta$. Say $\alpha\in\Gamma$, i.e., $\Gamma=\Gamma',\alpha,\Gamma''$:

$$\frac{\boxed{\Theta \,|\, \Gamma',\alpha,\Gamma'',\Delta}}{\boxed{R_\alpha yz,\Theta \,|\, \Gamma',\alpha_1(y),\alpha_2(z),\Gamma'',\Delta}}\ \alpha$$

Then, by the induction hypothesis, $\boxed{R_\alpha yz,\Theta \,|\, \Delta,\Gamma',\alpha_1(y),\alpha_2(z),\Gamma''}$ also closes, so that the statement of the lemma now obtains by another α-expansion.

[3] The current formulation may be considered a labeling of Abrusci's sequent calculus for cyclic linear logic in [1], where cyclic permutations were compiled away into the logical inferences.

3. $\mathscr{T}$ closes via some $\beta \in \Gamma, \Delta$. I.e., $\Gamma, \Delta = \Gamma_1, \Gamma_2, \beta, \Gamma_1', \Gamma_2'$ and $\Theta = R_\beta yz, \Theta, \Theta'$, with, for example, $\Gamma_2' = \Delta_1, \Delta_1'$, $\Gamma = \Gamma_1, \Gamma_2, \beta, \Gamma_1', \Delta_1$ and $\Delta = \Delta_1'$:

$$\dfrac{\boxed{R_\beta yz, \Theta, \Theta' \mid \Gamma_1, \Gamma_2, \beta, \Gamma_1', \Delta_1, \Delta_1'}}{\boxed{\Theta \mid \Gamma_1, \beta_1(y), \Gamma_1'} \qquad \boxed{\Theta' \mid \Gamma_2, \beta_2(z), \Delta_1, \Delta_1'}}\ \beta$$

By the induction hypothesis, $\boxed{\Theta' \mid \Delta_1', \Gamma_2, \beta_2(z), \Delta_1}$ closes, so that another β-expansion suffices:

$$\dfrac{\boxed{R_\beta yz, \Theta, \Theta' \mid \Gamma_1, \Delta_1', \Gamma_2, \beta, \Gamma_1', \Delta_1}}{\boxed{\Theta \mid \Gamma_1, \beta_1(y), \Gamma_1'} \qquad \boxed{\Theta' \mid \Delta_1', \Gamma_2, \beta_2(z), \Delta_1}}\ \beta$$

noting that if $\Gamma_1 \neq \varnothing$, then $\Delta (= \Delta_1, \Delta_1') = \varnothing$.

Example 23. We have a proof of $p \otimes (r \oslash ((p\backslash q) \obslash r)) \vdash q$, which served as a counterexample to Cut elimination in an earlier sequent calculus for **LG**.

$$\dfrac{\boxed{x \mid x : (p \otimes (r \oslash ((p\backslash q) \obslash r)))^\bullet, x : q^\circ}}{\dfrac{\boxed{R_\oplus uzv, R_\otimes xyz \mid y : p^\bullet, u : r^\bullet, v : ((p\backslash q) \obslash r)^\circ, x : q^\circ}}{\dfrac{\boxed{R_\otimes xyz \mid y : p^\bullet, z : (p\backslash q)^\bullet, x : q^\circ}}{\boxed{x \mid x : q^\bullet, x : q^\circ} \qquad \boxed{y \mid y : p^\bullet, y : p^\circ}}\ \beta \qquad \boxed{u \mid u : r^\bullet, u : r^\circ}}\ \beta}\ \alpha(\times 2)$$

Example 24. We have previously understood boxes as encodings for syntactic descriptions. We further illustrate this claim by representing the derivation of a simple transitive clause by a closed tableau. Consider the following lexicon for *He saw Pete*, consisting of a pairing of words with signed formulas:

he	saw	Pete
$(s/(np\backslash s))^\bullet$	$((np\backslash s)/np)^\bullet$	$np^\bullet$

The formula $(s/(np\backslash s))^\bullet$ for *he* was proposed by Lambek ([8]) in order to exclude occurrences in object positions. Grammaticality of the sentence under consideration w.r.t. a *goal* (signed) formula s° is now established by a closed tableau

$$\dfrac{\boxed{R_\otimes xhz, R_\otimes zwp \mid h : (s/(np\backslash s))^\bullet, w : ((np\backslash s)/np)^\bullet, p : np^\bullet, x : s^\circ}}{\boxed{p \mid p : np^\circ, p : np^\bullet} \qquad \dfrac{\boxed{R_\otimes xhz \mid h : (s/(np\backslash s))^\bullet, z : (np\backslash s)^\bullet, x : s^\circ}}{\dfrac{\boxed{z \mid z : (np\backslash s)^\circ, z : (np\backslash s)^\bullet}}{\dfrac{\boxed{R_\otimes uvz \mid u : s^\circ, v : np^\bullet, z : (np\backslash s)^\bullet}}{\boxed{u \mid u : s^\circ, u : s^\bullet} \qquad \boxed{v \mid v : np^\bullet, v : np^\circ}}\ \beta}\ \alpha \qquad \boxed{x \mid x : s^\bullet, x : s^\circ}}\ \beta}\ \beta$$

We note that, in **LG**, nothing prevents us from coupling words with output formulas. For example, the following lexicon would do just as well:

he	saw	Pete
$((np\backslash s) \oslash s)^\circ$	$(np \oslash (np\backslash s))^\circ$	$np^\bullet$

as witnessed by the tableau

$$\dfrac{\boxed{R_\oplus yhx, R_\oplus pwy \mid h : ((np\backslash s) \oslash s)^\circ, w : (np \oslash (np\backslash s))^\circ, p : np^\bullet, x : s^\circ}}{\dfrac{\boxed{R_\oplus pwy \mid y : (np\backslash s)^\circ, w : (np \oslash (np\backslash s))^\circ, p : np^\bullet}}{\dfrac{\boxed{R_\otimes uvy, R_\oplus pwy \mid u : s^\circ, v : np^\bullet, w : (np \oslash (np\backslash s))^\circ, p : np^\bullet}}{\dfrac{\boxed{R_\otimes uvy \mid u : s^\circ, v : np^\bullet, y : (np\backslash s)^\circ}}{\boxed{u \mid u : s^\circ, u : s^\bullet} \qquad \boxed{v \mid v : np^\bullet, v : np^\circ}}\,\beta \qquad \boxed{p \mid p : np^\circ, p : np^\bullet}}\,\beta}\,\alpha \qquad \boxed{x \mid x : s^\bullet, x : s^\circ}}\,\beta$$

3 Soundness and completeness

Let S be a finite set of labeled signed formulas and conditions $R_\otimes xyz, R_\oplus xyz$. An *interpretation* for S is a pair $I = \langle \mathscr{M}, \cdot^* \rangle$ with $\cdot^*$ a mapping of the variables occurring in S to the resources of $\mathscr{M}$. *Truth* w.r.t. I is defined by

1. $R_\delta xyz \in S$ ($\delta \in \{\otimes, \oplus\}$) is true w.r.t. I in case $R_\delta x^* y^* z^*$ in $\mathscr{M}$.
2. $x : A^\bullet \in S$ is true w.r.t. I if $x^* \in V(A)$ and false if $x^* \notin V(A)$
3. $y : B^\circ \in S$ is true w.r.t. I if $x^* \notin V(A)$ and false if $x^* \in V(A)$

Call S *satisfiable* if for some interpretation I, all elements of S are true w.r.t. I. The following observation, made w.r.t. arbitrary I, implies Lemma 31:[4]

$$\begin{array}{l} \alpha \text{ is true} \Leftrightarrow R_\alpha yz,\ \alpha_1(y) \text{ and } \alpha_2(z) \text{ are true for some } y, z \\ \beta \text{ is true} \Leftrightarrow R_\beta yz \text{ implies } \beta_1(y) \text{ or } \beta_2(z), \text{ for arbitrary } y, z \end{array}$$

Lemma 31. For any set S of labeled signed formulas and conditions $R_\otimes xyz, R_\oplus xyz$,

(a) If S is satisfiable and $\alpha \in S$, then for fresh y, z so is $S \bigcup \{R_\alpha yz, \alpha_1(y), \alpha_2(z)\}$
(b) If S is satisfiable and $\beta, R_\beta yz \in S$, then so is $S \bigcup \{\beta_1(y)\}$ or $S \bigcup \{\beta_2(z)\}$

Given a branch θ in a tableau, collect the elements of the Θ, Γ for each box $\boxed{\Theta \mid \Gamma}$ occurring in it in a single set S_θ (save for when $\Theta = \{x\}$). θ is satisfiable if S_θ is, and any tableau is satisfiable if one of its branches is. Lemma 31 implies

Theorem 32. If a tableau $\mathscr{T}$ is satisfiable, and $\mathscr{T}'$ is obtained from $\mathscr{T}$ by a single expansion, then $\mathscr{T}'$ is satisfiable.

The unsatisfiability of a closed tableau is now traced to its origin. Hence, provability of $A \vdash B$ means unsatisfiability of $\boxed{\varnothing \mid x : A^\bullet, x : B^\circ}$, yielding soundness:

Corollary 33. All models validate provable two-formula sequents $A \vdash B$.

For completeness, it suffices to show that we can simulate algebraic derivations:

$$\dfrac{}{A \le A}\,\mathit{Refl} \qquad \dfrac{A \le B \quad B \le C}{A \le C}\,\mathit{Trans} \qquad \begin{array}{c} A \le C/B \\ \hline\hline A \otimes B \le C \\ \hline\hline B \le A\backslash C \end{array}\,\begin{array}{c} r \\ r \end{array} \qquad \begin{array}{c} C \oslash B \le A \\ \hline\hline C \le A \oplus B \\ \hline\hline A \obslash C \le B \end{array}\,\begin{array}{c} cr \\ cr \end{array}$$

[4] These conditions are easily seen to be classically equivalent to those provided in §2.

already shown complete in [6]. That, for any A, a closed tableau of $\boxed{\varnothing}\boxed{x : A^\bullet, x : A^\circ}$ exists is a simple induction on A's degree. The following lemma tackles (*Trans*).

Lemma 34. The following expansion (*bivalence*) is admissible for closed tableaux $\mathscr{T}_1$ and $\mathscr{T}_2$ of $\boxed{\Theta}\boxed{\Gamma, \phi, \Gamma'}$ and $\boxed{\Theta'}\boxed{\Delta, \phi^\perp, \Delta'}$

$$\dfrac{\boxed{\Theta, \Theta'}\boxed{\Gamma, \Delta, \Gamma', \Delta'}}{\boxed{\Theta}\boxed{\Gamma, \phi, \Gamma'} \qquad \boxed{\Theta'}\boxed{\Delta, \phi^\perp, \Delta'}}\ \mathrm{B}$$

provided $N(\Theta) \cap N(\Theta') = \{u\}$, u being the label of $\phi, \phi^\perp$, and $\Gamma = \varnothing$ or $\Delta' = \varnothing$.

Proof. We proceed by induction on the degree of $\Gamma, \Delta, \Gamma', \Delta', \phi$.

1. One of $\boxed{\Theta}\boxed{\Gamma, \phi, \Gamma'}$ or $\boxed{\Theta'}\boxed{\Delta, \phi^\perp, \Delta'}$ equals $\boxed{x}\boxed{x : p^\bullet, x : p^\circ}$ or $\boxed{x}\boxed{x : p^\circ, x : p^\bullet}$. Immediate, save for cases like the following, where we apply Lemma 22.

$$\dfrac{\boxed{\Theta}\boxed{\Gamma', u : p^\circ}}{\boxed{\Theta}\boxed{u : p^\circ, \Gamma'} \qquad \boxed{u}\boxed{u : p^\bullet, u : p^\circ}}\ B$$

2. $\mathscr{T}_1$ does not close via ϕ. Suppose $\mathscr{T}_1$ closes via α. For example, $\Gamma = \Gamma_1, \alpha, \Gamma_1'$:

$$\dfrac{\boxed{\Theta, \Theta'}\boxed{\Gamma_1, \alpha, \Gamma_1', \Delta, \Gamma', \Delta'}}{\dfrac{\boxed{\Theta}\boxed{\Gamma_1, \alpha, \Gamma_1', \phi, \Gamma'}}{\boxed{R_\alpha yz, \Theta}\boxed{\Gamma_1, \alpha_1(y), \alpha_2(z), \Gamma_1', \phi, \Gamma'}}\ \alpha \qquad \boxed{\Theta'}\boxed{\Delta, \phi^\perp, \Delta'}}\ \mathrm{B}$$

Permuting B over α reduces the induction measure:

$$\dfrac{\dfrac{\boxed{\Theta, \Theta'}\boxed{\Gamma_1, \alpha, \Gamma_1', \Delta, \Gamma', \Delta'}}{\boxed{R_\alpha yz, \Theta, \Theta'}\boxed{\Gamma_1, \alpha_1(y), \alpha_2(z), \Gamma_1', \Delta, \Gamma', \Delta'}}\ \alpha}{\boxed{R_\alpha yz, \Theta}\boxed{\Gamma_1, \alpha_1(y), \alpha_2(z), \Gamma_1', \phi, \Gamma'} \qquad \boxed{\Theta'}\boxed{\Delta, \phi^\perp, \Delta'}}\ \mathrm{B}$$

Otherwise, $\mathscr{T}_1$ closes via β. For example, $\Gamma = \Gamma_1, \Gamma_2, \beta, \Gamma_1'$ (in which case Δ' must be empty), $\Gamma' = \Gamma_1'', \Gamma_2'$ and $\Theta = \Theta_1, \Theta_2$:

$$\dfrac{\boxed{R_\beta yz, \Theta_1, \Theta_2, \Theta'}\boxed{\Gamma_1, \Gamma_2, \beta, \Gamma_1', \Delta, \Gamma_1'', \Gamma_2'}}{\dfrac{\boxed{R_\beta yz, \Theta_1, \Theta_2}\boxed{\Gamma_1, \Gamma_2, \beta, \Gamma_1', \phi, \Gamma_1'', \Gamma_2'}}{\boxed{\Theta_1}\boxed{\Gamma_1, \beta_1(y), \Gamma_1', \phi, \Gamma_1''} \qquad \boxed{\Theta_2}\boxed{\Gamma_2, \beta_2(z), \Gamma_2'}}\ \beta \qquad \boxed{\Theta'}\boxed{\Delta, \phi^\perp}}\ \mathrm{B}$$

Permuting (B) with (β) reduces the induction measure:

$$\dfrac{\boxed{R_\beta yz, \Theta_1, \Theta_2, \Theta'}\boxed{\Gamma_1, \Gamma_2, \beta, \Gamma_1', \Delta, \Gamma_1'', \Gamma_2'}}{\dfrac{\boxed{\Theta_1, \Theta'}\boxed{\Gamma_1, \beta_1(y), \Gamma_1', \Delta, \Gamma_1''}}{\boxed{\Theta_1}\boxed{\Gamma_1, \beta_1(y), \Gamma_1', \phi, \Gamma_1''} \qquad \boxed{\Theta'}\boxed{\Delta, \phi^\perp}}\ \mathrm{B} \qquad \boxed{\Theta_2}\boxed{\Gamma_2, \beta_2(z), \Gamma_2'}}\ \beta$$

3. $\mathscr{T}_2$ does not close via $\phi^\perp$. Similar to case (2).

4. $\mathscr{T}_1$ and $\mathscr{T}_2$ close via ϕ and $\phi^\perp$ respectively. Say ϕ is a β, in which case $\phi^\perp$ is an α. Then $\Theta = R_\beta yz, \Theta_1, \Theta_2$, $\Gamma = \Gamma_1, \Gamma_2$ and $\Gamma' = \Gamma_1', \Gamma_2'$:

$$\dfrac{\boxed{R_\beta yz, \Theta_1, \Theta_2, \Theta' \mid \Gamma_1, \Gamma_2, \Delta, \Gamma_1', \Gamma_2', \Delta'}}{\dfrac{\boxed{R_\beta yz, \Theta_1, \Theta_2 \mid \Gamma_1, \Gamma_2, \beta, \Gamma_1', \Gamma_2'}}{\boxed{\Theta_1 \mid \Gamma_1, \beta_1(y), \Gamma_1'} \quad \boxed{\Theta_2 \mid \Gamma_2, \beta_2(z), \Gamma_2'}}\,\beta \qquad \dfrac{\boxed{\Theta' \mid \Delta, \alpha, \Delta'}}{\boxed{R_\alpha yz, \Theta' \mid \Delta, \alpha_1(y), \alpha_2(z), \Delta'}}\,\alpha}\,\mathrm{B}$$

We invoke the induction hypothesis twice by replacing with B-expansions on $\beta_1(u), \alpha_1(u)$ and $\beta_2(v), \alpha_2(v)$, each of lower degree:

$$\dfrac{\boxed{R_\alpha yz, \Theta_1, \Theta_2, \Theta' \mid \Gamma, \Gamma_2, \Delta, \Gamma_1', \Gamma_2', \Delta'}}{\boxed{\Theta_1 \mid \Gamma_1, \beta_1(y), \Gamma_1'} \qquad \dfrac{\boxed{R_\alpha yz, \Theta_2, \Theta' \mid \Gamma_2, \Delta, \alpha_1(y), \Gamma_2', \Delta'}}{\boxed{\Theta_2 \mid \Gamma_2, \beta_2(z), \Gamma_2'} \quad \boxed{R_\alpha yz, \Theta' \mid \Delta, \alpha_1(y), \alpha_2(z), \Delta'}}\,\mathrm{B}}\,\mathrm{B}$$

Simulation of transitivity immediately follows. Lemma 34 also applies in showing (co)residuation derivable. For example, suppose we have a closed tableau of $\boxed{z \mid z : C^\bullet, z : (A \oplus B)^\circ}$. Then for some (fresh) y, $\boxed{y \mid y : (A \oslash C)^\bullet, y : B^\circ}$ also closes:

$$\dfrac{\dfrac{\boxed{y \mid y : (A \oslash C)^\bullet, y : B^\circ}}{\boxed{R_\oplus zxy \mid x : A^\circ, z : C^\bullet, y : B^\circ}}\,\alpha}{\dfrac{\boxed{R_\oplus zxy \mid x : A^\circ, z : (A \oplus B)^\bullet, y : B^\circ}}{\boxed{y \mid y : B^\bullet, y : B^\circ} \quad \boxed{x \mid x : A^\circ, x : A^\bullet}}\,\beta \qquad \boxed{z \mid z : C^\bullet, z : (A \oplus B)^\circ}}\,\mathrm{B}$$

The above observations imply

Theorem 35. The tableau method for **LG** is complete.

The following is now an easy consequence of the subformula property:

Corollary 36. **LG** conservatively extends **NL**.

4 Lambek-Grishin grammars are context-free

We use our tableau method to establish context-freeness of Lambek-Grishin grammars. Following the strategy laid out in [13] and [4], we rely on an interpolation property proven in Lemma 41.

By an **LG** *grammar* $\mathscr{G}$ we shall understand a tuple $\langle \mathscr{A}, L, g^\diamond \rangle$ consisting of: a set of *words* $\mathscr{A}$; a *lexicon* L mapping words to (finite) sets of signed (!) formulas; and a signed atomic *goal* formula $g^\diamond$ ($\diamond \in \{\bullet, \circ\}$). The *language* $\mathscr{L}(\mathscr{G})$ recognized by $\mathscr{G}$ we then define by the set of lists $w_1, \ldots, w_n$ of words $w_i \in \mathscr{A}$ $(1 \leq i \leq n)$ such that, for some $A_1^{\diamond 1} \in L(w_1), \ldots, A_n^{\diamond n} \in L(w_n)$ $(\diamond 1, \ldots, \diamond n \in \{\bullet, \circ\})$ and tree Θ, $\boxed{\Theta \mid x_1 : A_1^{\diamond 1}, \ldots, x_n : A_n^{\diamond n}, x : g^\diamond}$ closes.

We proceed to show context-freeness of **LG** grammars. Recognizability of context-free languages is a consequence of Kandulski's results for **NL** and Corollary 36. Our strategy for showing that every **LG** grammar also has an equivalent context-free grammar follows closely that of [4], inspired in turn by [13]. We first prove an interpolation property for our tableaux.

Lemma 41. Suppose $\boxed{\Theta,\Theta' \mid \Gamma,\Delta,\Gamma'}$ closes s.t. $N(\Theta)\cap N(\Theta') = \{u\}$, and the variables in Γ and Γ' (Δ) draw from Θ (Θ'). Then for some $\phi = u : C^{\diamond}$, with $\diamond \in \{\bullet,\circ\}$ depending on whether $u \in H(\Theta)$ or $u \in C(\Theta)$, and with C a subformula of (a formula in) Γ,Δ,Γ', $\boxed{\Theta \mid \Gamma,\phi,\Gamma'}$ and $\boxed{\Theta' \mid \Delta,\phi^{\perp}}$ close.

Proof. We refer to ϕ and C interchangeably as the *witness* for Δ (borrowing terminology from [4]). We proceed by induction on the degree of Γ,Δ,Γ'. If $S = \boxed{\Theta,\Theta' \mid \Gamma,\Delta,\Gamma'}$ is already of the form $\boxed{u \mid u:p^{\bullet},u:p^{\circ}}$ or $\boxed{u \mid u:p^{\circ},u:p^{\bullet}}$, take $C = p$. Otherwise, the tableau $\mathscr{T}$ for S closes via some ψ in Γ, Γ' or Δ.

1. $\phi \in \Delta$. If $\psi = \alpha$, $\Delta = \Delta_1,\alpha,\Delta_2$ with $\mathscr{T}$ taking the form

$$\frac{\boxed{\Theta,\Theta' \mid \Gamma,\Delta_1,\alpha,\Delta_2,\Gamma'}}{\boxed{R_\alpha yz,\Theta,\Theta' \mid \Gamma,\Delta_1,\alpha_1(y),\alpha_2(z),\Delta_2,\Gamma'}}\,\alpha$$

Apply the induction hypothesis to obtain a witness ω for $\Delta_1,\alpha_1(y),\alpha_2(z),\Delta_2$, i.e., $\boxed{\Theta \mid \Gamma,\omega,\Gamma'}$ and $\boxed{R_\alpha yz,\Theta' \mid \Delta_1,\alpha_1(y),\alpha_2(z),\Delta_2,\omega^{\perp}}$ close. We can take $\phi = \omega$. Indeed, we obtain a closed tableau for $\boxed{\Theta' \mid \Delta_1,\alpha,\Delta_2,\omega^{\perp}}$ by an α-expansion:

$$\frac{\boxed{\Theta' \mid \Delta_1,\alpha,\Delta_2,\omega^{\perp}}}{\boxed{R_\alpha yz,\Theta' \mid \Delta_1,\alpha_1(y),\alpha_2(z),\Delta_2,\omega^{\perp}}}\,\alpha$$

If $\psi = \beta$, we must consider two subcases. If $\Delta = \psi$, then $\Theta' = \{u\}$ and we may take $\phi = \psi^{\perp}$. Otherwise, $\Theta' = R_\beta yz,\Theta_1,\Theta_2$ ($N(\Theta_1)\cap N(\Theta_2) = \varnothing$, $N(\Theta_1)\cap N(R_\beta yz) = y$ and $N(\Theta_2)\cap N(R_\beta yz) = \{z\}$), and either $u \in N(\Theta_1)$ or $u \in N(\Theta_2)$. In the former case, $\Delta = \Delta_1,\Delta_2,\beta,\Delta_3$ with $\mathscr{T}$ taking the form

$$\frac{\boxed{R_\beta yz,\Theta,\Theta_1,\Theta_2 \mid \Gamma,\Delta_1,\Delta_2,\beta,\Delta_3,\Gamma'}}{\boxed{\Theta,\Theta_1 \mid \Gamma,\Delta_1,\beta_1(y),\Delta_3,\Gamma'} \qquad \boxed{\Theta_2 \mid \Delta_2,\beta_2(z)}}\,\beta$$

Apply the induction hypothesis to find a witness ω for $\Delta_1,\beta_1(y),\Delta_3$, i.e., so that $\boxed{\Theta \mid \Gamma,\omega,\Gamma'}$ and $\boxed{\Theta_1 \mid \Delta_1,\beta_1(y),\Delta_3,\omega^{\perp}}$ close. We may take $\phi = \omega$. Indeed, a closed tableau for $\boxed{\Theta_1,\Theta_2 \mid \Delta_1,\Delta_2,\beta_2(z),\Delta_3,\omega^{\perp}}$ is found after a β-expansion:

$$\frac{\boxed{\Theta_1,\Theta_2 \mid \Delta_1,\Delta_2,\beta_2(z),\Delta_3,\omega^{\perp}}}{\boxed{\Theta_1 \mid \Delta_1,\beta_1(y),\Delta_3,\omega^{\perp}} \qquad \boxed{\Theta_2 \mid \Delta_2,\beta_2(z)}}\,\beta$$

If instead $u \in N(\Theta_2)$, then $\Delta = \Delta_1,\beta,\Delta_2,\Delta_3$ and $\mathscr{T}$ takes the form

$$\frac{\boxed{\Theta_1,\Theta_2 \mid \mathit{\Gamma},\mathit{\Delta}_1,\beta_2(z),\mathit{\Delta}_2,\mathit{\Delta}_3,\mathit{\Gamma}'}}{\boxed{\Theta_1 \mid \beta_1(y),\mathit{\Delta}_2} \qquad \boxed{\Theta,\Theta_2 \mid \mathit{\Gamma},\mathit{\Delta}_1,\beta_2(z),\mathit{\Delta}_3,\mathit{\Gamma}'}}\ \beta$$

This time, apply the induction hypothesis on $\boxed{\Theta,\Theta_2 \mid \mathit{\Gamma},\mathit{\Delta}_1,\beta_2(z),\mathit{\Delta}_3,\mathit{\Gamma}'}$.

2. $\psi \in \mathit{\Gamma}$. The case where $\psi = \alpha$ is easy (similar to when $\psi = \alpha \in \mathit{\Delta}$). So suppose $\psi = \beta$. If $\mathit{\Gamma} = \psi$, then $\Theta' = R_\beta yz,\Theta_1,\Theta_2$, $\mathit{\Delta} = \mathit{\Delta}_1,\mathit{\Delta}_2$ with $\mathscr{T}$ taking the form

$$\frac{\boxed{R_\beta yz,\Theta_1,\Theta_2 \mid \beta,\mathit{\Delta}_1,\mathit{\Delta}_2}}{\boxed{\Theta_1 \mid \beta_1(y),\mathit{\Delta}_1} \qquad \boxed{\Theta_2 \mid \beta_2(z),\mathit{\Delta}_2}}\ \beta$$

Note that $\mathit{\Gamma}' = \varnothing$ as $u \notin N(\Theta_1) \bigcup N(\Theta_2)$, so also $\Theta = \{u\}$. Evidently, we may take $\phi = \psi$. Now suppose $\Theta = R_\beta yz,\Theta_1,\Theta_2$ with $N(\Theta_1) \bigcap N(\Theta_2) = \varnothing$, $N(\Theta_1) \bigcap N(R_\beta yz) = \{y\}$ and $N(\Theta_2) \bigcap N(R_\beta yz) = \{z\}$. We must consider the cases $u \in N(\Theta_1)$ and $u \in N(\Theta_2)$. We consider the former, the latter being handled similarly (although then $\mathit{\Gamma}' = \varnothing$). Now $\mathit{\Gamma} = \mathit{\Gamma}_1,\mathit{\Gamma}_2,\beta,\mathit{\Gamma}_3$ and $\mathit{\Gamma}' = \mathit{\Gamma}_1',\mathit{\Gamma}_2'$ such that $\mathscr{T}$ takes the form

$$\frac{\boxed{R_\beta yz,\Theta_1,\Theta_2,\Theta' \mid \mathit{\Gamma}_1,\mathit{\Gamma}_2,\beta,\mathit{\Gamma}_3,\mathit{\Delta},\mathit{\Gamma}_1',\mathit{\Gamma}_2'}}{\boxed{\Theta_1,\Theta \mid \mathit{\Gamma}_1,\beta_1(y),\mathit{\Gamma}_3,\mathit{\Delta},\mathit{\Gamma}_1'} \qquad \boxed{\Theta_2 \mid \mathit{\Gamma}_2,\beta_2(z),\mathit{\Gamma}_2'}}\ \beta$$

and we may apply the induction hypothesis on $\boxed{\Theta_1,\Theta \mid \mathit{\Gamma}_1,\beta_1(y),\mathit{\Gamma}_3,\mathit{\Delta},\mathit{\Gamma}_1'}$ to find an ω for which $\boxed{\Theta_1 \mid \mathit{\Gamma}_1,\beta_1(y),\mathit{\Gamma}_3,\omega,\mathit{\Gamma}_1'}$ and $\boxed{\Theta' \mid \mathit{\Delta},\omega^\perp}$ close. We take $\phi = \omega$. Indeed, we find a closed tableau for $\boxed{R_\beta yz,\Theta_1,\Theta_2 \mid \mathit{\Gamma},\omega,\mathit{\Gamma}'}$ as follows:

$$\frac{\boxed{R_\beta yz,\Theta_1,\Theta_2 \mid \mathit{\Gamma}_1,\mathit{\Gamma}_2,\beta,\mathit{\Gamma}_3,\omega,\mathit{\Gamma}_1',\mathit{\Gamma}_2'}}{\boxed{\Theta_1 \mid \mathit{\Gamma}_1,\beta_1(y),\mathit{\Gamma}_3,\omega,\mathit{\Gamma}_1'} \qquad \boxed{\Theta_2 \mid \mathit{\Gamma}_2,\beta_2(z),\mathit{\Gamma}_2'}}\ \beta$$

3. $\phi \in \mathit{\Gamma}'$. Similar to the previous case.

Lemma 42. For T a set of formulas, closed under taking subformulas, define

$$\begin{aligned}
\mathbf{LG}_T =_{def} & \{S = \boxed{x \mid x : A^\bullet, B^\circ} \mid A,B \in T \ \&\ S \text{ closes}\}\\
\bigcup\ & \{S = \boxed{x \mid x : B^\circ, x : A^\bullet} \mid A,B \in T \ \&\ S \text{ closes}\}\\
\bigcup\ & \{S = \boxed{R_\otimes xyz \mid y : A^\bullet, z : B^\bullet, x : C^\circ} \mid A,B \in T \ \&\ S \text{ closes}\}\\
\bigcup\ & \{S = \boxed{R_\otimes xyz \mid x : C^\circ, y : A^\bullet, z : B^\bullet} \mid A,B \in T \ \&\ S \text{ closes}\}\\
\bigcup\ & \{S = \boxed{R_\otimes xyz \mid z : B^\bullet, x : C^\circ, y : A^\bullet} \mid A,B \in T \ \&\ S \text{ closes}\}\\
\bigcup\ & \{S = \boxed{R_\oplus xyz \mid y : A^\circ, z : B^\circ, x : C^\bullet} \mid A,B \in T \ \&\ S \text{ closes}\}\\
\bigcup\ & \{S = \boxed{R_\oplus xyz \mid x : C^\bullet, y : A^\circ, z : B^\circ} \mid A,B \in T \ \&\ S \text{ closes}\}\\
\bigcup\ & \{S = \boxed{R_\oplus xyz \mid z : B^\circ, x : C^\bullet, y : A^\circ} \mid A,B \in T \ \&\ S \text{ closes}\}
\end{aligned}$$

Now suppose $\boxed{\Theta \mid \mathit{\Gamma}}$ closes with all formulas of $\mathit{\Gamma}$ in T. Then $\boxed{\Theta \mid \mathit{\Gamma}}$ has a tableau whose branches end in members of $\mathbf{LG}_T$ and with the following instance of bivalence (B) as the sole type of expansion, provided $\mathit{\Gamma}'$ is not empty.

$$\frac{\boxed{\Theta,\Theta' \mid \mathit{\Gamma},\mathit{\Delta},\mathit{\Gamma}'}}{\boxed{\Theta \mid \mathit{\Gamma},\phi,\mathit{\Gamma}'} \qquad \boxed{\Theta' \mid \mathit{\Delta},\phi^\perp}}\ \mathrm{B}$$

Proof. By induction on the cardinality of Θ. In the base case, Θ equals $\{x\}$, $\{R_{\otimes}xyz\}$ or $\{R_{\oplus}xyz\}$, and $\boxed{\Theta|\Gamma} \in \mathbf{LG}_T$ by definition. Now suppose $\Theta = \Theta_1, \Theta_2$, $N(\Theta_1) \cap N(\Theta_2) = \{u\}$, both $\Theta_1 \neq \{u\}$ and $\Theta_2 \neq \{u\}$, $\Gamma = \Gamma_1, \Gamma_2, \Gamma_3$ and the variables in Γ_1 and Γ_3 (Γ_2) draw from Θ_1 (Θ_2). By Lemma 41, there now exists ϕ with label u s.t. $\boxed{\Theta_1|\Gamma_1, \phi, \Gamma_3}$ and $\boxed{\Theta_2|\Gamma_2, \phi^{\perp}}$ close. We can assume Γ_3 is not empty, as otherwise we could have picked Γ_1 for instantiating Δ in (B) as opposed to Γ_2. Since the cardinalities of Θ_1 and Θ_2 are strictly smaller than that of Θ_1, Θ_2 (as $\Theta_2 \neq \{u\}$), the induction hypothesis applies to $\boxed{\Theta_1|\Gamma_1, \phi, \Gamma_3}$ and $\boxed{\Theta_2|\Gamma_2, \phi^{\perp}}$. The statement of the lemma obtains after another application of (B).

Theorem 43. For every **LG**-grammar $\mathscr{G}$, $\mathscr{L}(\mathscr{G})$ is context-free.

Proof. Suppose we have an **LG**-grammar $\mathscr{G}_1 = \langle \mathscr{A}, L, g^{\diamond} \rangle$. Refer by T to the set of formulas in the range of L, closed under taking subformulas. We now construct the following context-free grammar $\mathscr{G}_2$: its set of terminals coincides with $\mathscr{A}$; its nonterminals are specified by $\{A^{\bullet} \mid A \in T\} \cup \{A^{\circ} \mid A \in T\}$; its start symbol is $g^{\diamond\perp}$ for $\diamond\perp = \bullet$ if $\diamond = \circ$ and $\diamond\perp = \circ$ if $\diamond = \bullet$; and its productions are given by

$$
\begin{array}{rl}
& \{B^{\bullet} \to A^{\bullet} \mid \boxed{x|x : A^{\bullet}, x : B^{\circ}} \in \mathbf{LG}_T\} \\
\cup & \{A^{\circ} \to B^{\circ} \mid \boxed{x|x : B^{\circ}, x : A^{\bullet}} \in \mathbf{LG}_T\} \\
\cup & \{C^{\bullet} \to A^{\bullet}, B^{\bullet} \mid \boxed{R_{\otimes}xyz|y : A^{\bullet}, z : B^{\bullet}, x : C^{\circ}} \in \mathbf{LG}_T\} \\
\cup & \{B^{\circ} \to C^{\circ}, A^{\bullet} \mid \boxed{R_{\otimes}xyz|x : C^{\circ}, y : A^{\bullet}, z : B^{\bullet}} \in \mathbf{LG}_T\} \\
\cup & \{A^{\circ} \to B^{\bullet}, C^{\circ} \mid \boxed{R_{\otimes}xyz|z : B^{\bullet}, x : C^{\circ}, y : A^{\bullet}} \in \mathbf{LG}_T\} \\
\cup & \{C^{\circ} \to A^{\circ}, B^{\circ} \mid \boxed{R_{\oplus}xyz|y : A^{\circ}, z : B^{\circ}, x : C^{\bullet}} \in \mathbf{LG}_T\} \\
\cup & \{B^{\bullet} \to C^{\bullet}, A^{\circ} \mid \boxed{R_{\oplus}xyz|x : C^{\bullet}, y : A^{\circ}, z : B^{\circ}} \in \mathbf{LG}_T\} \\
\cup & \{A^{\bullet} \to B^{\circ}, C^{\bullet} \mid \boxed{R_{\oplus}xyz|z : B^{\circ}, x : C^{\bullet}, y : A^{\circ}} \in \mathbf{LG}_T\} \\
\cup & \{A^{\bullet} \to w \mid w \in \mathscr{A},\ A^{\bullet} \in L(w)\} \\
\cup & \{A^{\circ} \to w \mid w \in \mathscr{A},\ A^{\circ} \in L(w)\}
\end{array}
$$

We claim $\mathscr{G}_1$ and $\mathscr{G}_2$ recognize the same languages.

- Going from left to right, assume $\mathscr{G}_1$ recognizes $w_1, \ldots, w_n$. Then for some $A_1^{\diamond 1} \in L(w_1), \ldots, A_n^{\diamond n} \in L(w_n)$ and Θ, $S = \boxed{\Theta|x_1 : A_1^{\diamond 1}, \ldots, x_n : A_n^{\diamond n}, x : g^{\diamond}}$ closes. We claim $g^{\diamond\perp} \to^* A_1^{\diamond 1}, \ldots, A_n^{\diamond n}$, and hence $g^{\diamond\perp} \to^* w_1, \ldots, w_n$. This follows from an inductive argument on the tableau of S constructed by Lemma 42, proving that if $\boxed{\Theta|y_1 : B_1^{\diamond' 1}, \ldots, y_m : B_m^{\diamond' m}, y : B^{\diamond'}}$ closes, then $B^{\diamond'\perp} \to^* B_1^{\diamond' 1}, \ldots, B_m^{\diamond' m}$. The base cases follow from the construction of $\mathscr{G}_2$, while the sole inductive case depends on the transitive closure of $\to^*$.
- Conversely, suppose $g^{\diamond\perp} \to^* w_1, \ldots, w_n$. Then, by the construction of $\mathscr{G}_2$, $g^{\diamond\perp} \to^* A_1^{\diamond 1}, \ldots, A_n^{\diamond n}$ for some $A_1^{\diamond 1} \in L(w_1), \ldots, A_n^{\diamond n} \in L(w_n)$. Since all production rules involved draw from elements of $\mathbf{LG}_T$, a straightforward inductive argument constructs a closed tableau of $\boxed{\Theta|x_1 : A_1^{\diamond 1}, \ldots, x_n : A_n^{\diamond n}, x : g^{\diamond}}$ for some Θ using B-expansions, and we remove the latter one by one from bottom to top through repeated applications of Lemma 34.

In [11], a slightly different notion of **LG**-grammars is used. Stated as a special case of our grammars $\mathscr{G} = \langle \mathscr{A}, L, g^{\diamond} \rangle$, $\diamond$ is fixed at $\circ$ and the range of L is restricted to signed formulas $A^\bullet$. Moreover, the language $\mathscr{L}(\mathscr{G})$ recognized by $\mathscr{G}$ now reads as the set of lists $w_1, \ldots, w_n$ of words s.t. for some $A_1^\bullet \in L(w_1), \ldots, A_n^\bullet \in L(w_n)$ and tree Θ, $S = \boxed{\Theta \boxed{x_1 : A_1^\bullet, \ldots, x_n : A_n^\bullet, x : g^\circ}}$ closes, provided Θ lacks conditions of the form $R_\oplus xyz$. Note, though, that the latter kind of conditions may still appear further down in the tableau for S. In particular, $A_1, \ldots, A_n$ may freely contain connectives from the coresiduated family $\{\oplus, \oslash, \obslash\}$. Seeing as the above definitions constitute special cases of ours, context-freeness is preserved.

Acknowledgements. This work has benefited from discussions with Michael Moortgat, Jeroen Bransen and Vincent van Oostrom, as well as from comments from an anonymous referee. All remaining errors are my own.

References

1. Vito Michele Abrusci. Classical conservative extensions of Lambek calculus. *Studia Logica*, 71(3):277–314, 2002.
2. Philippe De Groote and François Lamarche. Classical Non Associative Lambek Calculus. *Studia Logica*, 71:355–388, 2002.
3. V.N. Grishin. On a generalization of the Ajdukiewicz-Lambek system. In A. I. Mikhailov, editor, *Studies in Nonclassical Logics and Formal Systems*, pages 315–334, Nauka, Moscow, 1983.
4. Gerhard Jäger. Residuation, structural rules and context freeness. *Journal of Logic, Language and Information*, 13(1):47–59, 2004.
5. Natasha Kurtonina. *Frames and Labels. A Modal Analysis of Categorial Inference.* PhD thesis, UiL OTS, and ILLC Amsterdam, 1995.
6. Natasha Kurtonina and Michael Moortgat. Relational semantics for the Lambek-Grishin calculus. In Christian Ebert, Gerhard Jäger, and Jens Michaelis, editors, *MOL 10/11. Selected papers from the 10th and 11th Mathematics of Language Meetings, Los Angeles 2007, Bielefeld 2009*, volume 6149 of *LNCS*, pages 210–222, Heidelberg, 2010. Springer.
7. François Lamarche. On the Algebra of Structural Contexts. *Mathematical Structures in Computer Science*, 2003.
8. Joachim Lambek. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–169, 1958.
9. Joachim Lambek. On the calculus of syntactic types. In Roman Jakobson, editor, *Structure of Language and its Mathematical Aspects, Proceedings of the Twelfth Symposium in Applied Mathematics*, 1961.
10. Matthijs Melissen. The generative capacity of the Lambek–Grishin calculus: A new lower bound. In P. de Groote, editor, *Proceedings of the 14th Formal Grammar Conference*, volume 5591 of *Lecture Notes in Computer Science.* Springer, 2009. to appear.
11. Michael Moortgat. Symmetric categorial grammar. *Journal of Philosophical Logic*, 38(6):681–710, 2009.
12. Richard Moot. Proof nets for display logic. *CoRR*, abs/0711.2444, 2007.
13. M. Pentus. Lambek calculus and formal grammars. In *Provability, Complexity, Grammars*, number 192 in American Mathematical Society Translations–Series 2, pages 57–86. American Mathematical Society, Providence, Rhode Island, 1999.
14. Raymond M. Smullyan. *First–Order Logic.* Springer-Verlag, 1968. Revised edition, Dover Press, NY, 1994.

Inquisitive Semantics and Legal Discourse

Martin Aher

University of Osnabrück

Introduction

Many authors in argumentation theory and linguistics have recognized the importance of studying the language of law, albeit few researchers have investigated the semantics and pragmatics of law within precise logical frameworks. One might mention deontic logics as a notable exception and an example of how the use of language in law is both an interesting area for reflection on language itself and the manner in which legal discourse is inextricably connected with ordinary language use.

Previous formal accounts of legal discourse have often been reductive, reinterpreting utterances as the force of their speech act. For example, Arno Lodder, in his PhD thesis[Lodder, 1998], attempts to model legal discourse with the use of only four conversational "moves." These are claims (assertions such as "He is guilty."), questions (understood as rejections of claims, such as "Is that true?"), withdrawals (changing one's mind) and acceptances of the claims of others. It is true that participants do make assertions and utter other speech acts, yet it might be interesting to attempt to model legal discourse closer to the textual account. For example, one could attempt to model all uses of questions uttered in legal discourse, not merely the ones that cast doubt on assertions. Furthermore, assertions and their acceptances can be captured in semantics and pragmatics using the stalnakerian notion of common ground. One such framework is Inquisitive Semantics[Groenendijk and Roelofsen, 2009] and this paper explores the feasibility of utilizing it for modeling legal discourse.

A yet untapped corpus of natural language examples for the analysis of legal discourse can be found in the public and accessible World Trade Organisation (hereafter referred to as the WTO) panel reports, an international trade law equivalent of legal opinion papers or court rulings. These panel reports are summaries of legal discourse conducted mainly via letters between the country that brought the complaint to the WTO (hereafter the complainant), the respondent and the panel itself, a group of judges called the appellate body panellists. For our purposes, we can consider the complainant, the respondent and the panel as three interlocutors in a dialogue, and this is also the format that the panel report mimics. Yet, a reported interpretation of written communication differs from spoken dialogue in many ways and this should be kept in mind during the analysis. The crucial added value of using panel reports to study the interpretation of disjunction is that each utterance is followed by a reaction by the intended addressee which provides the analyst with valuable subjective data about how to interpret utterances in context.

The Example

Disjunction is a core notion for inquisitive semantics which makes it appropriate to investigate an example that gravitates around disjunction for possible discrepencies.[1] A legal dispute is complex and shrouded in legal terminology but perhaps the following elucidation will provide the background information necessary for the interpretation of the examples.[2] The banana dispute discussed here stems from the fact that the complainant can produce bananas cheaper than the respondent and so the respondent's bananas are no longer being sold. The respondent reacted by placing a tax on all bananas but exempted his own. Yet, the respondent also provided a way to avoid the tax. If one buys the respondent's bananas and then sells them within the respondent's country for profit, one will be allocated licences that exempt them from the tax because they will be considered to be inside the "tariff quota." The complainant finds this unfair as selling under the "tariff quota" still reduces his profits because he needs to first buy more expensive bananas instead of directly selling his own cheaper ones. This *status quo* led to the following exchange between the complainant, respondent and the panel of judges.

Example 1. **Complainant:** "[the respondent is] inconsistent with Article III:4 of GATT because this licence allocation amounts to a requirement or incentive to purchase [the respondent's] bananas"

Example 2. **Respondent:** "[the respondent] does not force any trader to purchase any quantity of [the respondent's] bananas"

Before providing the reaction of the panel, one should note that the panel is here understood as a single interlocutor over several disputes and panels. The principle of jurisprudence establishes that panels must be aware of prior panel reports and would need to explicitly mark uttering a statement that is inconsistent with what has been decided in a prior panel report. This is very similar to how a person would need to mark changing their opinion. Yet, if a prior panel has uttered something relevant, it needs to be explicitly brought to the attention of the interlocutors of the current dispute. This is what they do when they remind the complainant and respondent of the following. (All subsequent quotes came with the voice of the panel but they merely restated a prior panel report.)

Example 3. **Panel:** "operators wishing to increase their future share of bananas benefiting from the tariff quota would be required to increase their current purchases" [of the respondent's bananas.]

[1] Merely the relevant sections are more than four pages long, so I shall constrain the examples to the bare minimum.

[2] This summary attempts to avoid taking the side of either interlocutor and will thus probably annoy both. Apologies.

Example 4. **Panel:** [quoting the relevant GATT article III.4] "The products of the territory of any Member imported into the territory of any other Member shall be accorded treatment no less favourable than that accorded to like products of national origin in respect of all laws, regulations and requirements affecting their internal sale, offering for sale, purchase, transportation, distribution or use."

Example 5. **Panel:** "this obligation [described in 4] applies to any requirement imposed by a contracting party, including requirements 'which an enterprise voluntarily accepts to obtain an advantage from the government"'

Example 6. **Panel:** "The Panel then proceeded to examine the [respondent's] licensing scheme in the light of the incentive provided"

Example 7. **Panel:** "In the view of the Panel, a requirement to purchase a domestic product in order to obtain the right to import a product at a lower rate of duty under a tariff quota is therefore a requirement affecting the purchase of a product within the meaning of Article III:4"

The panel took note of the utterances of the complainant and respondent and found that the complainant's case was justified. The manner in which they reasoned was not by utilizing the language of acceptances and rejections that one would expect if the utterances were to be reduced to their speech acts. In fact, they provide an account which accommodates the positions of both the complainant and respondent and builds on these. It has thus become expedient to now introduce the relevant elements of the framework of inquisitive semantics to determine the feasibility of modeling the preceding exchange.

Inquisitive Semantics and Pragmatics

The following is merely a brief overview of the notions from inquisitive semantics and pragmatics that are needed for the following analysis. At the core of inquisitive semantics lies an innovative account of disjunction in which uttering the following does not merely assert the fact, but also raises an issue.[Groenendijk and Roelofsen, 2009]

Example 8. It is raining or snowing.

Inquisitive semantics couples the notion of informativeness with that of inquisitiveness so that the effect of uttering (8) becomes twofold. Firstly, it eliminates possible worlds that are incompatible with the utterance, in this case the world where "it is not raining and it is not snowing." This can be called the informative content of disjunction. Yet, disjunction also provides two possibilities, modeled as sets of possible worlds, to account for the fact that the interlocutor is requested to choose between the alternative disjuncts. This choice provides the inquisitive content of the utterance. As disjunction in inquisitive semantics also raises an issue, questions can be modeled via disjunction in the semantics, rather than in a separate syntax. For this, Radical Inquisitive Semantics[Groenendijk and Roelofsen., 2010] defines the notion of counter-possibilities which captures the negative responses for a proposition. For example, "Is it raining?" could be modeled as $p \vee \div p$ to represent the possibility that it is raining and the counter-possibility that it is not. Inquisitive Semantics also provides for non-inquisitive uses of disjunction, for more on this see [Mascarenhas, 2009; van Gool and Roelofsen, 2009].

Inquisitive semantics also provides a stalnakerian pragmatic account. (8) provides us with an example that is generally uttered in the context of information exchange which is widely considered the standard context of language use. Within inquisitive semantics, the aim of interlocutors is to enhance the common ground which might enhance the information that an individual interlocutor possesses or bring awareness of the common ground to all interlocutors so as to facilitate coordinated action.

Each conversational participant has an information state that embodies what the participant takes to be the case. An information state is represented in the traditional way by a set of possible worlds or, in other words, ways in which the participant can imagine the world to be. If the set is empty, the information state is inconsistent. Once the participants exchange inquisitive and informative statements, they shall establish common ground.[Stalnaker, 2002, 1978]

When an interlocutor utters a statement, he or she proposes to update the common ground with the information or alternatives in the statement. An informative update provides information about how the world is and thus eliminates possible worlds from among the possibilities in the common ground. An inquisitive update would provide possibilities between which an interlocutor may choose. Groenendijk and Roelofsen propose that in pragmatic dialogue management, an utterance does not immediately update the common ground; instead, the hearer must either directly or indirectly accept or support the statement. [Groenendijk and Roelofsen, 2009] If an update with an utterance would be contradictory with what the interlocutor is aware of knowing, an explicit cancellation is required to maintain the common ground. [Groenendijk and Roelofsen, 2009, p. 12] Through accepted updates, the common ground ultimately grows in information and shrinks in terms of the number of possible worlds it embodies.

The coherence of discourse is governed by the principle of compliance, which judges the relatedness of utterances to one another. There are two ways in which a subsequent utterance can be compliant with an initiative. It either partially resolves an issue, which happens when it provides information, or the following utterance provides a sub-question that would be easier to answer.[3] For example,

[3] Easier to answer is taken to mean that the sub-question would provide a partial answer and that all information states that support the original question also support the sub-question.

the way in which to be compliant with a question, or a hybrid such as disjunction, is either to assert an answer or to pose a sub-question.

The purpose of (8) is to provide the interlocutor with a choice between disjuncts. Yet, the interlocutor may disagree with the proposed update, negating the disjunction as a whole. To do so, one is required to negate both disjuncts. In fact, uttering merely the negation of one of the disjuncts has quite the contrary meaning. It would allow one to establish, via inference, that the other disjunct is true. The rationale behind this is that (8) was uttered by someone who had good reason to establish that either or both of the disjuncts is the case. The negation of one disjunct does not negate the disjunction and thus provides, by default, an acceptance of the proposal to update the common ground. This establishes the grounds to use the original disjunction and the negation of one disjunct as two premises for establishing the remaining disjunct via elimination.

Discussion

This section sketches a possible way of modeling the way in which the panel resolved the dispute. As inquisitive semantics is still in a developmental phase with new additions published quarterly so this paper needs to be constrained to merely highlighting the promising aspects and likely points of contention of using inquisitive semantics to model legal discourse.

Please recall the example in which the complainant used the disjunctive (1) to accuse the respondent of maintaining a discriminatory requirement or an unfair incentive. For simplicity, let p stand for the proposition that "this licence allocation amounts to a requirement to purchase bananas" (hereafter referred to simply as "requirement") and q for "this licence allocation amounts to an incentive to purchase bananas" (hereafter referred to as "incentive"). In this case, the relevant utterance has the form $p \vee q$.[4] The entire utterance was intended to establish an implication relation from the disjunction to the proposition of "being inconsistent with Article III.4". This could be captured in the following manner: $p \vee q \rightarrow r$.

If the complainant is successful and the complaint is accepted into the common ground the respondent will incur some disadvantages. Assuming that the respondent will try to avoid these, we could predict that they shall negate the disjunction as a whole. But this is not what they do, probably because they are limited by the maxim of sincerity[Groenendijk and Roelofsen, 2009, p. 18] or, in Gricean terms, the maxim of Quality[Grice, 1989, p. 27-28], to only uttering statements supported by their own belief state.

The most salient intepretation for the response in (2) is that it negates the disjunct "requirement" or, in other words, it has the effect of uttering $\neg p$.[5] One would expect that the judges of the panel have their task made easy by the

[4] Due to the length of the paper, metalinguistic and other possible uses of disjunction are not taken under consideration in this paper.

[5] A less plausible interpretation would be that the respondent utters no response to the disjunction, marked by the use of the lexically different "force."

respondent. If one utters $p \vee q$ and another utters $\neg p$ one could immediately utter something of the kind: “then q is the case” and claim that there is no real dispute as the respondent seems to have accepted the disjunction and negated only one disjunct. As we discussed earlier, when a disjunction is accepted and one disjunct is eliminated, the other disjunct can be established in the common ground. Yet, inquisitive semantics provides an analysis that takes into account that the disjunctive was part of the antecedant of an implication which provides a different interpretation.

The following analysis rests on the assumption that the entire process is directed at resolving the issue $?r$ or whether the respondent's licence allocation system is inconsistent with the relevant article of legal text. Thus, and this is a possible point of contention, any complaint is going to be interpreted by the panel as an issue formed of the proposition and its counter-possibility. This can be modeled using inquisitive semantics as $(p \vee q) \rightarrow ?r$ or, equivalently, as $(p \rightarrow ?r) \wedge (q \rightarrow ?r)$. Assuming that the respondent understands this, the utterance of $\neg p$ can be seen as rejecting the question behind the $p \rightarrow ?r$ as the supposed antecedant could not be the case. This reduces the complex issue to $q \rightarrow ?r$. The text of article III.4 does not lexically specify “incentives” which provides for hope that no implication from incentives to inconsistency will be found.

It is notable that the the complainant and respondent play no further part in the exchange but the panel makes several utterances. If the issue were, for example, “Is it raining?” then one would expect a straightforward answer that picked whichever alternative is supported by the belief state of the utterer. Yet, a panel seems to be unable to provide a verdict based on merely a belief, it must be derived from accepted facts. Thus, the panel does not utter r or $\neg r$ but instead attempts to reach either alternative via a process of eliminating possible worlds through a sequence of utterances. The entire process is not straightforwardly predicted by the notion of compliance as providing the process of reasoning is overinformative, yet the reasoning itself utilizes the familiar *modus ponens.*

The panel's first utterance (3) establishes that an “incentive” is the case[6] because one must increase the amount of bananas purchased from the respondent or face a tax. At this point, the panel does not conclude the case solved even though they have established q which is in itself sufficient to find the complainant's $p \vee q$ to be the case.[7] Instead, onse should recall that the disjunction was uttered with the intent of resolving $?r$ and for this it becomes relevant to investigate $q \rightarrow ?r$.

The panel then utters (4) to note that article III.4 makes no explicit mention of “incentives,” but it does explicitly list “requirements.” This has the effect of uttering $p \rightarrow r$ but it also suggests that there is no textual reason to provide

[6] The wording is complex but the panel reinforces the reading of 3 by soon afterwards uttering 6.

[7] This is reinfoced by a separate panel report: “panels are not required to make a finding on every claim raised, but rather panels may practise "judicial economy" and make findings on only those claims necessary to resolve a dispute.”[Summary of DS33]

a link between q and r. Without introducing new variables, the panel has the remaining issue of whether there exists a relation from "incentives" to "requirements." This issue is reinforced by the use of "to require" in (3). The panel finds in (5) that providing an "incentive" to accept some "requirement" can still be referred to as establishing a "requirement." The idea behind this reasoning is that to qualify for the advantages which provide an incentive, one must accept a requirement. Or in this case, to avoid the tax imposed by the respondent, the complainant must meet the requirement of purchasing some bananas from the respondent. Once requirements can be understood as something that one voluntarily accepted, the difference between "requirement" and "incentive" dissolves. Whenever an incentive is the case, a requirement must also be the case. This is what the panel also makes explicit in an introductory paragraph to the conclusion (6).

In brief, one can sketch the panel's actions as the following:

$$\begin{array}{rcl} Complainant & p \vee q & (1) \\ Respondent & \neg p & (2) \\ Panel & q & (3) \\ Panel & p \rightarrow r & (4) \\ Panel & q \rightarrow p & (5) \\ Panel & p, r & (7) \end{array}$$

Conclusions

This analysis cannot be considered exhaustive and its contribution relies mainly in highlighting the advantages of using semantics and pragmatics in the programme of modeling legal discourse. The sketched model captured most parts of the discourse. It also brought into focus two aspects of legal pragmatics that require careful consideration and further analysis.

Firstly, the initial statements of the complainant and respondent do not seem to have the same discoure effect as they would in other contexts. The crucial point here is that the panel was directed at the resolution of whether there existed an inconsistency with article III.4. This could have been part of the effect of the respondent's reaction yet it could also be a feature of the context that the panel exists to determine whether such inconsistencies exist. Furthermore, if we were to say that the respondent produced the utterance $\neg p$ then one must account for the fact that the respondent could not reject the final conclusion of p in (7) from entering the common ground.

Secondly, the panel's response to the dispute was, under the notion of compliance, overinformative as it did not merely answer the issue at hand but produced the entire reasoning process which consisted of utterances that were not directly compliant (although entirely relevant) with the aim of resolving the issue $?r$.

As to the feasibility of modeling legal discourse with inquisitive semantics and pragmatics, this limited analysis found no major obstacles.

Bibliography

Herbert Paul Grice. Logic and conversation. In *Studies in the Way of Words*. Harvard University Press, 1989.

Jeroen Groenendijk and Floris Roelofsen. Inquisitive semantics and pragmatics. In *Standford Workshop on Language, Communication and Rational Agency*, 2009.

Jeroen Groenendijk and Floris Roelofsen. Radical inquisitive semantics. Preliminary version, presented at the Colloquium of the Institute for Cognitive Science, University of Osnabrueck, January 2010.

Arno Lodder. *Dialaw on Legal Justification and Dialog Games*. PhD thesis, Universiteit Maastricht, 1998.

Salvador Mascarenhas. Inquistive semantics and logic (unpublished). Master's thesis, University of Amsterdam, 2009.

Robert Stalnaker. Assertion. *Syntax and Semantics*, 9(P Cole):315–332, 1978.

Robert Stalnaker. Common ground. *Linguistics and Philosophy*, 25:701–721, 2002.

Sam van Gool and Floris Roelofsen. Disjunctive questions, intonation and highlighting. In *Proceedings of the Amsterdam Colloquium*, 2009.

Posters

Fictional Names

Fiora Salis

LOGOS-University of Barcelona

When you and I recollect the content of Dostoyevsky's *The Brothers Karamazov* we could say:

1. Aleksey Karamazov is a likeable young man.

I assume that reports about the content of a fictional story are best understood as implicitly prefixed by an intensional operator *F*, which makes the proposition expressed by (1) true if it is part of the story content. Hence, (1) is of the more complex form 'It is part of the fiction that S'. Fictional truths are those that have been explicitly stipulated by the author in story-telling or those that are implicitly derivable from the explicit truths of the story[1]. I also assume that fictional names are referential expressions, even though they do not refer to anything, because there exist no fictional entities. It is a widespread view among referentialists that an utterance of a sentence containing fictional names that do not refer does not express any proposition. An original referentialist view has been advanced by so-called gappy proposition theorists such as David Braun (2005). I will show the drawbacks of Braun's proposal and defend a new version of gappy propositions as structured intensions with lexical items as relevant aspects of their truth-conditional content.

Sentences have syntactic structures that are given by the output of a transformational grammar and are usually called Phrase Structure Markers (henceforth PSMs). PSMs that are the objects of semantic interpretation are also called *logical forms.* There are two ways in which philosophers can think of PSMs: i) as the output of a transformational grammar; ii) as the output of a transformational grammar with lexical items inserted in it. (i) is certainly an adequate way of thinking of logical forms in general. But, as I will argue in what follows, (ii) guarantees a more adequate account of the

[1]Important aspects of the notion of truth in fiction such as the mechanisms of derivation of implicit fictional truths would exceed the aim of this short paper and will be therefore left out of the discussion.

structure of an ordered set together with lexical items:

1.a $[_S[_{NP}$ Aleksey Karamazov$]$ $[_{VP}$ is a likeable young man$]]$

An interpreted PSM is a corresponding ordered set with expressions replaced by intensions[2]. Names constitute an exception because, coherently with referentialism, the semantic interpretation of a proper name assigns it a fixed individual entity. Fictional names, however, have no referents. So, an interpretation of (1.a) is the following gappy proposition where $\boldsymbol{g}$ (the gap) is a dedicated element of the universe and $\boldsymbol{M}$ is the intension of the predicate expression:

1.b $[_S[_{NP}\ \boldsymbol{g}_{Aleksey\ Karamazov}]\ [_{VP}\ \boldsymbol{M}_{is\ a\ likeable\ young\ man}]]$

On this account structure, lexical items and semantic values are all different aspects of the truth-conditions of (1). (1.b) is a structured proposition together with its linguistic mode of presentation.

Now, let me motivate my choice of endorsing option (ii) with the following brief arguments. David Braun (2005) put forward a view under the name of gappy-proposition theory that assumed option (i), namely PSMs without lexical items. On this view we have the following problems. First, (1) and (2) express exactly the same structured proposition, yet they seem to say different things:

1. Aleksey Karamazov is a likeable young man.

1.b* $[_S[_{NP}\ \boldsymbol{g}]\ [_{VP}\ \boldsymbol{M}]]$

2. Smerdyakov is a likeable man.

2.b $[_S[_{NP}\ \boldsymbol{g}]\ [_{VP}\ \boldsymbol{M}]]$

Apparently, (1) says of Aleksey that he is a likeable man, while (2) says of somebody else, Smerdyakov, that he is a likeable man. Of course, this is only true in the fiction. There are no fictional individuals. Second, as a consequence of this, (1) and (2) would amount to the same fictional truth. And yet, we would like to say that (1) is fictionally true, while (2) is not part of the story content and hence it is false. Third, readers of *The Brothers Karamazov* would certainly assent to (1) as a correct report of the story content, while they would reject (2) as a misreport. Smerdyakov is not pleasant at all, in the story he is sullen and morose. So, the gappy proposition theory built on option (i) cannot explain the rationality of a speaker assenting to and denying one and the same

[2]On similar views of structured intensions see for instance Lewis (1972) and Richard (1990) among others.

a subject can have towards the same propositional object. In the present case, Braun would have to say that when assenting to (1) a speaker imagines the gappy proposition (1.b*) in an *Aleksey-ish* way, whereas when denying (2.b) the same speaker imagines the same gappy proposition, but in a *Smerdyakov-ish* way. But even so, as long as speakers can assent to and deny one and the same content, their rationality cannot be preserved. Therefore, Braun's solution is no solution at all.

A gappy proposition theory like the one I am defending that is built on option (ii) does not incur in the same problems (in fact, it rather solves them by adding lexical items as linguistic modes of presentation of different fictional individuals). But it may face at least two further difficulties. The first has to do with the notion of same-saying. There is a clear sense in which we would produce exactly the same correct report when uttering (1) and (3):

3. Alyosha is a likeable young man.

In Russian, 'Alyosha' is one of the diminutives (together with 'Alioshka' and 'Lióshechka') of 'Aleksey'. So it seems that there is a sense in which (1) and (3) say exactly the same thing (as would 'Alioshka is a likeable man' and 'Lióshechka is a likeable man'). And yet if we include the expressions 'Aleksey' and 'Alyosha' in the proposition as two different names of the same individual (although only fictionally so), there may be no same-saying in this case. The problem becomes more cogent when it comes to the related topic of translation. The fictional report in (1) expressed in English can be translated into Italian as:

4. *Aleksej Karamazov è un giovane piacevole.*

(1) and (4) must share a common content to be translated in the different languages, but their linguistic modes of presentation are clearly different. To maintain option (ii) and give an appropriate answer to these puzzling aspects, Id like to introduce an original way of individuating names.

There are two alternative metaphysical conceptions of words that Kaplan (1990) has called the *orthographic* and the *common currency* conception. The common currency conception presupposes a principle of continuity in accordance with which a word retains its identity through processes of change (corruption and translations) along the history of its uses. I propose to individuate a name through the history of its uses (or, what using practice is individuated by the first act of introduction of the name into public language. In particular, a fictional name is introduced by the author of fiction with no intention to refer, but only to pretend to refer to something. By reading the story, producing reports and critical judgments, the name is passed from the author to his readers in communication chains where each utterance is causally linked to the original source. I hypothesize that the individuation of the lexical items of logical forms takes place in what John Perry (2001) theorized as presemantic uses of context where context provides information for identifying an utterance, i.e. which words, in which language, with which syntactic structure and even with which meanings they are used. We can distinguish 'Aleksey Karamazov' from 'Smerdyakov' not because they are orthographically different names, but because their uses are linked to different name-using practices originated by Dostoyevsky's introductions in his storytelling. And we can identify 'Aleksey Karamazov', 'Alyosha Karamazov', 'Aleksej Karamazov' etc. as one and the same name because they are all causally linked to the same first act of stipulative introduction and they are all constrained (and sustained) by the same name-using practice.

References

Braun, D. (2005) Empty Names, Fictional Names, Mythical Names. *Nos* 39: 596-631.

Lewis, D.K. (1972) General Semantics. In D. Davidson and G. Harman (eds), *Semantics of Natural Language.* Dordrecht: Reidel.

Kaplan, D. (1990). Words. *The Aristotelian Society*, Supplementary Volume LXIV

Perry, J. (2001). *Reference and Reflexivity.* Stanford: CSLI Publications.

Richard, M. 1990. *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them.* Cambridge University Press.

Sainsbury, M. 2005. *Reference without Referents.* Oxford University Press.

[3]The term was first introduced by Sainsbury (2005), but the notion that I sketch here is rather different from his because Sainsbury does not use the common currency conception of names and does not use name-using practices to individuate names. On the contrary, on Sainsbury's proposal nothing prevents the possibility of generating a new name-using practice when the corrupted version of a name is used, or when a name is translated into a different language.

Implicit Arguments in Minimalist Grammars

Walter Pedersen

McGill University

1 Introduction

This paper is concerned with a particular widespread phenomenon in natural language, the presence of implicit arguments. Very roughly, an implicit argument can be said to be present in a sentence when a word expressing an *n*-ary relation appears with fewer than *n* phonologically-overt syntactic arguments. While this definition embraces a diverse array of phenomena such as null subjects and subjects of infinitives in control constructions, we shall here limit ourselves to considering implicit complements in simple transitive sentences. For example, the verb *eat* expresses a relation between an eater and a thing eaten, yet the latter may be left unexpressed, as in the sentence *Godzilla ate*. Implicit complements pose a challenge for any formal grammatical theory. Their distribution is for a large part idiosyncratic and lexically-based. Furthermore, we can distinguish a number of distinct types of implicit complements based on their interpretation and distribution.

The goal of this paper is to show how implicit complements can be incorporated into a minimalist-style grammar, which is a type of grammar in which lexical features play a primary role in building syntactic structure. The general attributes of minimalist grammars were first proposed in Chomsky [2]; some of these ideas were developed formally in Stabler [6]. It will be shown how modifications to the standard grammatical rules and features of minimalist grammars (such as those found in [6]) can allow for the introduction of implicit complements. §2 will provide a brief characterization of the data, and the proposal will follow in §3.

2 The Diversity and Distribution of Implicit Complements

It has been noted in the descriptive and theoretical literature (see for example [1], [3], [5]) that different verbs license different types of implicit complements; these are distinguished by the inferences that can be drawn from sentences containing them. In the case of a sentence like *Godzilla ate*, we find an implicit complement that is interpreted as an existential indefinite, with the sentence being synonymous (or nearly so) with the sentence *Godzilla ate something*. Thus, *eat* licenses an existential implicit complement.

Like *eat,* the verb *wash* expresses a binary relation (between a washer and thing washed), and is optionally transitive. However, the sentence *Godzilla washed* is not synonymous with *Godzilla washed something*, but rather with *Godzilla washed himself*. Thus, *wash* can be said to license a reflexive implicit complement, rather than an existential one. In addition to existential and reflexive implicit complements, we also find verbs that license reciprocal implicit complements (such as *kiss* in *The boy and girl kissed*), and verbs that license contextual definite implicit complements (such as *arrive* in *The boss arrived*, understood as e.g. *The boss arrived here/there*).

Note that only certain verbs allow for implicit complements; minimal pairs of verbs can be found that demonstrate the highly lexical nature of the phenomenon. For example, *Godzilla ate* contrasts with the ungrammatical **Godzilla devoured*, and *Godzilla hid* contrasts with **Godzilla concealed*.

In the following section we shall, for expository purposes, confine our attention to optionally transitive verbs that allow either an existential or a reflexive implicit argument (in particular, to the verbs *eat* and *wash*).

3 Incorporating Implicit Complements into Minimalist Grammars

In this section, it will be shown how the standard features and rules of minimalist-style grammars, when suitably modified and enhanced, can serve to introduce existential and reflexive implicit complements into a formal grammar. To demonstrate this prospect, we will develop a minimalist-style grammar for a fragment of English. The grammar developed here is presented in the Bare Grammar format of Keenan & Stabler [4]. The minimalist feature structure used here is a simplified version of that found in Stabler [6], in that only selectional (and not movement) features are used. The current proposal may be modified to fit any formalism in which lexical items bear selectional features.

Following the Bare Grammar format, let a grammar G = $<Lex_G, Rule_G>$, where Lex_G is a set of expressions and $Rule_G$ is a set of structure building functions. L_G, the language generated by G, is the closure of Lex_G under $Rule_G$.

We shall call the particular grammar we are developing here 'IMP'; IMP = $<Lex_{IMP}, Rule_{IMP}>$. Following the minimalist tradition, expressions of IMP are 3-tuples [v; x; f*] consisting of a phonological string v, a category feature x and a (possibly empty) string of selectional features f*. The lexicon of IMP is given below. Lex_{IMP} =

[*devoured*; s; $n^l n^r$], [*ate*; s; $n^l(n^r)$], [*washed*; s; $n^l<n^r>$], [*John*; n; ∅], [*Godzilla*; n; ∅]

Note that we allow expressions of IMP to bear bracketed selectional features as well as the usual non-bracketed ones. Bracketed features mark lexical items as those which allow implicit complements, with the ( ) and < > brackets corresponding respectively to existential and reflexive implicit complements.

Let $Rule_{IMP}$ = {•, E, R}. $Rule_{IMP}$ contains the usual 'merge' function •, as well as functions E, R for dealing with implicit complements.

• is a function that takes a pair of expressions and returns a single 'merged' expression. Two expressions are in the domain of merge iff the rightmost selectional feature of the first expression matches the category feature of the second; the ^{l} and ^{r} on selectional features determine the order in which the phonological strings of the input expressions are concatenated in the output expression. The formal definition of • is as follows:

$\langle A, B\rangle \in \mathrm{Dom}(\bullet)$ iff
$A = [v;\ x;\ f^*y]$ & $B = [w;\ z;\ \varnothing]$ & $y \in \{z^r, z^l, (z^r), (z^l), <z^r>, <z^l>\}$ for some strings v, w, features x, y z and possibly-empty string of features f^*

Where defined:
$[v;\ x;\ f^*y] \bullet [w;\ z;\ \varnothing] = [vw;\ x;\ f^*]$, where $y \in \{z^r, (z^r), <z^r>\}$
$[v;\ x;\ f^*y] \bullet [w;\ z;\ \varnothing] = [wv;\ x;\ f^*]$, where $y \in \{z^l, (z^l), <z^l>\}$

Let us define a sentence of L_{IMP} as an expression of the form [v; s; ∅]. L_{IMP} (i.e. the closure of Lex_{IMP} under $Rule_{IMP}$) contains the sentence [*Godzilla devoured John;* s; ∅]. This is shown by the following 'argument':

1. [*devoured*; s; n^ln^r] • [*John*; n; ∅] = [*devoured John*; s; n^l]
2. [*devoured John*; s; n^l] • [*Godzilla*; n; ∅] = [*Godzilla devoured John*; s; ∅]

• is a structure-building function similar to that found in any minimalist-style grammar. In addition to •, however, $Rule_{IMP}$ contains the unary functions E and R; it is by means of these functions that implicit argument phenomena are incorporated into the grammar. An expression is in the domain of E iff its rightmost UF is of the form (x); it is in the domain of R iff its rightmost UF is of the form $<x>$. Syntactically, both functions serve to delete the bracketed feature.

$A \in \mathrm{Dom}(\mathrm{E})$ iff $A = [v;\ x;\ y(z)]$ for some v, x, y, z
where defined $\mathrm{E}(\ [v;\ x;\ y(z)]\ ) = [v;\ x;\ y]$

$A \in \mathrm{Dom}(\mathrm{R})$ iff $A = [v;\ x;\ y<z>]$ for some v, x, y, z
where defined, $\mathrm{R}(\ [v;\ x;\ y<z>]\ ) = [v;\ x;\ y]$

Given these rules, L_{IMP} will contain the sentence [*Godzilla ate*; s; ∅], as the following argument shows.

1. E([*ate*; s; $n^l(n^r)$]) = [*ate*; s; n^l]
2. [*ate*; s; n^l] • [*Godzilla*; n; ∅] = [*Godzilla ate*; s; ∅]

It can be easily verified that L_{IMP} also contains the sentence [*Godzilla ate John*; s; ∅], as well as [*Godzilla washed John*; s; ∅] and [*Godzilla washed*; s; ∅]. However, L_{IMP} does not contain [*Godzilla devoured*; s; ∅], which is as desired, as the verb *devour* cannot be used intransitively in English.

Turning now to the semantics of IMP, let $M = \{\mathrm{D}, m\}$ be a model for IMP iff $\mathrm{D} \neq \varnothing$ and m is function whose domain is $\{A\colon A \in Lex_{IMP}\}$, and which satisfies the following conditions:

if $A = [v;\ \mathrm{n};\ \varnothing]$ then $m(A) \in \mathrm{D}$
if $A = [v;\ \mathrm{s};\ xy]$ then $m(A)$ is a function from D to functions from D to {0,1}
(where v is a string, and x, y are features)

In other words, nouns denote entities, and transitive verbs denote type <e<et>> functions.

For each model $M = \{D, m\}$, an interpretation $\| \, \|_m$ of L_{IMP} is a function that extends m and satisfies:

1. $\| A \bullet B \|_m = \|A\|_m(\|B\|_m)$
2. $\| E(A) \|_m = [\lambda x.\ \exists z : \|A\|_m (z)(x) = 1]$
3. $\| R(A) \|_m = [\lambda x. \|A\|_m(x)(x) = 1]$

The latter two conditions ensure that sentences such as *Godzilla ate* and *Godzilla washed* receive the appropriate implicit complement interpretations. Using SMALLCAPS to stand for the interpretation a lexical item receives in a model, it follows that for all models M of IMP:

$\|[\textit{Godzilla ate John}; s; \varnothing]\|_m = 1$ iff $\text{EAT}_m(\text{JOHN})(\text{GODZILLA}_m) = 1$
$\|[\textit{Godzilla washed John}; s; \varnothing]\|_m = 1$ iff $\text{WASH}_m(\text{JOHN}_m)(\text{GODZILLA}_m) = 1$

$\| E([\textit{ate}; s; n^l(n^r)]) \|_m = [\lambda x.\ \exists z: \text{EAT}_m(z)(x) = 1]$
$\| [\textit{Godzilla ate}; s; \varnothing]\|_m = 1$ iff $\exists z: \text{EAT}_m(z)(\text{GODZILLA}_m) = 1$

$\| R([\textit{washed}; s; n^r\langle n^l\rangle]) \| = [\lambda x.\ \text{WASH}_m(x)(x) = 1]$
$\| [\textit{Godzilla washed}; s; \varnothing \|_m = 1$ iff $\text{WASH}_m(\text{GODZILLA}_m)(\text{GODZILLA}_m) = 1$

Though quantified NPs and anaphors were not included in the fragment given here, it can be easily seen that *Godzilla ate* is predicted to be synonymous with *Godzilla ate something*, and *Godzilla washed* with *Godzilla washed himself.*

While presented with respect to a very small fragment of English, the proposal developed here can be straightforwardly extended to include ditransitive verbs that permit implicit complements, as well as to nouns, adjectives and adverbs. It can also be extended to include the other two types of implicit complements by adding additional bracket types and unary rules with corresponding interpretations.

References

1. Allerton, D.J.: Valency and the English Verb. Academic Press, London (1982)
2. Chomsky, N. The Minimalist Program. MIT Press, Cambridge, MA (1995)
3. Herbst, T., Heath, D., Roe, I., Gotz, D.: A Valency Dictionary of English. Mouton de Gruyter, Berlin. (2004)
4. Keenan, E., Stabler, E.: Bare Grammar: Lectures on Linguistic Invariants. CSLI Publications, Stanford CA (2003).
5. Levin, B.: English Verb Classes and Alternations. University of Chicago Press, Chicago, IL. (1993)
6. Stabler, E.: Derivational Minimalism. In: Retoré, C. (ed.) Logical Aspects of Computational Linguistics pp.68-95. Springer-Verlag, Berlin (1997).

Linguistic Exchange and Consensus Bargaining
A Game-Theoretical Account of Some Conversational Maxims

Mathias Winther Madsen

University of Copenhagen / University of Amsterdam
mathias.winther@gmail.com

Abstract. In this paper, I interpret conversations as bargaining situations in which two players attempt to shape a consensus according to their self-interest. A simple game is used as a model, and some properties of its Nash equilibria are briefly discussed. A number of semantic and pragmatic phenomena such as consistency, relevance, and orderliness turn out to be explainable in terms of strategic choices rather than absolute constraints imposed on the players.

1 Introduction

The label "game-theoretical semantics" is often applied to logics in which the relation between sentences and truth values is given in game-theoretic terms. Hintikka's semantic games [5], e.g., conceptualize proofs in terms of a competition between a verifier ("Myself") and a falsifier ("Nature").

This is interesting from many standpoints, but it also leaves the fundamental assumptions of classical logic essentially unchanged: The acceptability of a sentence is still an objective feature based on universal rules of syntax and semantic atoms with indsiputable truth values; no answer is given to why the atomic sentences are excempted from dynamics of the game, or how they come about.

An alternative would be to use game theory not as a link between atoms and compounds, but as model of the negotiation process producing acceptability and analyzability. One suggestion in that direction is Robert Brandom's inferentialist theory [1], which takes the ethical notions of commitment and entitlement as basic while eplaining truth and material implication as a product of those.

The following can be seen as an attempt to spell out such a theory in more formal terms. I introduce a game in which two players attempt to maximize utility by shaping a common consensus according to their own interest. A number of semantic and pragmatic phenomena such as consistency, relevance, and orderliness emerge as rational strategic choices rather than rules or maxims. The idea behind the model is thus akin to the functionalist approach taken by e.g. [2].

2 The Consensus Game

In the following, I describe a two-player game G. G is an extensive game with perfect information, analogous to the games discussed in [4, Part II].

In a game of G, the two players take turns uttering sentences from a highly restricted language $\mathcal{L}$. This generates a consensus which may be more or less desirable to the each player. In formal terms, the components of G are:

1. Two **players**, denoted 1 and 2.
2. A **language** $\mathcal{L}$, defined recursively from the two syntactic atoms a and b:
   (a) $\varphi := a \mid \neg a \mid b \mid \neg b$
   (b) $\varphi := (\varphi \to \varphi)$
3. An "termination move," denoted by a square, $\Box$.
4. A set H of **histories**, which includes the following elements:
   (a) All elements of $\mathcal{L}^K$ for all $K \in \mathbb{N}_0$ (finite nonterminal histories).
   (b) All elements of $\mathcal{L}^K \times \{\Box\}$ for all $K \in \mathbb{N}_0$ (finite terminal histories).
   (c) All elements of $\mathcal{L}^{\mathbb{N}}$ (infinite and hence terminal histories).
5. The **player function** P, which states which player has the turn after a nonterminal history $h \in \mathcal{L}^K$. $P(h) = 1$ when K is even and $P(h) = 2$ otherwise.
6. Two **preference relations** $\gtrapprox_1$ and $\gtrapprox_2$, described below.

Note that $\mathcal{L}$ allows for indefinite nesting of implications, but excludes e.g. double negation. A game of G may thus give rise to e.g. the following terminal history, in which player 1 makes 3 moves and player 2 makes 2:

$$(a,\ (a \to \neg b),\ (a \to \neg b),\ a,\ ((b \to a) \to b),\ \Box) \tag{1}$$

In the following, I will omit outer parentheses around implications and refer to the moves player i makes in the course of a history h as $M_i(h)$. Thus, $M_i((a^k)_{k=1}^K) = \{a^m \in h \mid m \in \mathbb{N} \text{ and } P((a^k)_{k=1}^{m-1}) = i\}$.

I shall refer to the sentences a, $\neg a$, b, and $\neg b$ as **bits** and call sets of bits as **worlds**. I will refer to the set of worlds as W and use the notation $a\overline{b}$ as a shorthand for $\{a, \neg b\}$, etc.

The **consensus world** $C(h)$ **produced** by a history h is the set of bits to which both players have committed themselves, by which I mean the set of bits in the closure of $M_1(h) \cap M_2(h)$ with respect to modus ponens inference. For instance, in the history (1), $M_1(h) \cap M_2(h) = \{a, a \to \neg b\}$, and the corresponding consensus world is $a\overline{b}$. Note that some histories may produce "inconsistent" consensus worlds such as $a\overline{a}$.

All terminal histories h are thus associated with some consensus world $C(h)$. They further have a length $|h|$, which may be infinity. In the following, I will assume that the set $W \times \mathbb{N}$ is equipped with two complete, transitive, relfexive, binary relations $\succsim_1$ and $\succsim_2$, that may be used to define the players' preference relations $\gtrapprox_1$ and $\gtrapprox_2$ as follows:

$$h_1 \gtrapprox h_2 \iff (C(h_1), |h_1|) \succsim (C(h_2), |h_2|) \quad \text{for all } h_1, h_2 \in H.$$

In the following, I will sometimes use $x \succsim_i y$ as a shorthand for $(x, 1) \succsim_i (y, 1)$. I assume these relations satisfy the following conditions for all worlds $x, y, z \in W$, all lengths $t \in N \cup \infty$, all singleton worlds $p \in W$, and $i = 1, 2$ (cf. [4, 119]):

1. Time is valuable:
   (a) $(x,t) \succsim_i (x,t+1)$ with strict inequality if $(x,1) \not\sim_i (x,\emptyset)$.
2. Preferences are stationary:
   (a) $(x,t) \succsim_i (y,t+1)$ iff $(x,1) \succsim_i (y,2)$
   (b) $(x,t) \succsim_i (y,t)$ iff $(x,1) \succsim_i (y,1)$
3. Preferences are preserved at least 4 periods:
   (a) If $(x,t) \succsim_i (y,t)$, then $(x,t+4) \succsim_i (y,t)$.
4. Unions are order-preserving:
   (a) If $x \succsim_i y$, then $x \cup z \succsim_i y \cup z$; holds with strict inequalities, too.
   (b) If $x \cup y \succsim_i \emptyset$, then $x \succsim_i \emptyset$ or $y \succsim_i \emptyset$; holds with strict inequalities, too.
5. Negation cancels utility:
   (a) $p \cup \overline{p} \sim_i \emptyset$.

3 Equilibrium strategies

In G, a strategy for player i is a function $s_i : \{h \in H \mid P(h) = i\} \to L \cup \{\Box\}$. A Nash equilibrium of the game is a strategy pair (s_1, s_2) for which no change in s_i can lead to a strictly better outcome for player i, given that the other player's strategy is left unchanged. Extensive games typically have a very large set of equilibria, and this is also the case for G. The number and character of G's Nash equilibria depend on the preferences of the two players.

Let p and q refer to the worlds a, $\overline{a}$, b, and $\overline{b}$, so that $p \succeq_1 q \succeq_1 \emptyset$ and $p \neq q$. The different versions of G may then be characterized by player 2's preferences with respect to p and q; and in each case, the equilibria may be divided into classes according to the worlds they produce, as in Table 1.

No.	Case	Preference	PE	NE	SPE
I	Complete Agreement	$p, q \succeq_2 \emptyset$	pq	pq, p, q, $\emptyset$	pq, p, q, $\emptyset$
II	Minor Disagreement	$p \succeq_2 \overline{q} \succeq_2 \emptyset$	pq, $p\overline{q}$, p	pq, $p\overline{q}$, p, $\emptyset$	p, $\emptyset$
III	Major Disagreement	$\overline{q} \succeq_2 p \succeq_2 \emptyset$	pq, $p\overline{q}$, p	p, $\emptyset$	p, $\emptyset$
IV	Complete Disagreement	$\overline{p}, \overline{q} \succeq_2 \emptyset$	any world	$\emptyset$	$\emptyset$

Table 1. Four possible preference distrubutions for player 2 relative to player 1 and the consensus worlds produced by the strategy pairs that are Pareto efficient (PE), equilibria (NE), and subgame perfect equilibria (SPE).

Note that a strategy pair (s_1, s_2) cannot be an equilibrium if it produces an outcome strictly worse than $\emptyset$ for one of the players: The player in question always has the option of obtaining the outcome $\emptyset$ by switching to the constant-valued strategy $h \mapsto \Box$.

Most of the equilibria classes in G consist in strategy pairs that map out a shortest path to the relevant consensus world, and a set of actions assigned arbitrarily to histories outside that path. In Case I, e.g., there are 12 possible paths: two repetitions (in any order) of p and q, p and $p \to q$, or q and $q \to p$, and then a $\Box$; for all other histories h, $s_i(h)$ may be any move in $\mathcal{L} \cup \{\Box\}$.

The exception to this pattern is Case II, in which the classes pq and $p\overline{q}$ are characterized by a locally irrational but globally rational choices: Some moderately attractive subgames may be played in a suboptimal way to make them highly unattractive for the adversary. For instance, player 2 may choose to sabotage all paths not leading to the consensus world $p\overline{q}$ by playing $\Box$ in response to any action that deviates from those paths. When such a vindictive strategy is combined with a lenient response, it becomes an equilibrium pair.

Further, unlike Cases I, III, and IV, order matters in Case II: The equilibrium pairs in the $p\overline{q}$ case may e.g. produce the history $(p,\, p \to \overline{q},\, p \to \overline{q},\, p,\, \Box)$, but not $(p,\, p,\, p \to \overline{q},\, p \to \overline{q},\, \Box)$. In the latter case, player 1 would be strictly better off by playing $\Box$ after the history $(p,\, p)$, and the underlying strategies can thus not constitute an equilibrium pair.

4 Semantic and Pragmatic Properties of the Equilibria

The equilibrium strategy pairs of G exhibit a number of features which superficially make it seem as if they were governed by a set of rules along the lines of [3]. In particular, the Pareto efficient equilibrium strategies satisfy:

Quantity:	All paths have minimal length, i.e., $\lvert h\rvert = C(h)+1$; the players never repeat themselves.
Quality:	The players never utter a bit as well as its negation; $a\overline{a}$ consensus worlds never occur; no world becomes consensus if a player has a preference against it (although a bit might).
Manner:	Discounting the final $\Box$, at least half the history consists of bits; no utterance is more complex than an implication between two distinct bits (e.g., the types $p \to (q \to p)$ and $p \to \neg p$ are not used).
Politeness:	The players avoid all overt conflict; all raised issues are confirmed.

As we have seen, however, these properties are caused by the players' strategic attempts to promote their self-interest, not inherent linguistic constraints.

References

1. Robert B. Brandom. *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard University Press, 1998.
2. Penelope Brown and Stephen C. Levinson. *Politeness: Some universals in language usage*. Cambridge University Press, 1987.
3. H. Paul Grice. Logic and Conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics, Volume 3: Speech Acts*. Academic Press, New York.
4. Martin J. Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994.
5. Esa Saarinen, editor. *Game-Theoretical Semantics: Essays on Semantics*. D. Reidel Publishing Company, Dordrecht, Holland, 1979.

Frame-Evoking and Lexical Prefixes in Bulgarian

Svetlozara Leseva

Institute of Bulgarian Language, Bulgarian Academy of Sciences

Abstract. The paper outlines an approach to the representation of lexical prefixes with a view to what is known as their argument-structure changing properties. The prefixes are treated as abstract frame-evoking predicates that interact with and subordinate the frame of the constant. A case study illustrates the specifics of this interaction and its effects on the conceptual structure and the syntactic properties of the derived verbs as compared with those of the constant.

1 Introduction

It has been widely acknowledged that Slavic verbal prefixes are a heterogeneous group that can be divided in at least two major classes - lexical and superlexical (Svenonius 2004, among others). This distinction, based on the capacity of the lexical prefixes of affecting the inherent structure of the verb, is recognised across theoretical accounts, although phrased in different terms with respect to the level affected - logical structure (Van Valin & LaPolla 1997), argument structure (Svenonius 2004), lexical conceptual structure (Spencer & Zaretskaya 1998), etc. In the following paper I assume that lexical prefixes are frame-evoking elements in the sense adopted in the Framenet project (Rupenhofer et al. 2006). Following the idea of constants in Levin and Rappaport Hovav (1995 and subsequent work), I will refer to the predicates (as well as to other frame-evoking elements) that lexicalise the core idiosyncratic meaning subject to modification in the process of prefixation by that term or (where appropriate) also as verb roots.

The interaction between the constant's frame and the prefix' abstract frame-evoking properties results in the subordination of the former, as a result of which a change is effected in the lexical meaning of the constant, and possibly in its conceptual structure (frame). Since the semantics of the prefix is rather abstract, the preverb does not necessarily evoke a single frame, but rather any of a small set of frames sharing certain properties. It is the interaction of the properties of the constant and the prefix that determines which of these frames will be realised.

2 A case study - the locative meaning of the prefix *za-*

In what follows I will discuss a group of verbs formed with the locative prefix *za-*. It denotes affecting the surface(s) of an object, or the object itself (on or from (all) sides). The derived verbs fall into a small number of coherent semantic classes defined by the interaction between the idiosyncratic meaning of the constants and the meaning of the prefix, such as verbs of covering/filling (where the prefix evokes the frame Filling) or verbs of hiding (where the prefix evokes Hiding_objects).

Let us first consider the interaction between verb roots and prefixes. According to Filip (Filip 2008) Slavic prefixes 'add meaning components that contribute to specifying a criterion for ordering of events' in the denotation of verbs, i.e. they define a scale that orders the set of events 'based on the degree to which they possess a certain measurable property' (spatial, temporal, etc.) and an upper bound. From the definition of *za-* it follows that the nature of the scale specified by the prefix is in the spatial dimension and that the constants that combine with it have some spatial component to them.

Indeed, many of the constants that form *za-* verbs are verbs of covering/filling, i.e. in terms of the chosen framework, they evoke the frame Filling defined in Framenet as including 'words relating to filling containers and covering areas with some thing, things or substance...'. The frame identifies the following core elements:

Agent [Agt], a Sentient actor; **Cause [cau]**, an Event which brings about the filling of the Goal; **Goal [Goal]**, the area or container being filled, generally the NP Object;**Theme [Thm]** a Physical_object (or substance) which changes location.

Consider the prototypical verb for the frame Filling *pălnya* (fill). It denotes the process aimed at covering/filling or, on general reading, the result of covering/filling. The prefix *na-* yielding the pair *na-pălnya/na-pălvam*[1] contributes a meaning that corresponds to a scale specifying volume/quantity.

1a [Turistite]$_{AGENT}$ *pălnyat* [kufarite si]$_{GOAL}$ s [mestni stoki]$_{THEME}$, dokato gi na-pălnyat dogore/napolovina.

1b [Tourists]$_{AGENT}$ *fill* [their suitcases]$_{GOAL}$ with [locally-made goods]$_{THEME}$ until the latter are completely/half full.

As an anonymous reviewer points out judgements are not always quite clear as to actually reaching the limit, i.e. the upper bound. Indeed, examples such as 'na-pălnya dopolovina' (fill half of) point to that, but they specify a different scale (half), and the verb specify 'exhaustiveness' with respect to it. In the absence of a clear indicator suggesting otherwise, the most likely interpretation is that of reaching a maximal degree on the specified scale.

With respect to the pairs *na-pălnya/na-pălvam* and *za-pălnya/za-pălvam* the crucial difference between them lies in the fact that the latter prefix specifies a

[1] throughout the paper the aspectual pairs of derived verbs are given in the following order - perfective/secondary imperfective; the secondary imperfective is derived by means of an imperfective suffix from the derived perfective; despite the aspectual differences the derived pair shares common properties with respect to conceptual (argument) structure and the distinction is therefore considered irrelevant for the purposes of the present account

scale and an upper bound on a spatial locational dimension. As a result, with verb roots alternating between filling and placing interpretation, *za-* evokes the frame Filling and rules out Placing, hence 2f is ruled out. *Na-* verbs allow the Filling/Placing alternation (2a, 2c) in a similar manner as the constants do.

2a [Momcheto]$_{AGENT}$ *(na-)pălni* [săda]$_{GOAL}$ [s voda]$_{THEME}$.
2b [The boy]$_{AGENT}$ *filled* [the container]$_{GOAL}$ [with water]$_{THEME}$.
2c [Momcheto]$_{AGENT}$ *(na-)pălni* [voda]$_{THEME}$ [v metalnia săd]$_{GOAL}$.
2d [The boy]$_{AGENT}$ *poured* [water]$_{THEME}$ [in the container]$_{GOAL}$.
2e [Momcheto]$_{AGENT}$ *za-pălni* [metalnia săd]$_{GOAL}$ [s voda]$_{THEME}$.
2f *[Momcheto]$_{AGENT}$ *za-pălni* [voda]$_{THEME}$ [v metalnia săd]$_{GOAL}$.

To conclude, the prefix *za-* evokes frames related to affecting a surface in a manner that is compliant with the properties of the constant. In the case with filling constants the resulting verb evokes the same frame.
The semantics of the constants and the prefix in the above cases is to a great extent similar, which renders the result quite trivial. Let us now proceed to the interaction of the prefix *za-* with constants that evoke other frames. Consider for instance the verb pair *zastroya/zastroyavam* meaning 'cover an extent or area with buildings' derived from the verb *stroya (build)*. The constant evokes the frame Building that 'describes assembly or construction actions, where an **Agent [Agt]** joins **Components [Cmpnt]** together to form a **Created_entity [CrEnt]**, which is profiled, and hence the object of the verb.' Beside the three core elements, the frame features a number of peripheral and extra-thematic elements that describe different aspects of the situation. One of the peripheral FEs - Place, identifies the place where the building occurs, syntactically expressed either by a prepositional phrase or by an adverbial phrase with a locative meaning. (3) show the prefixed verb *postroya* where the prefix *po-* denotes the completion of the creation process and hence defines a scale related to the physical extent (Filip 2008).

3a [Rabotnitsite]$_{AGENT}$ *po-stroiha* [golyama kăshta]$_{CR\_ENT}$ ot [tuhli]$_{CMPNT}$ v [poleto/tam]$_{PLACE}$.
3b [The workers]$_{AGENT}$ *built* [a big house]$_{CR\_ENT}$ out of [bricks]$_{CMPNT}$ in [the field/there]$_{PLACE}$.

Let us go back to the interaction between the prefix *za-* and the constant *stroya*. In the case of *zastroya/zastroyavam* (4) an element, semantically identical to the non-core Place FE of the Building frame, is conceptualised as a core FE. The attachment of the prefix *za-* to the constant *stroya* results in the verb pair's evoking of the frame Filling:

4a [Rabotnitsite]$_{AGENT}$ *zastroiha* [poleto]$_{GOAL}$ s [kăshti]$_{THEME}$.
4b [The workers]$_{AGENT}$ *covered(by-building)* [the field]$_{GOAL}$ with [houses]$_{THM}$.

The juxtaposition of *stroya/postroya(vam)* and *zastroya(vam)* reveals the way in which a prefix may affect the conceptual structure and, in consequence, the related levels of semantic and syntactic description.

(i) The constant preserves its idiosyncratic structure of participants and relations while the participants are (re-)mapped onto the frame evoked by the prefixed verbs. In the case under consideration physical object(s) of a particular semantic type (Created_entity), that is/are created by means of joining components together, come to occupy (as an abstract act of placement - Theme) an extent of land (Goal) under the influence of an Agent.

(ii) The core frame element projected as an external argument in the frame Building belongs to the semantic type Agent, whereas the respective frame element of the Filling frame may either be Agent or Cause. The constant *stroya* imposes selectional restrictions upon the corresponding frame element instantiated by *zastroya(vam)*, resulting in the ruling out of the Cause examples.

(iii) The profiled core element Created_entity (Artefact) is conceptualised as Theme (Physical object). The entities that may be realised as the Theme element are subject to the semantic restrictions imposed by the constant *stroya* on the Building frame element Created_entity, therefore physical objects that are not artefacts are disallowed.

(iv) The core element Components of the Building frame is demoted to a non-core participant (Material (denoting components, ingredients, etc.)) of the frame Buildings (noun-evoking frame distinct from Building) evoked by the Theme element.

(v) Finally, the non-core element Place of the Building frame is conceptualised as the profiled core element Goal of the Filling frame. The selectional restrictions are those relevant for the Place element (areas and surfaces of land), so containers and other surfaces are not licensed.

To sum up the argument, the *za-* verb pair evokes a frame that is different from the constant's and one that is predictable from the conceptual properties of the prefix.

A stronger case is provided by verbs whose constants do not feature certain elements that are to be mapped onto the frame Filling. Consider the verbs *zalesya/zalesyavam (afforest), zatrevya/zatrevyavam (grass - 'cover with grass'), zamăglya/zamăglyavam (cloud, fog), zasnezha/zasnezhavam (snow up, cover with snow)*, derived respectively from the constants *forest, grass, fog, snow.* In abstract terms these verbs may be represented by the definition: 'cover some area with N', N being the constant, i.e. they also evoke the frame Filling. The constants share the property of having spatial extension, i.e. they occupy locations, spread over surfaces, etc. Obviously, they must evoke a frame that features some kind of locative relation FE, that is conceptualised as the Goal element of Filling.

Consider the verb *zalesya(vam).* As defined in Framenet, the constant forest evokes the Biological_area frame, that '[...] denote large ecological areas as well as smaller locations characterized by the type of life present - in other words, geography locations as defined by biota.' The frame has one core element: **Locale [Locl]** - a Location which 'identifies a stable bounded area, and is typically the designation of the nouns of Locale-derived frames':

5a V [dolinata]$_{LOCALE}$ raste [dăbova]$_{CNST}$ gora.
5b There is an [oak]$_{CNST}$ forest in the [valley]$_{LOCALE}$.

(i) The Locale frame element is conceptualised as the Goal argument of the derived pair. Being profiled, it is syntactically expressed as a direct object. Besides, the non-core element Constituent_parts [Cnst] of the Biological_area frame is conceptualised as the Theme element of the Filling frame (6):

6a [Rabotnitsite]$_{AGENT}$ zalesiha [dolinata]$_{GOAL}$ s [dăb]$_{THEME}$.
6b [The workers]$_{AGENT}$ afforested [the valley]$_{GOAL}$ with [oak]$_{THEME}$.

(ii) The Biological_area frame does not feature any FE that corresponds to the Agent/Cause of the Filling frame, hence this FE should be attributed to the frame-evoking properties of the prefix.

On the basis of the above observations it may be concluded that the prefix evokes frame(s) that subordinate the constant's frame where subordination includes the mapping of FEs of the constant's frame to FEs of the prefix' frame, and/or the suppression of FEs, and/or the 'introduction' of FEs not available in the constant's frame. Mapping may either be straightforward, re-mapping (non-profiled core to profiled core FE; profiled core FE to non-profiled core FE) or may include promotion (non-core to core movement) or demotion (core to non-core movement) of FEs.

I. Straightforward frame mapping (e.g. *pălnya - zapălnya/zapălvam)* where:
  I.1. the frames evoked by the root (henceforth - REF (root-evoked frame)) and the derived verb (henceforth - DVF (derived-verb frame)) are identical, i.e. the frame relations including the inventory and configurations of core and non-core FEs and their semantic types are the same.
  II. Non-trivial frame mapping where:
    II.1. REF non-core FEs are promoted to core position in the DVF; (e.g. the Place FE of the Building frame maps to the Goal FE of the Filling frame, the Constituent_parts FE of the Biological_area frame - to the Filling Theme)
    II.2. REF core FEs are demoted to DVF non-core position; (the Components FE of the Building frame - to the Material FE in the Buildings frame)
    II.3. REF core FEs are re-conceptualised and re-mapped onto core positions in DVF; (the Created_entity FE of the Building frame as Theme of the Filling frame, the Locale FE of Biological area frame as the Filling Goal FE)
    II.4. non-REF FEs are mapped to DVF positions (Agent FE of the Filling frame with *zalesya(vam)*).

Straightforward mapping involving lexical prefixes results in the derivation of verbs that lexicalise the natural culmination of a process/activity with the lexical meaning additionally elaborated by the lexical component of the prefix.

(i) AGENT [GOAL]*profiled* THEME ⇒ AGENT [GOAL]*profiled* THEME

Non-trivial mapping leads to verbs that may involve as a profiled FE a non-profiled core FE or a non-core FE. Non-REF FEs cannot be profiled. $FE_{LOCATIVE}$ in (ii) stands for the FE that is to be mapped on the Goal FE.

(ii) [...] $FE_{LOCATIVE}$ [...] ⇒ AGENT [GOAL]*profiled* THEME

Table 1 sums up the observation on a non-exhaustive list of constants' frames that derive Filling *za-* verbs, the relevant constants' FEs (CFE) that are to be mapped to the external argument, the Goal FE, the Theme FE, and examples.

Table 1. Constants and derived verbs

Constant Frame	External argument	CFE → Goal	CFE → Theme	Verbs
Filling	Ag/Cause → Ag/Cause	Goal	Theme	*zapălnya/zapălvam* *zatrup(v)am* (heap) *zagarna/zagrashtam* (wrap (up)) *zamazha/zamazvam* ((be)smear) *zastelya/zastilam* (cover) *zaseya/zasyavam* (plant)
Placing	Agent → Agent	Goal	Theme	*zametna/zamyatam* (throw over)
Building	Agent → Agent	Place	Cr_Ent	*zastroya/zastroyavam*
Motion	Carrier → Cause	Area/Goal	Theme	*zaveya/zavyavam* (blowing over)
Fluidic_motion	None → Agent/Cause	Area/Goal	Fluid	*zaleya/zalivam* (cover with liquid)
Biological_area	None → Agent/Cause	Locale	Cnst	*zalesya(vam)* (afforest) *zatrevya(vam)* (plant with grass)
Clothing_parts	None → Agent	Body_location	Subpart	*zakachulya/zakachulvam* (hood)
Clothing	None → Agent	Body_location	Garment	*zabradya/zabrazhdam* (scarf)
Precipitation	Precipitation/None → Cause/Agent	Place	Inc. theme	*zasnezha(vam)* (cover with snow) *zaslanya(vam)* (frost)
Other	None → Agent/Cause	Other	Other	*zamaglya(vam)* (befog) *zadimya(vam)* (fill with smoke) *zaprasha(vam)* (cover with dust) *zashumya(vam)* (cover with leaves) *zasmolya(vam)* (tar)

3 Frame correspondences induced by constants

The prefix *za-* is capable of evoking several frames that interact in a regular way with Filling - Abounding_with (together with an inchoative variant of the same frame not specified in Framenet), Adorning (and an inchoative variant of Adorning, also not specified).
The Abounding_with frame denotes the situation where: 'A Location is filled or covered with the Theme. The Location is realized as the External Argument, and the Theme [...] as PP complement [...].'. It designates a static relation which evokes verbs such as: teem, swarm, throng, as well as adjectives (participles) used predicatively (be) covered, (be) adorned, (be) coated, etc.:

7a [Nebeto]$_{LOCATION}$ *be zastlano/se zastla* [s oblatsi]$_{THEME}$.
7b [The sky]$_{LOCATION}$ *was covered/covered* with [clouds]$_{THEME}$.

The second verb in the examples has inchoative meaning that denotes the coming into the state. Leaving the aspectual properties aside, the inchoative variants share the conceptual properties of the stative frame, both specify two core FEs - a **Theme** (identical to the Filling Theme) and thus trivially mapped onto the relevant element, and a **Location** corresponding to the Filling Goal FE. No FE corresponds to the Agent/Cause.
The Adorning frame defines: 'a static (primarily spatial) relationship between a Location and a Theme. All of the verbs used statically in this frame can also occur in the frame Filling.' It is noted that the frame bears correspondence to Abounding_with. The difference lies in the point of view shift, i.e. while the Theme with Adorning is realised as the external argument and the Location is realised as an NP object, with Abounding_with it is vice versa. The inchoative variant is also given.

8a [Oblatsi]$_{THEME}$ *zastilaha/zastlaha* [nebeto]$_{LOCATION}$.
8b [Clouds]$_{THEME}$ *covered* the [sky]$_{LOCATION}$.

The relation between the three frames may be stated as follows: stative Abounding_with describes a state, the inchoative variant - a transition into a state, and Filling the causation of a transition into a state. For the inchoative variant the relation boils down to the causative-inchoative alternation. Unlike the Abounding_with where the focus is on the Location being (or coming to be) occupied by the Theme, Adorning conceptualises the relation of the Theme occupying or coming to occupy the Location.
The ability of a *za-* verb to evoke configurations of these frames lies in the interaction between the conceptual structure of the prefix and that of the constant, i.e. it depends on the idiosyncratic properties of the constant whether one or more of these frames will be evoked. For instance, agentive filling/covering verbs such as *zastroya(vam)* do not have an inchoative variant (Levin & Rappaport-Hovav 1995), as don't unaccusative motion verbs in the frames Motion and Fluidic_motion.
Other frames evoked by the prefix *za-* in the interaction with particular constants are Hiding_objects (and (or) its inchoative Eclipse, e.g. *zabulya/zabulvam*

(veil), zakriya/zakrivam (hide, conceal (behind/under)), etc., as well as Closure - *zahlupya/zahlupvam ('cover with a lid')*, *zapusha/zapushvam (stop up, plug up)*, *zatapya/zatapvam (seal, cork)*. As already noted above, it is the nature of the constants that influences the construal of the situation, and hence the particular frame evoked.

4 Conclusion

As was shown throughout the paper, the interaction between a given lexical prefix and certain constants may be captured in terms of a small number of conceptual frames and frame-to-frame relations that give insights into the semantic properties and the 'syntax' of lexical prefixes, on the one hand, and reveal systemic relations between certain frames and, respectively, classes of verbs.
The frame approach can be applied to polysemy owed to the interaction between alternating constants, e.g. *sadya (plant)* (Placing and Filling) and distinct senses of a prefix (e.g. *za-* (attachment) and spatial *za-*, respectively. Finally, it may also be extended to at least some superlexical prefixes in terms of defining (possibly subframal, i.e. lexical-unit to lexical-unit) relations between verbs, consider *za-sazhdam - (begin to plant)* formed with the inceptive prefix *za-*.

5 Acknowledgements

I would like to thank my supervisor prof. Svetla Koeva, Ivelina Stoyanova and two anonymous reviewers for valuable suggestions. This research has been supported by the European Social Fund through the Human Resource Development Operational Programme 2007-2013 under the project *Mathematical Logic and Computational Linguistics: Development and Permeation (Contract No. BG051PO001-3.3.04/27).*

References

Filip, H.: Events and Maximalization: the Case of Telicity and Perfectivity. In Rothstein, S. (ed.). Theoretical and Crosslinguistic Approaches to the Semantics of Aspect. John Benjamins Publishing Company. (2008)

Ivanova, K.: Nacini na glagolnoto deystvie v savremenniya balgarski ezik. Sofia (1974).

Levin, B., Rappaport Hovav, M.: Unaccusativity. Cambridge. MA: MIT Press. (1995)

Rupenhofer, Josef, Ellsworth, M. M.R.L. Petruck, Ch.R.: Johnson, Jan Scheffczyk.Framenet II. Extended Theory and Practice. Available at http://framenet.icsi.berkeley.edu/. (2006)

Spencer, A., Zaretskaya, M.: Verb prefixation in Russian as lexical subordination. In Linguistics 36, pp. 1-39. (1998)

Svenonius, P.: Slavic prefixes inside and outside VP. In Tromso Working Papers on Language and Linguistics: Nordlyd 32.2, Special issue on Slavic prefixes, edited by Peter Svenonius, pp. 205-253. University of Tromso, Tromso. Available at www.ub.uit.no/baser/nordlyd/. (2004)

Van Valin, R., LaPolla, R: Syntax: structure, meaning and function. Cambridge:Cambridge University Press. (1997)

Ontological Categories & Property Specification: The case with Turkish preschoolers

Hatice Bayındır
*PhD in progress at METU Cognitive Science Program & Instructor at Çankaya University, Ankara, Turkey

Abstract

An important issue in cognitive development has been how children develop an understanding of the entities around them and how they categorize these entities ontologically through a process of tracking and property specification. The current study was conducted to have a better understanding of Turkish children's ontological categories and their corresponding properties. Ontological categories included living things and nonliving things. The sample consisted of 25 preschoolers: 6 four-year-olds, 9 five-year-olds, and 10 six-year olds, all of which were selected randomly from a kindergarten, Arı Preschool, in Ankara, Turkey. There was also an adult control group of 8 people whose data provided the criteria based on which children's responses were scored. The real photos were directly presented to the children and their answers to the questions were recorded by circling yes or no in the answer section of the form and notes regarding children's extra explanations were written down in the extra spaces provided. The whole procedure took ± 10 minutes for each child. The current study indicated that Turkish preschoolers' cognition of ontological categories develop across the specified age range, 4-6 year-olds. There was a significant difference between 4- year-olds and 6-year-olds in their performance of category specification, whereas 5-year-olds' performance significantly differed from neither 4-year-olds nor 6-year-olds. These findings demonstrated children's capability of distinguishing entities around them with reference to some biological properties. For all age groups, subjects' performance differed across category types, with humans as the best specified category while animals and plants were the worst in the broader category of living things, and with man-made things as the better specified category over natural things in the broader category of nonliving things. They all did better in the nonliving things category in comparison with the living things, excluding the human category. These findings could be linked to habitual surroundings, cultural conditioning, informal learning opportunities, formal schooling, and limited experience and observation opportunities.

Key words: ontological categories, living vs. non living things, property specification, Turkish preschoolers

1. Introduction and literature overview

An important issue in cognitive development has been how children develop an understanding of the entities around them and how they categorize these entities ontologically through a process of tracking and property specification.

One extension of this discussion has centered around the question whether children can distinguish living from non-living things. Two opposing lines of research have emerged regarding children's understanding of living vs. nonliving things (Zhu & Fang, 2000). The studies in line with Piaget's position of children's inability to distinguish between living and nonliving things found that children didn't know what entities are exactly included in the category of living things and they didn't regard animals and plants as the members of the same category (Carey, 1985; Richards & Siegler, 1986; Stavy & Wax, 1989; cited in Zhu & Fang, 2000). On the other hand, there is also evidence of children's capability of distinguishing living from nonliving things with reference to certain biological properties (Backscheider et al., 1993; Bullock, 1985; Hatano & Inagaki, 1994; Inagaki & Hatano, 1996; Massey & Gelman, 1988; Siegal, 1988; Springer, 1992, 1996a,b; Springer & Keil, 1991; cited in Zhu & Fang, 2000; Rosengren et al., 1991).

This controversy has been argued to be linked to the methodological differences; that is, it is often the case that when kids are asked to classify items as alive or not, they tend to make mistakes; however, when they are led to think based on the properties specific to living things, they have better performance in distinguishing living vs. nonliving things and grouping items correctly (Zhu & Fang, 2000). Zhu & Fang (2000), in both open-ended and controlled experiments on growth and aliveness with Chinese preschoolers, found that preschoolers had some notion of biology. In more specific terms, four- and five-year olds showed an ability to distinguish living from non-living things while they didn't demonstrate an understanding of a link between aliveness and growth yet; however, six-year-olds understood the link between aliveness and growth (Zhu & Fang, 2000).

Though Zhu & Fang (2000) study is informative to a great extent in terms of explaining Chinese children's knowledge of living vs. nonliving things and the controversy in the literature, another extension of this issue could be a study of children's ontological categories, which requires further investigation of properties assigned to various entity categories by children. In this sense, the properties to be investigated are not only aliveness and growth, but also related issues such as origin of existence, having a respiratory system, or dying as well as intentionality (beliefs, plans, and desires) and expression of emotions.

There is also a need for further research on how different languages influence children's development of ontological categories (Zhu & Fang, 2000), and a study on Turkish preschoolers, whose language hasn't been explored in this sense yet, could be interesting.

In the current study, an exploratory experiment was conducted to have a better understanding of Turkish children's ontological categories and their corresponding properties. Ontological categories included living things which involved humans, animals and plants, and nonliving things which were either natural (both dynamic and static representatives) or man-made (both dynamic and static representatives).

In the specification of the category properties and subsequently in the design of the questions, human category was chosen as the reference category inspired by the finding that urban children tend to hold a conception of biological world treating human as a privileged inductive base due to their poor knowledge of animals other than humans (Carey, 1985) and that children's experience with a biological kind affects the strength of that kind as an inductive base (Atran et al., 2001; Ross et al., 2003; Medin & Waxman, in press; cited in Waxman & Medin, 2007). The suggestion could be extended to include plants due to urban children's limited exposure to such observation and experience.

Finally, the properties, which revealed themselves in the questions used in the experiment, allowed not only a discrimination between living vs. nonliving things but also further sub-branching.

2. Research questions and hypotheses

The research questions are:

(1) Do Turkish preschoolers have a general understanding of ontological categories?

(2) Is there an influence of age on Turkish preschoolers' level of understanding of ontological categories?

The hypotheses are:

H1. Turkish preschoolers are expected to have a general understanding of ontological categories.

H2. An age effect is expected for ontological category development in Turkish preschoolers.

3. Method

Sample: The sample consisted of 25 preschoolers: 6 four-year-olds, 9 five-year-olds, and 10 six-year-olds, all of which were selected randomly from a kindergarten, Arı Preschool. There were equal numbers of boys and girls and their parents had varying levels of educational backgrounds. There weren't equal numbers of participants from each age group due to either time limitations or exclusion of unreliable data. There was also an adult control group of 8 people (*M. Age=32*) whose data provided the criteria based on which children's responses were scored. The experiment was carried out by the experimenter herself and there was an occasionally present observer.

Experimental method and apparatus: The real photos were directly presented to the children and their answers to the questions were recorded by circling yes or no in the answer section of the form and notes regarding children's extra explanations were written down in the extra spaces provided.

Stimuli: The material was a series of real photos which included two representatives from each category for validity purposes: (1) humans: boy, girl, (2) animals: dog, turtle, (3) plants: flower, tree, (4) natural and static things: stone, hill, (5) man-made and static things: chair, building, (6) natural and dynamic things: cloud, body of water, (7) man-made and dynamic things: car, motorcycle. The representatives were selected based on several criteria: how familiar children are with these items, how stereotypical these items are, and how much variety within the category can be covered with the inclusion of these specific items. In each pair, the first items comprised the first set (Form A), and the second items comprised the second set (Form B). The photos were standardized for size and resolution quality. For each category, there were a set of questions to be asked: (1) Do X's grow? (2) Are X's alive? (3) Can X's be happy? (4) Can X's have a plan for tomorrow? (5) Can X's want something/some event? (6) Have X's been manufactured? (7) Can X's breathe? (8) Will X's die? ("Büyür mü? Canlı mı? Mutlu olur mu? Yarına bir planı var mı? Bir isteği var mı? Bu üretilmiş mi? Nefes alır mı? Ölür mü?" in Turkish) If the answer to any of these questions wasn't clear, further questions of why or how were asked to get more straightforward answers.

Procedure and location: Each child was tested individually in a quiet room, called the story-reading room, in their kindergarten. There was sufficient light in the room and the child and the experimenter were sitting on the floor on cushions as the children always did in that room. The questions were asked and their responses were recorded by the same female experimenter. There was occasionally an observer, the contact person who received the permission to conduct the experiment in the kindergarten. In the experiment, each subject was presented with one set of 7 photos and for each photo the abovementioned eight questions were asked. If his/her answers were not clear, they were asked further questions of why or how. The whole procedure took ± 10 minutes for each child.

Research design and variables: The children who gave a correct answer (i.e., saying *yes* to the question "Do boys grow?" while being shown the photo of the boy for the human category, saying *no* to the question "Do chairs grow?" while being shown the photo of the chair for the man-made and static category) and an acceptable reason (i.e., for the boy, because he eats a lot and gets bigger, and for the chair, because it can't eat anything or chairs don't change) for each question in each category got one point on that question. Thus the total possible score for one category was eight.

Subjects' scores were entered into a 3 (age 4,5,6) by 2 (Form A or Form B) by 7 (humans, animals, plants, natural and static things, man-made and static things, natural and dynamic things, and man-made and dynamic things) ANOVA, with category type as a within-subject variable and age and form as between-subject variables.

4. Results and Discussion

Subjects' scores were entered into a 3 (age 4,5,6) by 2 (Form A or Form B) by 7 (humans, animals, plants, natural and static things, man-made and static things, natural and dynamic things, and man-made and dynamic things) ANOVA, with category type as a within-subject variable and age and form as between-subject variables. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(20) = 74.62$, $p < .05$); therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .42$). Significant main effects were obtained for age ($F(2,19) = 6.05$, $p < .05$, $\omega^2 = .39$), and category type ($F(2.5,47.31) = 13.67$, $p < .05$, $\omega^2 = .42$). As a result, the first hypothesis that Turkish preschoolers have a general understanding of ontological categories and the second hypothesis that there is an effect of age on ontological category development in Turkish preschoolers are accepted. Thus, the current study indicated that Turkish preschoolers' cognition of ontological categories develop across the specified age range, which is in line with Zhu and Fang's findings (2000).

Form type didn't have a significant effect ($F(1,19) = 0.98$, $p > .05$, $\omega^2 = .05$), which demonstrated the validity of the both sets of representatives from each category.

There was also a significant category type by age interaction when sphericity wasn't assumed ($F(12,30) = 2.73$, $p < .05$, $\omega^2 = .52$). As *post-hoc* analyses revealed and as can be seen from Figure 1, there was a significant difference between 4-year-olds and 6-year-olds in their performance of category specification, whereas 5-year-olds' performance significantly differed from neither 4-year-olds nor 6-year-olds.

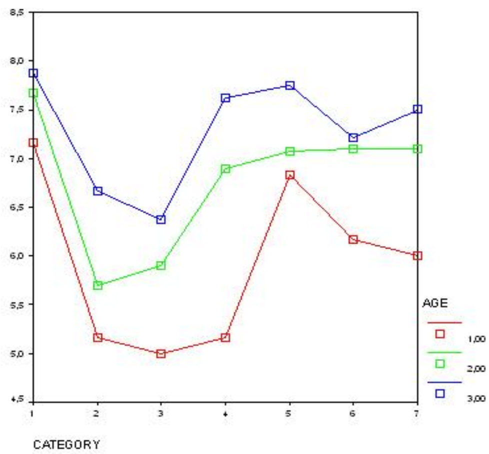


Figure 1. Category type by age (Category type: (1)human, (2)animal, (3)plant, (4)natural and static, (5)man-made and static, (6)natural and dynamic, (7)man-made and dynamic; Age: (1) age 4, (2) age5, (3) age 6)

For all age groups, subjects' performance differed across category types. In the broader category of living things which include humans, animals, and plants, both 4-year-olds and 6-year-olds performed better with humans and animals than with plants, while 5-year-olds performed better with humans and plants than with animals. All age groups did significantly better with humans, which suggests a confirmation of the finding that urban children use the human category as a reference category (Carey, 1985). On the other hand, in the broader category of nonliving things which include things that are natural vs. man-made and static vs. dynamic, 5-year-olds and 6-year-olds exhibited fairly consistent performance across categories and demonstrated better performance in the category of nonliving things as opposed to the category of living things excluding the main referent human category. 4-year-olds did relatively poorly in the category of natural things in comparison with the category of man-made things and similar to their elders, they did better in the category of nonliving things as opposed to animals and plants, excluding humans.

One reason for the relatively poorer performance in the categories of animals and plants in all age groups may be due to their habitual surroundings (urban, in this case) and limited access to experience with such biological kinds, as indicated in some previous studies (Atran et al., 2001; Ross et al., 2003; Medin & Waxman, in press; cited in Waxman & Medin, 2007). There could also be the influence of exposure to cultural beliefs and language input which lead the children to think, for instance, plants *have emotions* or *want water* to grow, or animals *have plans and desires to accomplish*, which is also illustrated in the literature by the cultural belief *vitalism (*a causal model existing in Japan and Australia which depends on the distinctly biological concept of energy) 5-8-year-old Japanese children relied on when understanding bodily processes (Inagaki & Hatano, 1993, 1996, 2002; Hatano & Inagaki, 1994,1999, 2000, cited in Waxman & Medin, 2007). In addition, when children's incorrect answers and their justifications were examined, in the animal category, by

41% of the children, human properties regarding intentionality were overgeneralized to include animals, and in the plant category, it was seen that 64% of the children didn't know plants have some sort of respiratory system.

Again in the same examination, it was seen that the reason for 4-year-olds' much poorer performance with natural nonliving things in comparison with that of 5-and 6-year-olds was due to that they didn't know how these things originated and that they thought these would also grow like living things. The problem might be caused by the fact that such information regarding the origin and nature of these things requires longer periods of observation and tracking changes/stabilities in them. In this sense, the reason for the comparatively better performance of all age groups with man-made nonliving things as opposed to natural nonliving things may be that they are all very familiar with these items and all have had enough life experience to observe them.

Finally, elder children's better performance in comparison with that of 4-year-olds might be due to the possibility that as a result of older age, they have had more learning opportunities either informal (e.g., hands-on experience such as farming, fishing, summer camp activities) or formal, and more access to videos, books, and visits to the zoo (Inagaki, 1990; Rosengren et al., 1991, cited in Waxman & Medin, 2007).

5. Conclusion

The current study indicated that Turkish preschoolers' cognition of ontological categories develop across the specified age range, 4-6 year-olds. There was a significant difference between 4-year-olds and 6-year-olds in their performance of category specification, whereas 5-year-olds' performance significantly differed from neither 4-year-olds nor 6-year-olds. These findings are in support of the previous studies which demonstrated children's capability of distinguishing entities around them with reference to some biological properties.

For all age groups, subjects' performance differed across category types, with humans as the best specified category while animals and plants were the worst in the broader category of living things, and with man-made things as the better specified category over natural things in the broader category of nonliving things. They all did better in the nonliving things category in comparison with the living things, excluding the human category. These findings could be linked to habitual surroundings, cultural conditioning, informal learning opportunities, formal schooling, and limited experience and observation opportunities.

However, there are also limitations to this study. One of the reasons why children had a tendency to overgeneralize human properties to other categories, especially to animals, might be that human category was the reference category in the specification of the properties and the resultant questions to be asked during the experiment. In a future replication of this study, properties specific to other categories could also be included. On the other hand, considering that time is another important limitation, it could be more practical to separate categories into subgroups and conduct successive experiments.

Future research might also do a contrastive analysis of ontological categories specified by rural vs. urban children in order to see if children with more access to experience in nature have any advantage over the others in any of the categories examined.

6. Bibliography

Atran, S., Medin, D.L, Lynch, E., Vapnarsky, V., Ucan Ek' & Sousa, P. (2001) Folkbiology doesn't come from folkpsychology: Evidence from Yukatec Maya in cross-cultural perspective. *Journal of Cognition and Culture* 1, 4-42.

Backscheider, A.G., Shatz, M., & Gelman, S.A. (1993). Preschoolers' ability to distinguish living kinds as a function of regrowth. *Child Development, 64,* 242–1257.

Bullock, M. (1985). Animism in childhood thinking: A new look at an old question. *Developmental Psychology*, *21*, 217–225.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition, 50,* 171–188.

Hatano, G., & Inagaki, K. (1999). A developmental perspective on informal biology. In D.L. Medin & S. Atran (Eds.), *Folkbiology* (pp.321-354). Cambridge, MA: MIT Press.

Hatano, G., & Inagaki, K. (2000). Domain-specific constraints of conceptual development. *International Journal of Behavioral Development, 24*, 267-275.

Inagaki, K. (1990). The effects of raising animals on children's biological knowledge. *British Journal of Developmental Psychology, 8*(2), 119-129.

Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind-body distinction. *Child Development*, 64, 1534-1549.

Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child Development*, *67*, 2823-2840.

Inagaki, K., & Hatano, G. (2002). *Young Children's Thinking About the Biological World*. New York: Psychology Press.

Massey, C.M., & Gelman, R. (1988). Preschooler's ability to decide whether a photographed unfamiliar object can move itself. *Developmental Psychology*, *24*, 307–317.

Medin, D. L. & Waxman, S. R. (in press). Interpreting asymmetries of projection in children's inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive Reasoning* . New York , NY: Cambridge University Press.

Richards, D.D., & Siegler, R.S. (1986). Children's understandings of the attributes of life. *Journal of Experimental Child Psychology*, *42*, 1–22.

Rosengren, K.S., Gelman, S.A., Kalish, C.W., & McCormick, M. (1991). As time goes by: Children's early understanding of growth in animals. *Child Development*, *62,* 1302–1320.

Ross, N., Medin, D.L., Coley, J.D. & Atran, S. (2003). Cultural and experiential differences in the development of folkbiological induction. *Cognitive Development, 18*, 25-47.

Siegal, M. (1988). Children's knowledge of contagion and contamination as causes of illness. *Child Development*, *59*, 1353–1359 .

Springer, K., & Keil, F.C. (1991). On the development of biologically specific beliefs: The case of inheritance. *Child Development*, *60*, 637–648.

Springer, K. (1992). Children's awareness of the biological implications of kinship. *Child Development*, *63*, 950–959.

3

Language and Computation

Long Papers

Ranked multidimensional dialogue act annotation

Marcin Włodarczak

Bielefeld University

Abstract. In this paper we propose a dialogue act annotation system allowing ranking of communicative functions of utterances in terms of their subjective importance. It is argued that multidimensional dialogue act annotation schemes, while allowing more than one tag per utterance, implicitly treat all functions as equally important. Consequently, they fail to capture the fact that in a given context some of the functions of an utterance may have a higher priority than its other functions. The present approach tries to improve on this deficiency. Preliminary results of an annotation experiment suggest that ranking communicative functions accurately reflects the communicative competence of language users.

1 Introduction

Multifunctionality of utterances is often acknowledged in modern dialogue studies [1–3]. It is argued that participants simultaneously address several aspects of communication such as providing feedback, managing the turn-taking process and repairing faulty utterances. Various kinds of implicit functions are an additional source of multifunctionality [4]. The requirement for accounting for multifunctionality of utterances is, of course, also valid for dialogue act annotation schemes. There the notion of multifunctionality is usually introduced explicitly in the form of multidimensional annotation schemes, which allow an utterance to be labelled with more than one tag. However, in such schemes each utterance is represented as an unstructured set of tags. Consequently, they do not reflect the hierarchical organisation of utterance functions determined by speakers' communicative goals. The approach presented here tries to enrich the existing frameworks with a notion of ranking of communicative functions. Importantly, it allows more than one highest-ranking function and more than two different ranks.

The paper has the following structure. In the following section the notion of multidimensional tagsets is introduced. In Sec. 3 existing annotation frameworks are presented alongside the alternative approach proposed in the present paper. The design and the results of an annotation task conducted to validate this framework are presented in Sec. 4, and are followed by conclusions in Sec. 5.

2 Multidimensional Tagsets

Unlike in one-dimensional tagsets, which only allow one tag per utterance, in multidimensional tagsets each utterance can be labelled with multiple tags, each representing a different communicative function. We adopt here the formal definitions of both kinds of tagsets given in [2].

Definition 1. *A one-dimensional tagset is a set* $A = \{a_1, a_2, \ldots, a_N\}$*, each utterance being tagged with exactly one elementary tag* $a_n \in A$*.*

Definition 2. *A multi-dimensional tagset is a collection of dimensions (or classes, categories, etc.)* $\mathcal{T} = \{A, B, \ldots\}$ *where each dimension is in turn a list of tags, say* $A = \{a_1, a_2, \ldots, a_M\}, B = \{b_1, b_2, \ldots, b_N\}$*. When a multi-dimensional tagset is used, each utterance is tagged with a composite* label *or* tuple *of tags* $(a_i, b_j, \ldots)$*.*

Obviously, this is a highly idealised view since it requires that for each utterance a tag is specified in each dimension. If, as is most often the case (see [4]), this requirement is not met and a tag is specified only in some dimensions, the empty tag $\emptyset$ must be added to each of the dimensions In such cases, the empty label $(\emptyset, \emptyset, \ldots, \emptyset)$ must be ruled out. The set of possible labels is then $(A \times B \times C \times \ldots) - (\emptyset, \emptyset, \ldots, \emptyset)$.

Alternatively, rather than employ the notion of the empty tag, only those dimensions can be considered in which a non-empty tag is applicable. This is the approach adopted in [7]:

Definition 3. *A multidimensional dialogue act assignment system is a 4-tuple* $A = (D, f, C, T)$ *where* $D = D_1, D_2, \ldots, D_m$ *is a dialogue act taxonomy with 'dimensions'* $D_1, D_2, \ldots . D_m$*,* f *is a function assigning tags to utterances,* C *is a set of constraints on admissible combination of tags, which additionally allow a dialogue utterance to be assigned a tag in each of the dimensions, but never more than one tag per dimension, and* T *is a set of additional labels that* f *may assign to utterances—*T *contains such labels as* inaudible *or* abandoned[1]*.*

Notably, the set C should be kept relatively small to make *orthogonality* of dimensions as high as possible. This ensures that any combination of tags from different dimensions is admissible [8].

3 Ranked Annotation System

As mentioned above, multifunctionality of utterances is a result of the fact that speakers simultaneously address several aspects of communication. Furthermore, it could be argued that depending on the context specific aspects might be more important than others, thereby forming a hierarchical ordering of functions, a

[1] It could be argued that a 5-tuple should be used instead. The additional element would define a domain of the function f—a set of utterances.

possibility hinted at already in [12]. However, it should be clear that multidimensional dialogue act schemes are not capable of capturing this notion. Instead, they implicitly treat all functions as equally important.

Surprisingly, the problem has received relatively little attention in literature. Bunt and Geertzen [13], discussing their modifications to the kappa statistic, remark that utterances may be argued to have a *primary function* and possibly several *secondary functions*, and note that disagreement about the former is usually more serious than about the latter.

Popescu-Belis observes that although multidimensional tagsets better reflect the multifunctionality of utterances, one-dimensional tagsets offer an advantage of having a much smaller search space, which leads to higher human and automatic annotation accuracy [5]. One of the ways of overcoming the trade-off between a rich pragmatic representation and a smaller search space is only considering the observed tag combinations. For example, the SWDB-DAMSL tagset [9] was developed by clustering 220 DAMSL [10] tag combinations which occurred in 205,000 utterances of the Switchboard corpus into 42 final mutually exclusive tags.

Instead, [5] proposes an alternative strategy. *Dominant Function Approximation* (DFA) assumes that a tagset specifies default values in every dimension based on linguistic and pragmatic grounds or on frequency counts, and states that at most one communicative function of an utterance is non-default (it is then called a *dominant function*). The author notes that while the DFA might be acceptable for current technological applications, it might not be sufficient for a detailed linguistic analysis.

Popescu-Belis tried to verify his hypothesis by checking the number of utterances with more than one non-default functions in existing annotations. Since the number was found to be relatively small (between 3 and 8%), the DFA seems to be correct. However, it could be argued that such findings might be a result of specific annotation guidelines, which often instruct annotators to only mark the most significant function. Indeed, it seems that the possibility of an utterance having several dominant functions cannot be ruled out *a priori*. Moreover, the binary distinction into dominant and default functions may well turn out to be too restrictive.

Alternative Approach. The present approach proposes to model the relative prominence of communicative functions by means of *greater or equal prominence* relation. The term *prominence* will be henceforth used to denote the significance of a communicative function relative to other functions of the same utterance. It is assumed that prominences of every two functions of the same utterance are comparable, i.e. it is possible to decide whether one of the functions is more prominent than the other or whether they are equally prominent. Consequently, the relation in question imposes a *non-strict linear order* on the set of functions of an utterance. Importantly, the ordering of functions is viewed here from the speaker's point of view, i.e. it is assumed that in a given context accomplishing some of the speaker's goals is of greater importance than accomplishing some other goals. The lower-ranking functions may either accomplish ancillary goals or be a means of accomplishing the higher-ranking goals. [17] suggest that entailment relations [4] between communicative functions might be a major factor influencing their relative prominence.

A set of functions of an utterance with equal prominences will be referred to as a *level of prominence*. It should be clear that each level of prominence is an equivalence class given an equivalence relation of *equal prominence*. Obviously, levels of prominence can be also ordered with respect to the prominence of their elements, i.e. one level of prominence precedes another level of prominence if the prominence of functions in the first is greater than the prominence of functions in the second (relation of *strict linear order*).

This approach might be thought of as a generalisation of the approaches outlined above by imposing fewer constraints on the number of levels of prominence. Specifically, multiple functions are allowed to have the same prominence, i.e. every level of prominence may have more than one element. One of the consequences of this is that many dominant (highest-ranking) functions are allowed. Therefore, the approach allows for more flexibility.

It should be also noted that, unlike in the DFA, the notion of default values is not employed here. Moreover, while the DFA was proposed to *simplify* the pragmatic representation of an utterance in order to improve the accuracy of automatic and manual tagging, the present approach aims at *enriching* the pragmatic representation for the needs of linguistic analysis.

Lastly, the concept of the ordering of communicative functions can be easily incorporated into the definition of Multidimensional Dialogue Act Assignment System (Def. 3) to capture the notion of the *Multidimensional Ranked Dialogue Act Assignment System*:

Definition 4. *A Multidimensional Ranked Dialogue Act Assignment System is a 5-tuple $A = (D, f, R, C, T)$ where D, f, C and T are as before, and R is a relation of* greater of equal prominence *holding between functions represented as tags which f assigns to an utterance.*

4 Experiment

The following experiment was conducted to investigate how many dominant functions and how many levels of prominence are identified by annotators. It is based on an analogous experiment proposed by Popescu-Belis [5], namely participants were asked to order functions assigned to segments with respect to their relative prominence. However, unlike in the original design, minimal constraints were imposed on the ordering of functions of utterances. Since approaches like the DFA impose much stricter constraints on an annotation scheme, they would be supported if under these conditions the proportion of utterances with more than one dominant function and more than two levels of prominence was relatively low. Otherwise, the alternative approach outlined above would be more appropriate.

4.1 Experimental settings

HCRC Map Task Corpus [14] was used. Map task dialogues are task related dialogues in which participants cooperate to reproduce a route drawn in one participant's map on the other participant's map. Differences between the maps are introduced to make the task more difficult. The total duration of the data selected for the experiments equalled 4 minutes and 43 seconds.

The tagset chosen for the experiment was the DIT^{++} dialogue act taxonomy [11]. It consists of ten dimensions related to managing the task domain (*Task/Activity*), feedback (*Allo-* and *Auto-feedback*), time requirements (*Time Structuring*), problems connected with production of utterances (*Own* and *Partner Communication Management*), attention (*Contact Management*), discourse structure (*Discourse Structuring*) and social conventions (*Social Obligations Management*).

The data were segmented into functional segments [2] in accordance with [16], and annotated by two experts. 136 functional segments were identified. Full agreement had been reached with regard to segmentation and annotation. Importantly, entailed feedback functions [4] were included in the annotations.

Four naive annotators took part in the experiment. The annotators were undergraduate students at Tilburg University. They had been introduced to the annotation scheme and the underlying theory while participating in a course on pragmatics. The course comprised approximately three hours of lectures and a few small annotation exercises on data other than map task dialogues.

All annotators accomplished both tasks individually, having received the materials (transcriptions and sound files) in electronic form. Time for the task was not limited. To encourage high quality of annotations the students were motivated by an award of 10% of the total grade for the pragmatic course.

The participants' task was to order utterance functions to order the functions assigned to utterances with respect to their relative importance. The ordering was done by assigning each function a numerical value from the set of *consecutive* natural numbers, starting from "1" as the most prominent function. The lowest possible rank was, therefore, equal to the number of utterance functions. However, more than one function could be assigned the same numerical value.

4.2 Results and Discussion

Since it was observed that participants failed to rank functions of some segments, the total number of analysed rankings was equal to 293 (243 and 55 for segments with two and three functions respectively). Cohen's kappa [18] was calculated for 54 segments (44 and 10 with two and three functions respectively) ranked properly by all four participants.

Inter-rater agreement values for functions assigned specific ranks are given in Tab. 1 and 2. As can be observed, mean kappa values indicate fair to moderate

[2] [15] defines a *functional segment* as a "minimal stretch of communicative behaviour that has one or more communicative functions."

agreement. It should be borne in mind, however, that while the participants had some experience using the DIT^{++} tagset, they were completely naive with regard to ranked annotation. It could be, therefore, hoped that more experienced annotators could achieve much higher agreement. Moreover, kappa values do not seem to decrease substantially across ranks. Indeed, while the number of segments with three functions was rather low, in Tab. 2 agreement for the third rank was higher than for the second rank. These results contrast sharply with the assumptions of the DFA, which would predict that agreement values should drop across ranks.

Table 1. Kappa coefficient values for functions assigned specific ranks in two-functional segments.

Annotators	Rank 1	Rank 2
1 & 2	0.46	0.35
1 & 3	0.64	0.67
1 & 4	0.34	0.32
2 & 3	0.27	0.21
2 & 4	0.41	0.37
3 & 4	0.47	0.49
Mean	0.43	0.40

Table 2. Kappa coefficient values for functions assigned specific ranks in three-functional segments.

Annotators	Rank 1	Rank 2	Rank 3
1 & 2	0.75	0.38	0.41
1 & 3	0.29	0.37	0.55
1 & 4	0.49	0.21	0.54
2 & 3	0.31	0.23	0.14
2 & 4	0.51	0.08	0.44
3 & 4	0.51	0.56	0.44
Mean	0.48	0.30	0.42

By comparison, [19] reports results of a ranking experiment using a simplified versions of the DIT^{++} tagset and ten completely naive raters. Perhaps not surprisingly, some the the observed kappa values were lower than those in the present study. More interestingly, however, the value for the first rank was substantially higher than for the remaining ranks. Specifically, it was found that for two-functional segments inter-rater agreement was equal to 0.39 for the first rank and 0.1 for the second rank. In the category of utterances with three functions kappa values equalled 0.18, 0.04 and 0.04 for the ranks of one, two and three. Notably, the values for the ranks of two and three are identical, which might indicate that functions with these ranks did not differ much with regard to their relative prominence. While these results are in accordance with the DFA, it should be noted that participants had no experience not only with ranking but with the tagset itself. This suggests that the DFA could prove more useful when completely naive annotators are used.

Proportions of utterances with different numbers of identified levels of prominence are presented in Fig. 1. Overall, in 97% of segments the number of identified levels of prominence was equal to the number of segment functions. Only in three out of 243 two-functional segments, and five out of 55 three-functional segments was it otherwise. Since minimally two levels of prominence were identified in three-functional segments, at most two functions were assigned the same rank. However, all these cases came from the same annotator, and might, therefore, be highly idiosyncratic.

The DFA predicts that the proportion of utterances with more than two levels of prominence should be small. Obviously, since utterances with two functions can be assigned the maximum of two distinct ranks, only three-functional segments are of interest in this respect. Although there were relatively few such segments, as much 91% of them would not be represented correctly if more restrictive annotation guidelines, such as the DFA, were adopted.

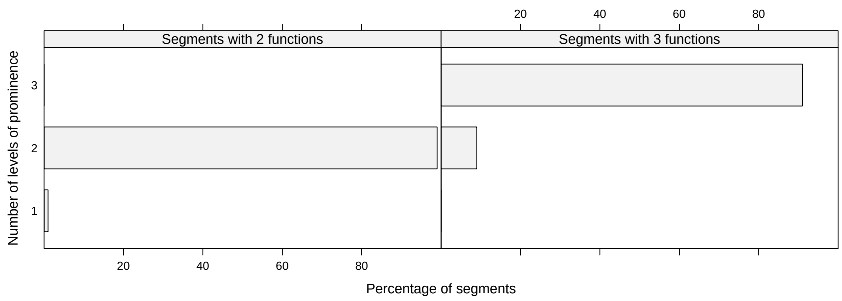


Fig. 1. Proportions of segments with different numbers of levels of prominence

Fig. 2 presents proportions of utterances with different numbers of identified dominant functions (i.e. functions assigned the rank of one). Here the overwhelming tendency is for a segment to have exactly one such function. This was the case for 99% of two-functional segments and 91% of three-functional segments. The remaining cases again came from the same annotator.

Considering the results regarding the numbers of dominant functions and levels of prominence together, it should be said that there is a very strong tendency for each function to be assigned a different rank. The relation of *greater or equal prominence* is, therefore, in most cases a relation of *greater prominence*, i.e. it is a relation of *strict* linear order. Consequently, the DFA is only partially correct. It is right in predicting one dominant function per segment but does not differentiate between the prominences of non-dominant functions. However, it is interesting to note that whenever the same rank was assigned to two functions, it was in fact the first rank in all but one case.

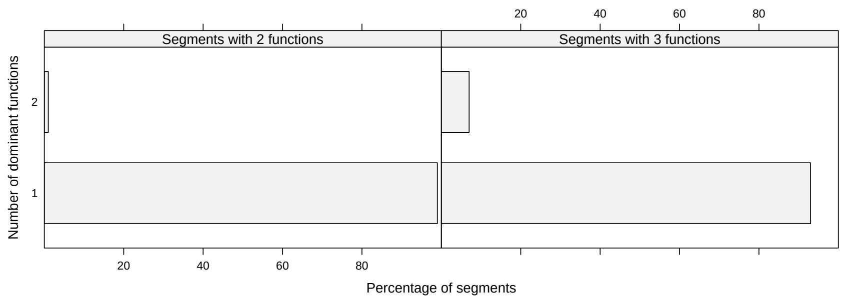


Fig. 2. Proportions of segments with different numbers of dominant functions

Figure 3 presents distributions of functions of two-functional segments belonging to specific dimensions across ranks. While functions from most dimensions are assigned the ranks of one and two with comparable frequencies, there is a noticeable difference between frequencies of *Turn Management* and *Feedback* functions. Specifically, *Feedback* functions are the most frequent of functions assigned the rank of one (38%), and *Turn Management* functions are the second most frequent (29%). By contrast, among functions assigned the rank of two it is the other way round with *Turn Management* functions comprising 43%, and *Feedback* functions comprising 30%. Additionally, *Task Management* functions have a higher frequency among the functions ranked second (18%) than among those ranked first (11%). Two-tailed Fisher's exact test was conducted to test whether the proportions between functions depend on the rank. The result was statistically significant with a p-value of 0.01.

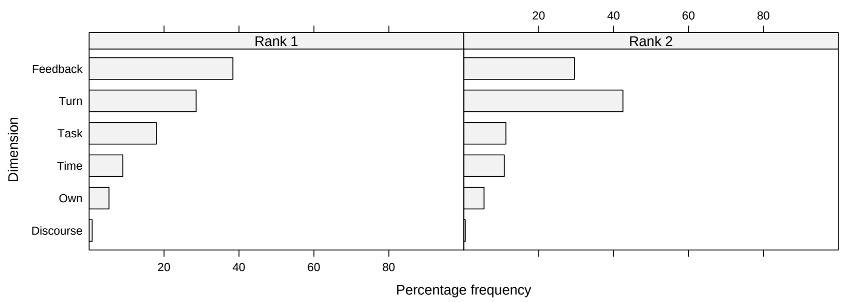


Fig. 3. Frequency distribution of ranks assigned to functions from specific dimensions in two-functional segments. The dimension names were abbreviated as follows: *Feedback–Auto-* and *Allo-feedback* clustered together, *Turn–Turn Management*, *Task–Task Management*, *Time–Time Management*, *Own–Own Communication Management*, *Discourse–Discourse Management*.

Analogous result for utterances with three functions are presented in Fig. 3. Here, except for minor differences among low frequency *Own Communication Management*, *Time Management*, and *Discourse Management* functions, the greatest differences concern functions from the *Feedback*, *Turn Management* and *Task* dimensions. While *Feedback* functions have the highest frequency across all three ranks but their dominance over the other two dimensions varies greatly depending on the rank. Among functions assigned the rank of one *Feedback* functions make up 41%, *Turn Management* functions make up 25%, and *Task* functions make up 14%. This difference is even larger among functions ranked second with respective frequencies of 57%, 14% and 10%, but is almost nonexistent in the category of functions ranked third, where their frequencies equal 28%, 26% and 26%. Two-tailed Fisher's exact test was again conducted to test whether the proportions between functions depend on rank. The result was statistically insignificant with a p-value of 0.06.

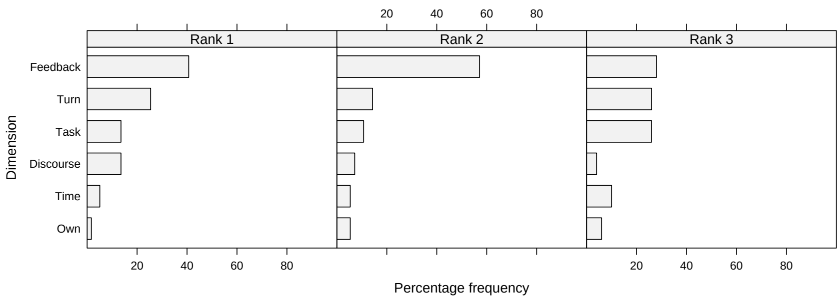


Fig. 4. Frequency distribution of ranks assigned to functions from specific dimensions in three-functional segments. For the explanation of the dimensions names abbreviations see Fig. 3.

5 Conclusions

The results reported above show clearly that in a great majority of cases the number of identified levels of prominence tends to be equal to the number of segment functions. In other words, each function is usually assigned a different rank. Therefore, the relation proposed in Sec. 3 was in most cases a relation of *strict* linear order. Apart from that, frequencies of functions from respective dimensions were found to depend on rank in case of two-functional segments, and to be independent of it in case of three-functional segments. However, since the analysed dataset (and, in particular, the number of segments with three functions) was relatively small, these results should be treated as preliminary.

In the light of these findings it must be said that the DFA is right in predicting that most segments have just one highest-ranking function but it fails to account for distinctions among lower-prominence functions. It is, of course, a question of specific research goals whether the resulting underspecification is considered acceptable. Regarding the notion of default values assumed in the DFA, the fact that each function was assigned a different rank in most of the three-functional segments seems to suggest that the usefulness of this notion is limited. Additionally, contrary to the assumptions of the DFA, inter-annotator agreement values were found to be similar across all ranks.

Obviously, the results obtained here should ideally be confirmed in a larger scale annotation experiment. In addition, a number of issues not discussed here could also be investigated. For example, rather that analyse frequencies of functions across ranks globally, relative prominences of specific *combinations* of functions could be analysed. This, in turn, should shed more light on the problem of default functions assumed in the DFA.

6 Acknowledgements

I would like to thank Harry Bunt and Volha Petukhova (Tilburg University) for granting me access to their data.

References

1. Bunt, H. (2000) *Dialogue pragmatics and content specification* In H. Bunt. and W. Black (eds.), Abduction, Belief and Context in Dialogue. Amsterdam: Benjamins
2. Popescu-Belis, A. (2005). *Dialogue Acts: One or More Dimensions?* ISSCO Working Paper, 62, University of Geneva
3. Allwood, J. (2000). *An activity-based approach to pragmatics* In H. Bunt. and W. Black (eds.), Abduction, Belief and Context in Dialgoue. Amsterdam: Benjamins
4. Bunt, H. (2009). Multifunctionality and multidimensional dialogue semantics. *Proceedings of DiaHolmia 2009*, Stockholm
5. Popescu-Belis, A. (2008). Dimensionality of dialogue act tagsets: an empirical analysis of large corpora. *Language Resource and Evaluation 42 (1)*
6. Bunt, H. (2006). Dimensions in Dialogue Act Annotation *Proceedings of LREC 2006*, Paris
7. Bunt, H. and Girard, Y. (2005). *Designing an Open, Multidimensional Dialogue Act Taxonomy.* In C. Gardent and B. Gaiffe (eds.) DIALOR'05, Proceedings of the Ninth International Workshop on the Semantics and Pragmatics of Dialogue, Nancy
8. Petukhova, V. and Bunt, H. (2009). The independence of dimensions in multidimensional dialogue act annotation. *Proceedings of the 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Boulder, Colorado
9. Jurafsky, D., Shriberg, E. and Biasca, D. (1997). *Switchboard SWBD-DAMSL Labelling Project Coder's Manual. Draft 13.* Technical Report 97-02. University of Colorado, Institute of Congnitive Science.
10. James Allen and Mark Core. (1997). *DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1).* Technical Report, Multiparty Discourse Group, Discourse Resource Initiative

11. Bunt, H. (2009). The DIT^{++} taxonomy for functional dialogue markup. *Proceedings of the AAMAS 2009 Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts" (EDAML 2009)*, Budapest
12. Jakobson, R. (1960). *Linguistics and Poetics.* In Sebeok (ed.) Style in Language, 350–377, MIT Press: Cambridge, MA
13. Geertzen, J. and Bunt, H. (2006). Measuring annotator agreement in a complex hierarchical dialogue act scheme. *Proceedings of SIGDIAL 2006*, Sydney
14. Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34
15. ISO (2009). *Semantic annotation framework (SemAF), Part 2: Dialogue acts.* ISO CD 24617-2, October 2009. ISO, Geneva.
16. Geertzen, J., Petukhova, V. and Bunt, H.. (2007). A Multidimensional Approach to Utterance Segmentation and Dialogue Act Classification. In: *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, Antwerp, pp. 140–149.
17. Włodarczak, M., Bunt, H. and Petukhova, V. (2010). Entailed feedback: evidence from a ranking experiment. In P. Łupkowski and M. Purver (eds.) *Aspects of Semantics and Pragmatics of Dialogue. Semdial 2010, 14th Workshop of the Semantics and Pragmatics of Dialogue*, Poznań
18. Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol.20, No.1, pp. 37–46.
19. Włodarczak, M. (2009). *Ranked multidimensional dialogue act annotation.* MA thesis, Adam Mickiewicz University, Poznan.

Predicting the Position of Attributive Adjectives in the French NP

Gwendoline Fox[1] and Juliette Thuilier[2]

[1] University of Paris 3 - Sorbonne Nouvelle (ILPGA) and EA 1483
[2] University of Paris 7 - Denis Diderot (UFRL) and ALPAGE (INRIA)

1 Introduction

French displays the possibility of both pre-nominal and post-nominal ordering of adjectives within the noun phrase (NP)[3].

(1) un magnifique tableau / un tableau magnifique
a magnificent painting / a painting magnificient
"a magnificient painting"

(2) un beau tableau / ??un tableau beau / un tableau très beau
a nice painting / a painting nice / a painting very nice
"a nice painting"

The above examples show that the positionning of attributive adjectives is a complex phenomenon: in (1) the adjective *magnifique* may be in both positions while *beau* strongly prefers anteposition, unless it is modified by an adverb, as seen in (2).

The question of adjective alternation has led to many studies in French linguistics ([1,2,3,4,5,6] among others). The constraints playing a role in this phenomenon are said to be phonological, morphological, syntactic, semantic, discursive and also pragmatic. Only one of the proposed constraints is categorical in the sense that it imposes a specific position to an attributive adjective: the presence of a post-adjectival complement (3) or modifier (4) only allows postposition of the adjective.

[3] The position of the adjective can imply semantic change, for specific adjectives (i) or specific noun-adjective combinations (ii):

(i) un coffre ancien vs. un ancien coffre
a chest old a old chest
" an old chest vs. a former chest"

(ii) un gros fumeur / un fumeur gros / un gros chanteur
a big smoker a smoker big a big singer
"an heavy smoker / a fat smoker / a fat singer"

We do not take into consideration these kind of semantic changes in this article. Our work focuses on the form of the adjective.

(3) un homme fier de son fils / *un fier de son fils homme
a man proud of his son / a proud of his son man
"a man proud of his son"

(4) un entretien long de deux heures / *un long de deux heures entretien.
a interview long of two hours / a long of two hours interview
" a two hours long interview "

The other constraints participating in the alternation between anteposition and postposition are not categorical. For instance, as noted in the corpus studies of [1,2], length and frequency are preferential constraints: short adjectives, as well as the most frequent ones, tend to be anteposed.

In order to account for preferential constraints, we present along the same lines as [7], a quantitative study of the position of attributive adjectives, based on two corpora: the French Tree Bank (henceforth FTB) and the Est-Républicain corpus (henceforth ER). The aim of this article is to propose a prediction model based on interpretable constraints and to compare their prediction capacities in order to better estimate their respective contribution in the choices guiding the placement of adjectives.

2 Methodology

Building the datatable The first step of this work is to collect the data concerning adjectives and capture the constraints found in the literature. The study is based on the functionnally annotated subset of the FTB corpus [8][4], which contains 12351 sentences, 24098 word types and 385458 tokens. It is, for the moment, the only existing treebank for French. We extracted all the occurrences of attributive adjectives from this corpus[5], and filtered out numeral adjectives[6], adjectives appearing in dates[7], abbreviations[8] and incorrectly annotated occurrences. We also discarded the 438 adjectives occurring with a post-adjectival dependent since postposition is imposed (see (3) and (4)) by a categorical constraint that overrides any other preferential constraint. The remaining adjectives constitute the basis of the datatable, to wich we have added information on the position of each adjective with respect to the noun it modifies, and 11 other variables that we describe in section 3.

Three variables of our study (Freq, CollocAnt and CollocPost) were extracted from the ER corpus for more reliable counts. The raw corpus is a 147,934,722 tokens corpus, available on the ATLIF website[9]. It was tagged and

[4] This subset corresponds to the part that was manually corrected.

[5] We identified attributive adjectives using the following pattern in the treebank : an adjective occuring with an nominal head within a NP is an attributive adjective.

[6] Cardinal numerals such as *trois* 'three', *vingt* 'twenty', *soixante* 'sixty'... are sometimes annotated as adjectives in the FTB.

[7] Examples of dates containing adjectives : "$[\mathit{13}]_{ADJ}$ $[\mathit{mars}]_N$", "$[\mathit{lundi}]_N$ $[\mathit{31}]_{ADJ}$".

[8] Nouns or adjectives are viewed as abbreviations if their last letter is a capital letter.

[9] *http://www.cnrtl.fr/corpus/estrepublicain/*

lemmatized with the *Morfette* system [9] adapted for French. This corpus was used to compute the frequency of every adjectival lemma as well as Adjective-Noun and Noun-Adjective collocations.

The datatable contains 14804 occurrences corresponding to 1920 adjectival lemmas. 4227 (28.6%) tokens appear in anteposition, and 10577 (71.4%) in postposition. Table 1 shows that the adjectival lemmas displaying position alternance represent only 9.5% of all lemmas, yet these few lemmas correspond to 5473 occurrences, i.e. 37.0% of the datatable, which means that they are highly frequent adjectives.

Note that among the alternating adjectives (occurring in both positions), the ratio between anteposed and postposed occurrences is the reverse from that of all adjectives: there are 3727 anteposed (68,1%) and 1746 postposed (31,9%) adjectives. Alternating adjectives thus show a preference for anteposition. The general pattern is therefore that postposed adjectives tend to be infrequent lemmas occurring only in postposition, whereas alternating adjectives tend to be frequent and to prefer anteposition.

	anteposed	*postposed*	*both positions*	*Overall*
number of lemmas	125	1613	182	1920
	6.5%	*84.0%*	*9.5%*	*100%*
tokens	500	8831	5473	14804
	3.4%	*59.7%*	*37.0%*	*100%*

Table 1. Distribution of adjectival lemmas and tokens according to position

Statistical inference and logistic regression We used logistic regression models [10] in order to best estimate the distribution of adjective positions using the variables from the datatable. Formally, a logistic regression is a function for which values can be interpreted as conditional probabilities. Its analytical form is as follows:

$$\pi_{\text{ante}} = \frac{e^{\beta \mathbf{X}}}{1 + e^{\beta \mathbf{X}}} \tag{1}$$

where, in our case, π_{ante} is the probability for the adjective to be anteposed and β corresponds to the abbreviation of the sequence of regression coefficients α, β_0... β_{n}, respectively associated with the predicting variables X_0... X_{n}. Given a scatter plot, the calculation of regression consists in the maximum likelihood estimation of α and β_{i} parameters for each variable in a *logit* space.

This type of modelling consists in the combining of several explicative variables (binary or continuous) to predict the behaviour of a single binary variable, here the position of the adjective. More precisely, we estimate the probability of anteposition as a function of 11 variables. Given one adjectival occurrence and the value of the 11 explanatory variables attributed to this occurrence, the model predicts postposition if the probabilty is lower than 0.5, and anteposition if the probability is higher or equal to 0.5.

In order to evaluate the relevance of the constraints, we compare prediction models based on different constraint clusters. We use a 10-fold cross-validation to compute the accuracy of each model (noted μ and its standard deviation σ). The accuracy represents the proportion of data that is correctly predicted.

The comparison of the different models takes as a reference the accuracy of the *baseline model*: $\mu = 71.4\% (\sigma = 0.019)$. This model does not contain any explanatory variables and systematically predicts postposition. Its accuracy thus corresponds to the proportion of postposed adjectives in the datatable.

3 Variables

The variables we use in our logistic regression models are derived from the constraints found in the literature on attributive adjectives in French. They are summarized in table 2. Each model is based on different sets of constraints according to specific properties. The first set (COORD and ADV) concerns the syntactic environment of the adjective, the second is based on length and frequency (ADJ-LENGTH, AP-LENGTH and FREQ), the third one on the lexical properties of the adjectival item (DERIVED, NATIO, COLOUR and INDEF). Finally, the fourth group examines collocationnal effects of the Noun - Adjective combination (COLLOCANT and COLLOCPOST).

Variables	Types	Description
COORD	*bool*	adjective in coordination or not
ADV	*bool*	adjective with pre-modifying adverb or not
ADJ-LENGTH	*real*	length of the adjective in syllables
AP-LENGTH	*real*	length of the AP in syllables
FREQ	*real*	adjective frequency in the ER corpus
DERIVED	*bool*	derived adjective or not
NATIO	*bool*	adjective of nationality or not
COLOUR	*bool*	adjective of colour or not
INDEF	*bool*	indefinite adjective or not
COLLOCANT	*real*	score for the adjective-noun bigram
COLLOCPOST	*real*	score for the noun adjective bigram

Table 2. Summary table of variables and their values (*bool* = boolean and *real* = real number)

Coordination (COORD*)* In a competence account of attributive position like in [5], the position of coordinated adjectives is not restricted, as can be seen in example (5) (from [5]).

(5) une belle et longue table / une table belle et longue
a beautiful and long table / a table beautiful and long
"a long and beautiful table"

However, 94.6% of coordinated adjectival occurrences (i.e. 758 occurrences) are postposed in our data. Usage-based data thus suggests that coordination is a factor that strongly favours postposition.

*Presence of a premodifying adverb (*ADV*)* The general constraint is the same as for coordination: the presence of a pre-adjectival modifier does not restrict the position of the modified adjective (example (6)).

(6) une très longue table / une table très longue
a very long table / a table very long
"a very long table"

[5] point out that adjectives can be postposed with any adverb whereas only a small set of adverbs allows anteposition. This is confirmed in our datatable: 11 types of adverb[10] are observed with anteposed adjectives, while 119 different types appear with adjectives in postposition. Furthermore, the adverbs found with adjectives in anteposition are not specific to this position, they also appear with postposed occurrences. From a general quantitative point of view, 74.9% of the premodified adjectival occurrences are in postposition.

Length Numerous works on word order use the notion of length: for attributive adjectives in French [1,2], for word [11,12] and constituent [13,14,15,7] alternation in other languages. The main idea is expressed by the principle *short comes first*, i.e. short elements tend to appear first. Here, we consider length in terms of number of syllables and we introduce two variables: length of the adjective (ADJ-LENGTH) and length of the adjectival phrase (AP) (AP-LENGTH)[11].

*Lemma frequency (*FREQ*)* In his corpus study, [2] observes that high frequency is correlated with anteposition. In this work, we built a dictionary of frequency for each adjectival lemma in the ER corpus. We consider that frequency in ER better estimates the probability of use of an adjective than frequency in FTB, given the importance of the data (almost 1.5 million words for ER vs. 385,000 words for FTB)[12].

*Derived adjectives (*DERIVED*)* Adjectives may be derived from other parts-of-speech: e.g. from verbs (past participles, present participle, by suffixation: *-ible* 'faillible' (*faillible) / -able* 'faisable' (*doable*) /*if* 'attractif' (*attractive*) or from

[10] The 11 adverbs are : 'encore' *again*, 'désormais' *from now on*, 'moins' *less*, 'peu' *not much*, 'plus' *more*, 'si' *so*, 'tout' *very*, 'très' *very*, 'trop' *too*, 'bien' *well*, 'aussi' *also*.

[11] We obtain the number of syllables using the speech synthesis software ELITE [16]. It counts the number of syllables for every token, taking into account the actual form of the adjective (feminine versus masculine, for instance) as well as the possible effects of sandhi phenomena, like the *liaison* phenomenon. The value associated to each adjectival type corresponds to the mean of all its tokens length.

[12] The FREQ value of an adjective is 0 if the adjective is in the datatable but not in the ER corpus. Frequency being considered as a mere estimator, the model will handle such data similarly to very low non-null values.

nouns ('métallique' (*made of metal*), 'scolaire' (*academic*), 'présidentiel' (*presidential*)). These adjectives are described as prefering postposition. We marked them with the variable DERIVED[13].

Lexico-semantic classes Most reference grammars state that objective adjectives (i.e. adjectives for which the semantic content is perceptible or can be infered from direct observation) are postposed. Objective adjectives are classified into sub-groups like form, colour, physical property, nationality, technical terms... In order to estimate the relevance of lexico-semantic classes for the placement of adjectives, we test the predictive capacity of two classes by means of two variables: NATIO for nationality[14] and COLOUR for colour[15].

We also added the class of indefinite adjectives. These adjectives are special in the fact that their syntacitc properties show a hybrid behaviour between determiners and adjectives. On the one hand, indefinite adjectives may introduce and actualise the noun, like determiners. On the other hand, they may co-occur with a determiner and can be placed in post-nominal position, even though they favour anteposition. These latter properties are specific to attributive adjectives. The adjectives we identified as indefinite in our datatable are: 'tel' (*such*), 'autre' (*other*), 'certain' (*some/sure*), 'quelques' (*few*), 'divers' (*various*), 'différent' (*different*), 'maint' (*numerous*), 'nul' (*null/lousy*), 'quelconque' (*any/ordinary*), 'même' (*same/itself*). They are marked by means of the variable INDEF.

Collocations It is well known that the nature of some Adjective-Noun combinations is strongly collocational. This implies that the position of attributive adjectives in French should also be influenced by collocational effects. The collocation score in our datatable corresponds to the frequency of the Adjective-Noun (COLLOCANT) and Noun-Adjective (COLLOCPOST) bigrams in the ER corpus. We use raw frequency relying on the idea that the frequency of stored elements directly affects the representation of these elements [19]. As a further support, the experiment conducted by [20] shows that judgements elicited from human subjects about adjective-noun pairs in English are highly correlated with the co-occurrence frequency[16]. This suggests that frequency is a good association measure of collocational bigrams.

[13] The adjectives derived from another part-of-speech (noun or verb) are collected using the software of derivational morphological analysis DERIF [17].

[14] Using the dictionnary PROLEXBASE [18].

[15] Using the dictionnary CHROMA: *http://pourpre.com/chroma/*.

[16] The authors pointed out that frequency has the best correlation score compared to other association scores: conditional probability of the noun given the adjective, log-likelihood ratio, selectional association measure.

$$\pi_{\text{ante}} = \frac{e^{\mathbf{X}\beta}}{1+e^{\mathbf{X}\beta}}, \text{ where}$$

$$\begin{array}{rlll}
X\beta = & +0.86 & & *** \\
& -0.53 & \textsc{Coord} = 1 & * \\
& -0.55 & \textsc{adj-length} & *** \\
& -0.41 & \text{AP-}\textsc{length} & *** \\
& +0.00003 & \textsc{Freq} & *** \\
& -0.47 & \textsc{derived} = 1 & *** \\
& +1.74 & \textsc{Indef} = 1 & *** \\
& -5.25 & \textsc{natio} = 1 & *** \\
& -15.05 & \textsc{colour} = 1 & \\
& +0.003 & \textsc{CollocAnt} & *** \\
& -0.003 & \textsc{CollocPost} & ***
\end{array}$$

Fig. 1. Formula of prediction model, significant effects are coded *** p<0.001, ** p<0.01, * p<0.1

4 Prediction model of attributive adjective position

The prediction model is built with all the variables described in part 3 and maximized with a backward elimination procedure based on AIC criterion [21][17]. The ADV constraint's contribution to the model is not significant according to the procedure. It was thus eliminated. The model is presented in figure 1.

As we expected, the variables COORD, ADJ-LENGTH, AP-LENGTH, DERIVED, NATIO, COLOUR and COLLOCPOST tend to favour postposition, whereas FREQ, INDEF and COLLOCANT vote for anteposition.

Compared to the *baseline model* performances ($\mu = 71.4\%, \sigma = 0.019$), this model has a significantly better accuracy ($\mu = 88.6\%$, $\sigma = 0.01$). The prediction performances are presented in table 7.

In order to compare the effect of different constraint clusters, we built 4 prediction models based on different groups of variables: a *Syntactic model* containing COORD[18]; a *Lexical property model* with NATIO, COLOUR, INDEF and DERIVED; a *Frequency-Length model* containing the variables ADJ-LENGTH, AP-LENGTH and FREQ and a *Collocations model* containing COLLOCANT and COLLOCPOST.

*Syntactic model (*COORD*).* First, the comparison shows that the effect of the syntactic constraint COORD is insignificant when it is not combined with other constraints (*Syntactic model* accuracy: $\mu = 71.4, \sigma = 0.02$). Moreover, the elimination of the ADV constraint from the prediction model strengthens the idea that syntactic constraints have no important predictive power. This can be partly ex-

[17] Note that for this particular model forward selection procedure gives the same results.

[18] We do not integrate the ADV constraint in the *Syntactic model* because the elimination with the AIC procedure already shows its lack of predictive power within the global model.

		Predicted position		% Correct
		P	A	
observed	P	10208	369	96.5%
position	A	1323	2904	68.7%

Overall accuracy: $\mu = 88.6\%$ ($\sigma = 0.001$)

Table 3. Classification table for *prediction model*

plained by the fact that these two variables are relevant for a very small set of data: ADV and COORD represent respectively 5.2% and 5.4% of all the data.

		Predicted position		% Correct
		P	A	
observed	P	10574	3	99.9%
position	A	4227	0	0%

Overall accuracy: $\mu = 71.4$ ($\sigma = 0.02$)

Table 4. Classification table for *Syntactic model*

*Lexical properties model (*NATIO, COLOUR, INDEF *and* DERIVED*).* Second, lexical properties are relevant when they are not combined with the other constraints (*Lexical properties model* accuracy: $\mu = 74.7, \sigma = 0.02$). This observation encourages us to extend the number of semantic classes in order to improve our modelling.

		Predicted position		% Correct
		P	A	
observed	P	10506	71	99.3%
position	A	3681	546	12.9%

Overall accuracy: $\mu = 74.7$ ($\sigma = 0.02$)

Table 5. Classification table for *Lexical properties model*

*Frequency-Length model (*ADJ-LENGTH, AP-LENGTH *and* FREQ*).* Third, we note that the variables of length and frequency have the most important predictive power (*Frequency-Length model* accuracy: $\mu = 85.8, \sigma = 0.009$). It is interesting that both variables are said to be led by processing ease. Length is positively correlated with complexity, be it from an articulatory point of view [22,23] or a syntactic one [14,24]. Based on the idea that the general processing cost is reduced when less complex elements precede more complex ones, it is expected that short adjectives should favour anteposition. Likewise, as [19] argues, each use of a word is stored in the memory of locutors and added to the mental representation that they have of it. Each occurrence reinforces the mental representation and makes it more accessible for the locutor to process. In other words, a highly frequent item is highly accessible, and thus easy to process. Consequently, highly frequent adjectives are also expected to be anteposed for the general processing ease of the NP. The importance of both constraints in the prediction may be viewed as a support for these claims[19].

		Predicted position		% Correct
		P	A	
observed	P	10122	455	95.7%
position	A	1654	2573	60.9%

Overall accuracy: $\mu = 85.8$ ($\sigma = 0.009$)

Table 6. Classification table for *Frequency-Length model*

*Collocations model (*COLLOCANT *and* COLLOCPOST*).* Fourth, the *Collocations model* shows that the frequency of bigrams also represents a good predictor (*Collocations model* accuracy: $\mu = 79.7, \sigma = 0.013$). This observation is linked to the above-mentioned idea that frequency of use has an effect on mental representations. Indeed, as in [19] and in works on construction grammar [25,26], we may hypothesize that units larger than words are stored.

		Predicted position		% Correct
		P	A	
observed	P	10495	82	99.2%
position	A	2930	1297	30.7%

Overall accuracy: $\mu = 79.7$ ($\sigma = 0.013$)

Table 7. Classification table for *Collocations model*

To conclude this section, note that a large proportion of constraints playing a significant role in the studied phenomenon relates to the adjectival item (ADJ-LENGTH, FREQ, NATIO, COLOUR, INDEF and DERIVED). The choice of position thus seems largely determined by the use of a specific item.

[19] Note that frequencies are biased by the journalistic nature of corpora: adjectives of nationality are frequent despite the fact that they are postposed in most cases. Nevertheless, the variable NATIO of the global prediction model votes for postposition, which neutralizes the frequency effect.

5 Conclusion

We examined the question of position alternation of attributive adjectives in French using quantitative methods applied to corpora. One can draw several conclusions from the logistic regression models that we proposed. First the satisfactory results of our general model show that a good part of the modeling can be done on the basis of the form without considering the semantics due to position. Moreover, this work points out that lexical properties, including semantic classes, are relatively good predictors. These conclusions suggest that our knowledge on the very nature of adjectival items plays an important role in their positionning. Nevertheless, the prediction performances may be improved by taking more semantics into account: adding information for other semantic classes should naturally enhance the model. The question however remains on how to capture and formalise semantic relations in a quantitative study. Finally, our work outlines the importance of length and frequency-based constraints. This confirms the role of the nature of adjectives, but it also shows that usage strongly participates to the building of linguistic knowledge, and hence to the positionning of adjectives.

References

1. Forsgren, M.: La place de l'adjectif épithète en français contemporain, étude quantitative et sémantique. Almqvist & Wilksell, Stockholm (1978)
2. Wilmet, M.: La place de l'épithète qualificative en français contemporain : étude grammaticale et stylistique. Revue de linguistique romane **45** (1981) 17–73
3. Delbecque, N.: Word order as a reflexion of alternate conceptual construals in french and spanish. similarities and divergences in adjective position. Cognitive Linguistics **1** (1990) 349–416
4. Nølke, H.: Où placer l'adjectif épithète? focalisation et modularité. Langue française **111** (1996) 38–57
5. Abeillé, A., Godard, D.: La position de l'adjectif èpithète en français : le poids des mots. Recherches linguistiques de Vincennes **28** (1999) 9–32
6. Noailly, M.: L'adjectif en français. Ophrys (1999)
7. Bresnan, J., Cueni, A., Nikitina, T., Baayen., H.: Predicting the dative alternation. In Boume, G., Kraemer, I., Zwarts, J., eds.: Cognitive Foundations of Interpretation. Royal Netherlands Academy of Science, Amsterdam (2007)
8. Abeillé, A., Barrier, N.: Enriching a french treebank. In: Proceedings of Language Ressources and Evaluation Conference (LREC), Lisbon (2004)
9. Grzegorz Chrupala, G.D., van Genabith, J.: Learning morphology with morfette. In Calzolari, N., *et al*, eds.: Proceedings of LREC'08, Morocco, ELRA (2008)
10. Agresti, A.: An introduction to categorical data analysis. Wiley interscience (2007)
11. Cooper, W.E., Ross, J.R.: World order. In: Papers from the Parasession on Functionalism. Chicago Linguistic Society (1975) 63–111
12. Benor, S.B., Levy, R.: The chicken or the egg? a probabilistic analysis of english binomials. Language **82(2)** (2006) 28–55
13. Hawkins, J.: The relative order of prepositional phrases in english: Going beyond manner-place-time. Language Variation and Change **11** (2000) 231–266
14. Wasow, T.: Postverbal behavior. CSLI publications (2002)

15. Rosenbach, A.: Animacy versus weight as determinants of grammatical variation in english. Language **81(3)** (2005) 613–644
16. Beaufort, R., Ruelle, A.: elite : système de synthèse de la parole à orientation linguistique. In: Actes des XXVI journées d'études sur la parole, Dinard (2006)
17. Namer, F.: Acquisition automatique de sens à partir d'opérations morphologiques en français : étude de cas. In: Traitement Automatique de la Langue Naturelle (TALN). (2002)
18. Tran, M., Maurel, D.: Prolexbase : Un dictionnaire relationnel multilingue de noms propres. Traitement automatique des langues **47**(3) (2006) 115–139
19. Bybee, J.: The emergent lexicon. In Gruber, M.C., Higgins, D., Olson, K.S., Wysocki, T., eds.: CLS 34: The panels, University of Chicago, Chicago Linguistic Society (1998)
20. Lapata, M., McDonald, S., Keller, F.: Determinants of adjective-noun plausibility. In: Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL'95). (1999) 30–6
21. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**(6) (1974) 716–723
22. Cutler, A., Cooper, W.E.: Phoneme-monitoring in the context of different phonetic sequences. Journal of Verbal Learning and Verbal Behavior **18(4)** (1978) 497–508
23. Pinker, S., Birdsong, D.: Speakers' sensitivity to rules of frozen word order. Journal of Verbal Learning and Verbal Behavior **18(4)** (1979) 497–508
24. Hawkins, J.: Efficiency and Complexity in Grammars. Oxford University Press (2004)
25. Croft, W.: Radical Construction Grammar. Oxford University Press (2001)
26. Goldberg, A.: Constructions at Work: the nature of generalization in language. Oxford University Press (2006)

Embodied determiners

Simon Pauw and Michael Spranger

Sony Computer Science Laboratory Paris {simon|spranger}@csl.sony.fr

Abstract. In this paper we test the dominant paradigm for modeling the semantics of determined noun phrases called *Generalized Quantifier Theory* in embodied interactions with robots. We contrast the traditional approach with a new approach, called *Clustering Determination*, which is heavily inspired by research on grounding of sensorimotor categories, and we show that our approach performs better in noisy, real world, referential communication.

1 Introduction

This paper focusses on semantic models of the grammatical class of determiners (e.g., *all*, *the*) and their use in determined noun phrases for embodied artificial agents that communicate in and interact with the real world. The dominant paradigm for modeling the semantics of determined noun phrases is *generalized quantifier theory* [Barwise and Cooper, 1981], which is a theory deeply rooted in logical theories of quantification [Mostowski, 1957]. It has been successfully applied to formal semantics, i.e. truth conditional semantics, with substantial influence on linguistics over many years [Westerståhl, 1995, Lappin, 1997]. In applying this theory to natural language, determined noun phrases are modeled as assertions over sets that have some property.

Here, we study the applicability of the theory to situated interaction scenarios, where embodied artificial systems try to reach goals in the real world using language. For instance, in *spatial language games* [Steels, 2001] robots try to draw the attention to an object in the environment using a phrase like *the left block*. Attention is confirmed by the hearer through pointing at the object or giving it to the speaker. Such a model of communication understands language as a tool to reach certain aims, and it has provided many insights into how agents can self-organize sensorimotor categorical systems, for instance, for color, postures and space [Bleys et al., 2009, Spranger and Loetzsch, 2009, Spranger et al., 2010b]. In this line of research, insights from *prototype theory* [Rosch and Lloyd, 1978], which stress graded category membership, are the key to making systems adaptive and flexible enough to succeed in real world communication.

In this paper we present a concrete implementation of Generalized Quantifiers (GQ) that allow agents to take part in situated interactions. In a second approach we extend the insights from sensorimotor categorization to incorporate complex, compositional semantic structure, and propose an implementation of determiners that differs substantially from the Generalized Quantifier approach. We show that our approach called *Clustering determination* performs better and more reliably in real world communicative situations.

2 Embodied interaction

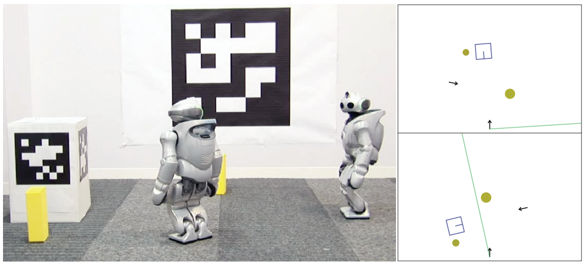


Fig. 1. Left: example scene consisting of various objects, e.g., robots, blocks and boxes. Right: world models constructed by each robot reflecting the current knowledge about the environment.

Figure 1 shows an example scene with two robots interacting in a shared environment. Each robot perceives the world through its own onboard sensors, e.g., the camera and proprioceptive sensors. The vision system [Spranger, 2008] gathers information from the sensors into a world model, that reflects the current belief of a robot about the state of the environment. Starting from this model, the speaker agent will choose a topic, which can be any object or subsets of objects in the environment, collectively referred to as topic hereafter. The goal of the speaker in the scenario described in this papers is to draw the attention of the interlocutor to the topic and make him point to it or to each of the objects that are the topic.

In order to reach their communicative goal, speakers go through a planning process which progressively builds semantic structure that most likely helps to reach the particular goal. Since the communicative goal is to help the hearer identify the topic, semantic structure represents a series of operations that the hearer has to go through in order to single out the objects that are the topic. Figure 2 shows the semantic structure underlying the utterance *the left block*. This particular structure consists of an operation that introduces the current context `get-context`, followed by `apply-class` which applies the object class `block`, the result of which is fed to the operation `apply-spatial-category`, which processes the data using the spatial category `left`, and finally, `apply-selector` which will compute the topic using the selector `unique`[1]. We will see later how these operations are precisely implemented for the two approaches, discussed in this paper.

For verbalizing, semantic structure agents are equipped with a rule based system called FCG [Steels and De Beule, 2006] that maps semantic structure to syntactic structure and back using the linguistic knowledge of an agent. Here, we

[1] Much more can be said about this structure and how it is constructed by agents. The interested reader is referred to Spranger et al. [2010a] for more information.

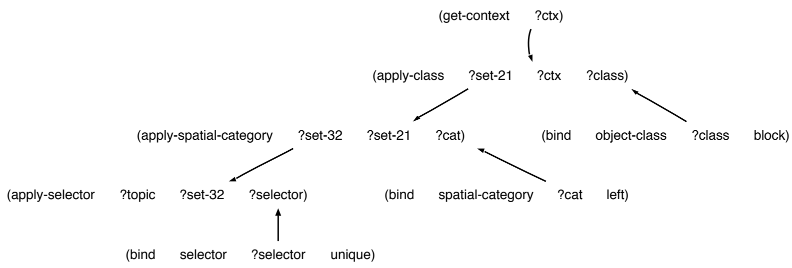


Fig. 2. Semantic structure representing the meaning of the utterance *the left block.* The network contains bind-statements that introduce semantic entities (e.g., the object class `block`), as well as operations that define what to do with these semantic entities. Links in the network are defined by variables (starting with a `?`), e.g., the output of operation `apply-class` is linked to `apply-spatial-category` through the variable `?set-21`.

equipped agents with lexical items for spatial categories (e.g., *left, back, front, right*), object classes (e.g., *block, box, robot, thing*) and selectors (e.g., *the* and *all*, where *the* maps to `unique`). Moreover, rules for determined adjective noun phrases, determined noun phrases, and determined spatial phrases like *all blocks left of the box* are provided [Spranger et al., 2010b].

3 Generalized Quantifiers

We now turn to the actual implementation of operations. Our first approach to modeling determined phrases is based on the notion of generalized quantifiers [Barwise and Cooper, 1981]. There, a noun or verb phrase denotes a property that can be represented as a function from entities to truth values, in other words, as the set of entities for which the property holds. Consequently, the interpretation of *ball* is the set of all balls B in a context, and *are red* is the set of all things that are red R. Determiners then are understood as set relations. For instance, the sentence *all blocks are red* can be modeled as $B \subseteq R$. The determined noun phrase is therefore modeled as a function from a set to truth values, in other words, a generalized quantifier. For example, the determined noun phrase *all blocks* is interpreted as a function $f(P)$ that is true iff $B \subseteq P$.

In accordance with this model, we implement the operations recruited in semantic structure (as seen in Figure 2) based on the notion of set membership. For instance, the operation `apply-class`, when passed an input set and an object class like `block`, will return a set with all elements that belong to that class. The decision on set membership is based on the visual system, which tracks different classes of objects over time using dedicated visual algorithms tracking blocks, robots and boxes separately. In the semantic structure in Figure 2, the output set of the `apply-class` operation is passed to `apply-spatial-category`, which will filter all elements that are in this set *and* are left and thus return all left blocks. Here, all objects that are within a cone of 90˚ to the left of the reference system are considered. Next the operation `apply-selector` is executed. This operation computes all the sets for which the generalized quantifier yields true. In the case of the determiner *the*, it tests if the set of left blocks contains precisely one element and if so, makes that element the referent of the noun phrase. For the determiner *all*, it is checked if the set of left blocks is non-empty and subsequently returns the entire set as referent.[2]

One might be tempted to question how general this particular modelling approach is. For the concrete filtering operations, i.e., those for applying categories of various kind, probably other mechanisms are possible, but the main point from this section is that no matter what the particular implementation is, set membership is at the heart of it, because GQ are defined in terms of sets or properties. However, having a strong set notion of category membership quickly leads to problems in embodied interaction scenarios, because two agents might disagree on the category membership of an entity close to the category boundary due to perceptual noise and different viewpoints on the scene. In the next section we propose a model that tries to deal with such cases by introducing graded category membership and by a mechanism that postpones the decision of determining the referent until all categories have applied.

4 Clustering Based Determination

Our second approach to determined phrases is inspired by ideas from prototype theory about how humans categorize objects [Rosch et al., 2004, Lakoff, 1987, Langacker, 1988]. In contrast to traditional theories about categorization, these theories propose models of graded category membership, with some objects being more central to a category than others. Much work has been done on grounding graded categories, and it was shown that domains such as color [Berlin and Kay, 1993, Bleys et al., 2009], but also actions [Spranger and Loetzsch, 2009] are best understood in terms of prototypicality effects. However, these approaches do not study determined phrases. Here, we propose a concrete implementation of graded category membership that is extended by determination.

We represent graded categories using a scoring mechanism. In terms of the semantic structure in Figure 2, an operation that applies a category to a set of objects, such as the operation `apply-class`, will score the objects in the set as to how similar they are with respect to the class or category in question (with a score of 0 denoting no similarity and 1 meaning most similar). The objects are categorized by copying them to the output set and by multiplying their original score with the score reflecting the currently processed category membership. Hence, in contrast to the generalized quantifier approach objects are not filtered but copied and scored. Let us, for instance, look at what happens, when the category `left` is applied, which in the semantic structure of Figure 2 follows the application of the object class `block`. The input set to the operation `apply-spatial-category`, thus, consists of all objects in the context, with a

[2] For efficiency reasons we only compute those properties that are a subset of the noun, which can be safely done for conservative quantifiers.

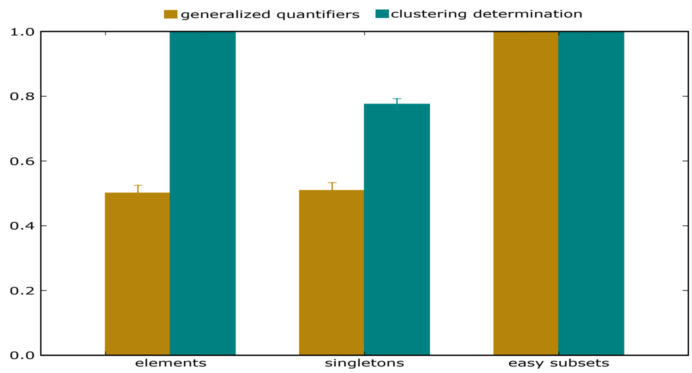


Fig. 3. Average communicative success (and variance) for 8 times 2000 interactions on spatial scenes such as in Figure 1. Three experimental conditions were tested to compare the generalized quantifiers (GQ) and clustering determination approach.

score reflecting their membership to the object class `block`. The category `left` is implemented as a direction vector pointing to the left, and similarity to the category is computed via the angle distance between object and category (using an exponential decay enveloped angle distance). Hence, if the object is lying to the right it will get a score close to 0, and a score of 1, if it lies directly to the left of the coordinate system. Since all objects are already scored by their similarity with `block`, that score with is multiplied with the score for the category `left`.

In order to select entities from this set using a determiner, the particular communicative goal of an agent is considered. For instance, the determiner *the* (selector `unique`) which is used to pick out one particular object from the context, is implemented as picking the object with the highest score from the input set to the operation `apply-selector`. Hence, the referent of a phrase like *the right yellow block* is the object which is most similar to the categories right, yellow and block. The *all* determiner on the other hand is implemented as a clustering algorithm, which segments data into two classes: the objects belonging to the category combination denoted by such an utterance and the ones not. We used a k-means [Lloyd, 1982] algorithm with $k = 2$, which segments the input data set based on the scores of objects. As a result, the input set is divided into two possible classes, one of which has a higher score centroid and thus is the set denoted by the phrase. `apply-selector` returns the set with the highest score centroid. Notice that k-means can fail when, for instance, when all objects have the same score. In that case we consider the whole input set.

5 Results

Due to space constraints we were only able to sketch the implementation of particular prototypes, object classes and selectors for the two approaches discussed in this paper. But, it must be stressed that the same principles apply to other object classes, e.g., `box`, `robot`, `thing` and spatial categories e.g., `right`, `front`, `back`. Equipped with the machinery for parsing and producing utterances, agents play many language games in different spatial setups (Figure 1 is an example of roughly 250 spatial scenes). Each interaction is either a success, if the hearer correctly points to all topic objects, or a failure, otherwise.

Figure 4 shows how the two approaches presented in this paper perform. We compare the communicative success of GQ with clustering based determination in three experimental conditions: *elements*, *singletons* and *easy subsets*. The graphs show that for the easy subsets condition both approaches perform well and reach success in all interactions, which immediately follows from the fact that easy subsets only include *all blocks*, *all boxes* or *all robots*. The vision system tracks these classes reliably without any errors, which implies that generalized quantifiers work well, when the knowledge about the state of the world is extremely accurate and precise.

The two interesting conditions are the elements and singletons conditions, which test the performance of *the* and *all* determination on single blocks or a set containing a single block respectively. Since in both approaches agents have to talk about blocks they are bound to use spatial categories to discriminate between the two or more blocks present in every scene. The difference between the two conditions is that in in the elements condition they are forced to use the *the* determiner, whereas in the singletons case *all* is used. In both cases a clear advantage for clustering determination is apparent, with the elements condition showing this most salient (clustering determination: 100% success, GQ: around 50% success). In the singletons condition k-means clustering determination is directly compared with generalized quantifiers, which achieve success in roughly 50% of the interactions, compared to 80% success of k-means.

It is interesting to note, that while there is no difference in performance for generalized quantifiers across the two interesting conditions there is quite a substantial difference in performance for the clustering determination approach. This can be explained with additional information available to the hearer when confronted with a *the* determined phrase. In such cases, the hearer only has to find the best matching element in the context, that is, the object which is most similar with respect to the semantic structure of the utterance. This information is missing in the case of *all* determination. How many objects are potentially in the topic set, needs to be explicitly recovered by interpreting agents. On the other hand, this additional information makes no difference in the case of GQ, where the set denoted by the utterance is strictly determined by the categories and, hence, there is no dynamic segmentation of the context based on similarity.

6 Discussion and Conclusion

The experimental results clearly show that the clustering determination model performs better than generalized quantifiers in embodied situated interactions. But, one could wonder if our implementation of generalized quantifiers is general enough to draw any strong conclusions. Could we improve the results by choosing other categorization mechanisms? The general problem with generalized quantifiers is that we have to compute strict sets. Even if we look for other means of categorization, the problem remains that the determiner itself can, in some sense, not influence the manner in which sets are computed. The only task of the determiner is to reason over sets not construct sets to reason over. This is conceptually different in the second model, where it is essentially upon the determiner to construct sets. It is this design choice that in a nutshel is the difference between the two approaches.

Another point of concern is the generality of the clustering model. It should be clear at this point that for situated interactions we need a mechanism that can deal with vagueness, but there are other proposal for this such as *fuzzy set theory* [Zadeh, 1965] and *supervaluation theory* [Fine, 1975]. Why didn't we base ourselves on these? These theories provide a logical framework for graded category membership but do not depend on any specific procedure for computing the vague category borders. Here we provide a model that could well be integrated in either theory. To cast our model into either framework might be an interesting point of further research, but is not of interest for the presented experiment.

In the context of language games, semantics is not about truth but about communicative goals (e.g., performing actions, pointing out an object or event, sharing intention) [Steels, 2001]. If an agent chooses to use a specific word or grammar construction, he does so because it contributes to the communicative goal. Generalized quantifiers, on the other hand are concerned with establishing truth. This discrepancy is reflected in the experimental results. With the generalized quantifier model, the introduction of the determiner *the*, does not give rise to a higher communicative success. This is due to the fact that *the* is a special case of *all*, so it does not provide the agents with additional information that helps discriminating the referent.

We have demonstrated why we need a model for determined noun phrases that diverges from the standard generalized quantifier interpretation. Generalized quantifiers rely on strict category membership which, as we have shown by means of an experiment, performs poorly for a world model that is based on real world perception. Using prototype theory and standard clustering methods we propose an alternative model and were able to show that, indeed, this model is more robust in noisy real world communicative interactions.

Bibliography

J. Barwise and R. Cooper. Generalized quantifiers and natural language. *Linguistics and philosophy*, 4(2):159–219, 1981.

B. Berlin and P.. Kay. *Basic color terms; their universality and evolution*. London, Van Nostrand Reinhold [c1973], 1993.

Joris Bleys, Martin Loetzsch, Michael Spranger, and Luc Steels. The grounded colour naming game. To appear in Proceedings Roman-09, 2009.

K. Fine. Vagueness, truth and logic. *Synthese*, 30(3):265–300, 1975.

G. Lakoff. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press Chicago, 1987.

R.W. Langacker. Foundations of cognitive grammar, vols. 1 and 2. *Stanford: Stanford University Press*, 1991:2, 1988.

S. Lappin. The handbook of contemporary semantic theory. 1997.

S.P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

R. Montague. The proper treatment of quantification in ordinary English. *Formal Semantics: The Essential Readings*, 1974.

A. Mostowski. On a generalization of quantifiers. *Fundamenta mathematicae*, 44(2), 1957.

E. Rosch and B.B. Lloyd. *Cognition and categorization*. John Wiley & Sons Inc, 1978.

E. Rosch, C.B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology: Key Readings*, page 448, 2004.

M. Spranger, M. Loetzsch, and S. Pauw. Open-ended grounded semantics. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI-2010)*, 2010a.

Michael Spranger. World models for grounded language games. German diplom (master) thesis, Humboldt-Universitaet zu Berlin, March 2008.

Michael Spranger and Martin Loetzsch. The semantics of SIT, STAND, and LIE embodied in robots. In Niels A. Taatgen and Hedderik van Rijn, editors, *Proceedings of the 31th Annual Conference of the Cognitive Science Society (Cogsci09)*, pages 2546–2552, Austin, TX, 2009. Cognitive Science Society.

Michael Spranger, Simon Pauw, and Martin Loetzsch. Open-ended semantics co-evolving with spatial language. In *Proceedings of EVOLANG 8*, 2010b.

Luc Steels. Language games for autonomous robots. *IEEE Intelligent Systems*, 16(5):16–22, October 2001.

Luc Steels and Joachim De Beule. Unify and merge in Fluid Construction Grammar. In *Symbol Grounding and Beyond: Proceedings of EELC 2006*, LNAI 4211, pages 197–223, Berlin, 2006. Springer.

D. Westerståhl. Quantifiers in natural language: A survey of some recent work. *SYNTHESE LIBRARY*, pages 359–359, 1995.

LA Zadeh. Fuzzy sets*. *Information and control*, 8(3):338–353, 1965.

Phonetic Alignment Based on Sound Classes

A New Method for Sequence Comparison in Historical Linguistics

Johann-Mattis List (July 23, 2010)

Heinrich Heine University Düsseldorf

Abstract. In this paper, I present a new method for the automatic implementation of pairwise and multiple alignment analyses in historical linguistics which is based on sound classes and implemented as a Python library. While sound classes are usually employed in historical linguistics as a stochastic device for detecting possible sound correspondences among languages and the proof of genetic relationship among languages, it shall be shown that they are equally well apt for phonetic alignment tasks. Moreover, they have two further advantages: Firstly, due to the fact that sound classes constitute a rather small alphabet, they are perfectly apt for subsequent use in biological software tools for sequence alignment, which makes it possible to carry out quick pairwise and multiple alignment analyses. Secondly, since sound classes can be based on explicit historical considerations regarding phonetic similarity, the alignments are capable of yielding certain outputs which cannot be retrieved by applying similarity metrics which are solely based on synchronic phonetic resemblences.

1 Sequence Comparison and Alignment Analyses in Historical Linguistics

Among the different aspects of comparative-historical linguistics, sequence comparison plays a crucial role. It constitutes the basis of the comparative method which seeks to detect regular sound correspondences in lexical material of different languages in order to prove their genetic relationship and to uncover the unattested ancestor language by means of linguistic reconstruction [1]. Since sequences – in contrast to sets – consist of non-unique elements which retrieve their distinctive function only because of their order, sequence comparison is always based on phonetic alignment, i.e. the corresponding phonetic segments of two or more sequences are ordered in such a way that they are set against each other.

In the following, I shall present a new method for phonetic alignment, which is not only easy to implement and to modify but also explicitly historically oriented. The paper is structured as follows: After giving a short introduction into the basic algorithms which are usually employed when carrying out pairwise and multiple sequence alignments, I shall describe the method by presenting the basic idea behind the sound classes employed and their implementation in the Python library. In a further step, I shall discuss the performance of the method in contrast to an alternative proposal by G. Kondrak [2].

2 Basic Procedures for Alignment Analyses

2.1 Pairwise Alignments

Almost all procedures for alignment analyses in historical linguistics which have been proposed so far are based on the Dynamic Programing Algorithm (DPA) which was independently proposed by different scholars as a way to carry out sequence alignments in such different disciplines as biology, linguistics, and gas chromatography [3][1].

The basic idea of the DPA is to create a matrix which confronts all segments of the sequences under comparison either with each other or with alternative null-sequences (fills). In a further step, the algorithm seeks the path through the matrix which is of the lowest general cost. The general cost is cumulatively calculated by means of a specific scoring function that penalizes the matching of segments with each other and the insertion or deletion of segments in the two sequences. Figure 1 illustrates this process for the alignment of Engl. "heart" vs. Germ. "herz" based on the Levenshtein scoring function which penalizes both fills and mismatches with 1. The left matrix shows the shortest alignment path chosen by the algorithm, the right matrix reflects the cumulatively calculated costs for each cell (for a more detailed description of the algorithm, cf. [6]).

Fig. 1. DPA-Matrix for Engl. "heart" vs. Germ. "herz"

-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	h	-	e	-	a	-	r	-	t
h	-	h	-	h	-	h	-	h	-	h	-
-	-	-	h	-	e	-	a	-	r	-	t
e	-	e	-	e	-	e	-	e	-	e	-
-	-	-	h	-	e	-	a	-	r	-	t
r	-	r	-	r	-	r	-	r	-	r	-
-	-	-	h	-	e	-	a	-	r	-	t
z	-	z	-	z	-	z	-	z	-	z	-
-	-	-	h	-	e	-	a	-	r	-	t

0	1	2	3	4	5
1	0	1	2	3	4
2	1	0	1	2	3
3	2	1	1	1	2
4	3	2	2	2	2

Regarding alignment analyses in historical linguistics, three aspects of the DPA are of crucial importance: (1) the basic part of the algorithm which iterates over the matrix, (2) the scoring function for segment-to-segment comparison, and (3) the pre-segmentation of the linguistic units, i.e. the transcription of the phonetic characters, carried out when compiling the datasets for sequence comparison. All these aspects have a crucial impact on the quality of the alignments created by the procedure.

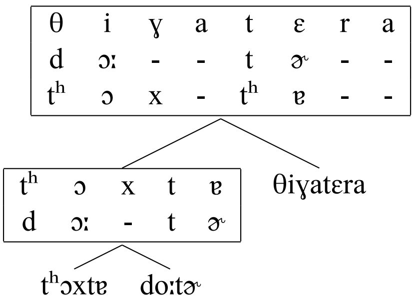
While earlier approaches mainly concentrated on the scoring function [7][2] or additional edit operations [2] the method proposed here divides sounds into specific classes whose members show a high probability of interchange during language evolution.

[1] As examples for different proposals regarding the DPA, cf. e.g. [4] and [5]

2.2 Multiple Alignments

While pairwise alignment analyses can be carried out without problems using the above-mentioned dynamic programming algorithm or certain of its extensions [8][9][10], multiple sequence alignments (MSA) have to make use of certain heuristics which do not guarantee that the optimal alignment for a set of sequences has been found, since the computational effort increases enormously with the number of sequences being analysed [6, 345]. The most common heuristics which is applied in computational biology are the so-called progressive algorithms which are based on a guide-tree that is reconstructed from the pairwise alignment scores of all sequences and along which the sequences are stepwise added to the multiple alignment [11, 143f] (see Figure 2).

Fig. 2. MSA Based on a Guide-Tree



While the original approach by Feng & Doolittle[12] for progressive MSA compares sequences only pairwise, thus taking a pair of sequences as representative for a whole multiple alignment, profile-based approaches allow for a more refined approach to align multiple sequence alignments to each other [11, 146f]. A profile consists of the relative frequency of all segments of a multiple alignment in all its positions [6, 337] (see Figure 3). Thus, a profile represents a multiple alignment as a sequence of vectors. Aligning profiles to profiles instead of aligning two represantative sequences of two given multiple alignments usually yields better results in MSA, since more information can be taken into account which would otherwise be ignored.

3 Sound Classes in Historical Linguistics

The main idea behind sound classes in historical linguistics is the assumption that it is possible "to divide sounds into such groups, that changes within the boundary of the groups are more probable than transitions from one group into another" [13, 272][2]. Thus, when comparing the dental consonants t, d, tʰ, θ with the velars k, g, kʰ, ɣ one can assume that

[2] My translation, original text: «[...] выделить такие группы звуков, что изменения в пределах группы более вероятны, чем переводы из одной группы в другую.»

Fig. 3. MSA and Profiles

	Multiple Alignment: Traditional Format						
	ʧ	-	l	o	vʲ	ɛ	k
	ʧ	-	-	o	v	ɛ	k
	ʧʲ	ɪ	l	ɐ	vʲ	ɛ	k
	ʧ	-	w	ɔ	vʲ	ɛ	k
	Multiple Alignment: Profile Representation						
ʧ	.75						
ʧʲ	.25						
l			.50				
o				.50			
v					.25		
vʲ					.75		
ɐ				.25			
ɛ						1.0	
ɪ		.25					
k							1.0
w			.25				
ɔ				.25			
-		.75	.25				

it is more probable that any of the dentals may change to a dental than to a velar sound, and vice-versa. This does, of course, not mean that a sound change from one class into another is impossible, yet most linguists would certainly agree that such a sound change would be rather unexpected and strange. Starting from this general assumption, A. B. Dolgopolsky [14] was the first to carry out empirical studies of the most typical sound changes in a large sample of languages. He proposed ten fundamental sound classes, which are given in Table 1.

Table 1. Dolgopolsky's Sound Classes

No.	Class	Description	Example
1	P	labial obstruents	p,b,f
2	T	dental obstruents	d,t,θ,ð
3	S	sibilants	s,z,ʃ,ʒ
4	K	velar obstruents, dental and alveolar affricates	k,g,ʦ,ʧ
5	M	labial nasal	m
6	N	remaining nasals	n,ɲ,ŋ
7	R	liquids	r,l
8	W	voiced labial fricative and initial rounded vowels	v,u
9	J	palatal approximant	j
10	ø	laryngeals and initial velar nasal	h,ɦ,ŋ

Sound classes have been employed in a couple of recent studies which largely deal with stochastical aspects of the prove of genetic relationship among languages [15] [16] or as a heuristic for the automatical implementation of cognate judgments [17] [3]. In contrary to the approach presented here, these studies are not based on sequence alignment, but rather check whether the first or the first two consonants of basic words match regarding their respective sound classes.

4 The Python Library for Sound-Class-Based Alignment

4.1 General Working Procedure

The method for sound-class-based alignment has been implemented as a Python library and can be invoked from the Python prompt or within Python scripts[4]. The core function of the library, the alignment function, executes the following operations: After tokenizing the input sequences (which should be in IPA-transcription), it first converts the input sequences into strings of capitals which represent the 11 sound classes employed by the method. These strings are then passed to a function that carries out an alignment analysis of the class-strings. The aligned strings are then converted back to their original IPA-transcription (see Figure 4).

The sound classes employed in the library are mainly based on the suggestions of Dolgopolsky [14], but they are extended to cover the full range of IPA, including the most common diacritics, and vowels (simple vowels and diphtongs), which are ignored in Dolgopolsky's original system, are included as an eleventh class of sounds.

4.2 Pairwise Alignments

Pairwise alignments are implemented by the pairwise2-module of the BioPython library [19], which allows one to carry out both local and global alignment analyses. While global alignment analyses carry out alignments for two entire strings, local alignment analyses, which are based on an extension of the DPA (the Smith-Waterman algorithm [8]), seek the two substrings which show the highest similarity and eventually leave prefixes and postfixes unaligned [11, 22-24].

In order to enhance the alignment analysis, a special matching dictionary has been prepared as an input for the scoring function (see Table 2). Note that the scoring function of pairwise2 is based on similarity of segments as apposed to distance. Segments which should be matched by the algorithm are therefore given higher scores than segments whose matching should be avoided.

[3] As a matter of fact, nearly all alignment analyses in historical linguistics, such as the ones carried out by the ASJP project [18] or the approach proposed by G. Kondrak [2], are based on "sound classes" in a broader sense, since they usually abstract from a strict phonetic notion to a broader phonemic one. Yet, appart from the algorithm proposed by Covington [7], none of these approaches makes use of historical knowledge regarding the probability of sound change processes when carrying out the similarity judgments.

[4] A preliminary version of the modules, including two testsets, is online available under http://www-public.rz.uni-duesseldorf.de/ jorom002/sca.zip.

Fig. 4. The Alignment Analysis of the Sound-Class-Approach

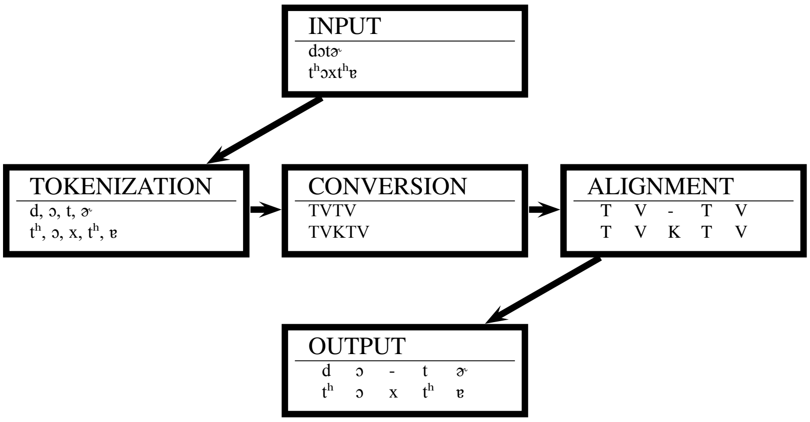


Table 2. Matching Dictionary for the Scoring Function of BioPython.pairwise2

Score	**Condition**	**Example**
5	Consonant-Class-Identity	K + K
4	Vowel-Class-Identity	V + V
-10	Vowel-Non-Vowel	V + K
-4	Non-Identity of Consonants	K + M
2	Specific Combinations	K + S, K + T, T + S
-1	Gaps	- + K

4.3 Multiple Alignments

In the current implementation of the library, both traditional MSA, based on the Feng-Doolittle-algorithm [12], as well as profile-based MSA roughly based on the CLUSTALW-implementation [20] for MSA in evolutionary biology is possible. The code for MSA analyses has been written by the author. The calculation of the guide-tree, which can be carried out either by the UPGMA clustering algorithm [21] or the Neighbor-joining method [22], makes use of the cogent.phylo-module of the PyCogent library [23]. The scoring function for profile-sequence and profile-profile alignments is based on the *sum of pairs score*, a standard way to score multiple alignments in evolutionary biology [11, 139f], which consists of the mean of the sums of all pairwise segment scores of two MSAs.

5 Performance of the Method

5.1 Pairwise Alignments

The method was tested (using local alignment) on a testset of 82 cognate pairs proposed by M. A. Covington [7], which was slightly modified in order to be appropriate for the input

Table 3. Comparison of the Sound-Class-Approach with ALINE's Alignments

No.	Pair	Sound-Class-Approach		ALINE
1	Engl. daughter / Old Grk. θυγατήρ “daughter”	‖d o - - t ə r‖ ‖tʰ u g a t e: r‖	‖d - - o t ə r‖ ‖tʰ u g a t e: r‖	‖d o t ə r‖ tʰu‖g a t e: r‖
2	Spa. decir / Fre. dire “say”	de‖θ i r‖ ‖d i r‖		‖d e θ i r‖ ‖d - - i r‖
3	Engl. this / Grm. dieses “this”	‖ð i s‖ ‖d i: z‖əs		‖ð i z‖ di:‖z ə s‖
4	Fox kiinwaawa / Menomini kenuaq “you (Pl.)”	‖k i: n w a: w a‖ ‖k e n - u - a‖ʔ		‖k i: n w a:‖wa ‖k e n u a‖ ʔ
5	Old Grk. δίδωμι / Lat. do “I give”	di‖d o:‖mi ‖d o:‖	‖d i‖do:mi ‖d o:‖	di‖do: mi‖ ‖do:‖
6	Engl. tooth / Lat. dentis “tooth”	‖t u - θ‖ ‖d e n t‖is		‖t u θ‖ den‖t i s‖
7	Engl. I / Lat. ego “I”	‖ai‖ ‖e‖go	‖ai‖ eg‖o‖	‖ai‖ ‖e‖go
8	Engl. eye / Grm. Auge “eye”	‖ai‖ ‖au‖gə	‖ai‖ aug‖ə‖	‖a i‖ ‖a u‖ge
9	Spa. todos / Fre. tous “all”	‖t o‖dos ‖t u‖	to‖d o‖s ‖t u‖	‖t o‖dos ‖t u‖
10	Engl. one / Lat. unus “one”	w‖ə n‖ ‖u: n‖us		‖wə n‖ ‖u: n‖us
11	Engl. round / Lat. rotundus “round”	‖r - - au n d‖ ‖r o t u n d‖us	‖r au - - n d‖ ‖r o t u n d‖us	‖r a - u n d‖ ‖r o t u n d‖us

format required for the sound class approach[5]. The results of the alignment analyses were compared to the ALINE algorithm of G. Kondrak [2] which shows the best performance of recently proposed alignment algorithms for linguistic purposes. Comparing the output of the sound-class approach to ALINE's alignments for the Covington testset, there are 71 cases, where both methods yield exactly the same results. Of the 11 ones which are aligned differently (see Table 3) there are six cases where the sound-class approach gives

[5] The non-IPA-characters of Covington's testset were converted to IPA-symbols and the halfvowels of the diphtongs, which were originally coded as glides were converted to the respective full vowels.

two equivalent outputs. In four out of these six cases, one of these double-outputs matches with ALINE's single-output (Nos. 5, 7, 8, and 9 in Table 3), in one case both outputs produced by the sound-class-approach are superior to ALINE's output (No. 1) and in the last one (No. 11), it cannot be decided, which of the outputs given by either of the approaches is better. In the remaining 5 cases of different output, there are three cases where ALINE performs better (Nos. 2, 4, and 10) and two cases where the sound-class-approach gives the better alignment (Nos. 3, and 6).

It becomes obvious that, apart from similar results in the majority of the cases, there are a couple of significant differences between the alignments of the sound-class approach and ALINE's alignments. Firstly, there are cases, where the sound-class approach yields multiple outputs while ALINE has a single one. These results are mostly due to the information loss accompanying the conversion of IPA-strings into sound-class-strings. This, however, does not constitute a general problem for the method, since the double-outputs occur mostly in cases which are problematic for alignments in general: No historical linguist would dare to align words such as Engl. 'I' and Lat. 'ego', lacking the relevant facts from other related languages. Secondly, there are cases which show a particular benefit of the sound-class approach: Since this approach is not based on a synchronic idea of phonetic similarity, but on a 'historical' notion of phonetic similarity, it yields convincing outputs in such challenging cases as Engl. 'daughter' vs. Old Grk. 'thugatēr', where it matches all consonants correctly, while ALINE opposes *d* and *g*. Table 4 gives some representative examples for cognate pairs (transcribed with more phonetic detail) where the alignments of the sound-class-approach are superior to those of ALINE.

Table 4. Additional Examples for Different Alignments of the Sound-Class-Approach and ALINE

	Sound-Class-Approach			**ALINE**		
1 Old Grk. καρδιά / Skr. hṛt "heart"						
		k a r d	ia	ka	r d	ia
		h - r t		h	r d	
2 Grm. Herz / Lat. cor "heart"						
		h ɛɐ	ʦʰ	hɛɐ	ʦʰ	
		k o	r		k	or
3 Mod. Grk. νέος / Rus. новый "new"						
		n e - o s			n e o s	
		n o v ɨ j			n - o v	ɨj
4 Mod. Grk. καρδιά/ Grm. Herz "heart"						
		k a r ð	ia	kar	ð i a	
		h ɛɐ - ʦʰ			h ɛ ɐ	ʦʰ

5.2 Multiple Alignments

Apart from the fact that the sound-class-approach is easy to implement and to modify, while yielding satisfying results comparable to that of more refined algorithms for sequence comparison, a major advantage of the approach lies in its flexibility to be adapted for more complex approaches. Thus, the implementation of more complex algorithms is far less complicated, since many functions available in biological software modules for python can be easily included. This makes it even possible to carry out multiple alignment analyses which are rarely implemented in the current algorithms for sequence comparison in historical linguistics, the only exception known to the author being a proposal by Prokič et al. (2009) [24].

Since MSA has only recently been added to the library, no full-size tests runs can be reported at the moment, yet the first tests on small samples of dialectal data and cognate sets of Indo-European languages yield quite promising results. Furthermore, it can be easily demonstrated that the profile-based approach yields better results than other progressive approaches, as the comparison on profile-based MSAs and MSAs based on the traditional Feng-Doolittle algorithm in Table 5 for Indo-European and Slavic cognate pairs demonstrates, where Old Church Slavonic дъщи "daughter" and Polish człowiek "human" are incorrectly aligned in the non-profile-based approach.

Table 5. Multiple Alignments Yielded by the Sound-Class Approach

No.	**Traditional MSA based on the Feng-Doolittle algorithm**						
1	Old Grk. θυγατήρ / Grm. Tochter / Engl. daughter / OCS дъщи / Skr. duhitār "daughter"						
	t^h	u	g	a	t	eː	r
	t	o	x	-	t	ə	r
	d	o	-	-	t	ə	r
	d	u	-	ʃ	t	i	-
	d	u	h	i	t	aː	r
2	Czech člověk / Bulgarian човек / Russian человек / Polish człowiek "human"						
	ʧ	-	l	o	v^j	ɛ	k
	ʧ	-	-	o	v	ɛ	k
	$ʧ^j$	ɪ	l	ɐ	v^j	ɛ	k
	ʧ	w	-	ɔ	v^j	ɛ	k
No.	**Profile-based MSA**						
1	Old Grk. θυγατήρ / Grm. Tochter / Engl. daughter / OCS дъщи / Skr. duhitār "daughter"						
	t^h	u	g	a	t	eː	r
	t	o	x	-	t	ə	r
	d	o	-	-	t	ə	r
	d	u	ʃ	-	t	i	-
	d	u	h	i	t	aː	r
2	Czech člověk / Bulgarian човек / Russian человек / Polish człowiek "human"						
	ʧ	-	l	o	v^j	ɛ	k
	ʧ	-	-	o	v	ɛ	k
	$ʧ^j$	ɪ	l	ɐ	v^j	ɛ	k
	ʧ	-	w	ɔ	v^j	ɛ	k

6 Conclusion

In this paper, I presented a new approach for pairwise and multiple sequence alignments in historical linguistics. Although the method is quite simple regarding its basic assumptions and its implementation as a Python library, the performance of the approach is not only comparable to that of previously proposed ones, but it even shows a better performance in very challenging alignment tasks, the reason being its explicit historical orientation regarding phonetic similarity.

References

1. Lass, R.: Historical linguistics and language change. Cambridge University Press, Cambridge (1997)
2. Kondrak, G.: Algorithms for language reconstruction. PhD thesis, University of Toronto, Toronto (2002)
3. Kruskal, J.B.: An overview of sequence comparison: Time warps, string edits, and macromolecules. SIAM Review **25**(2) (April 1983) 201–237
4. Wagner, R.A., Fischer, M.J.: The string-to-string correction problem. J. ACM **21**(1) (1974) 168–173
5. Needleman, S.B., Wunsch, C.D.: A gene method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology **48** (July 1970) 443–453
6. Gusfield, D.: Algorithms on Strings, Trees and Sequences. Cambridge University Press, Cambridge (1997)
7. Covington, M.A.: An algorithm to align words for historical comparison. Computational Linguistics **22**(4) (1996) 481–496
8. Smith, T.F., Waterman, M.S.: Identification of common molecular subsequences. Journal of Molecular Biology **1** (1981) 195–197
9. Gotoh, O.: An improved algorithm for matching biological sequences. Journal of Molecular Biology **162**(3) (1982) 705 – 708
10. Oommen, B.J.: String alignment with substitution, insertion, deletion, squashing, and expansion operations. Inf. Sci. Inf. Comput. Sci. **83**(1-2) (1995) 89–107
11. Durbin, R.: Biological sequence analysis. Probabilistic models of proteins and nucleic acids. 7th print edn. Cambridge University Press, Cambridge (2002)
12. Feng, D.F., Doolittle, R.F.: Progressive sequence alignment as a prerequisite to correct phylogenetic trees. Journal of Molecular Evolution **25**(4) (1987) 351–360
13. Burlak, S.A., Starostin, S.A.: Sravnitel'no-istoričeskoe jazykoznanie (Comparative-historical linguistics). Akademia, Moskva (2005)
14. Dolgopolsky, A.B.: A probabilistic hypothesis concerning the oldest relationships among the language families of northern Eurasia. In Shevoroshkin, Vitaly V.; Markey, T.L., ed.: Typology Relationship and Time. Notes on Linguistics. Karoma Publisher, Inc. (1986) 27–50 Originally published in Russian as "Gipoteza drevnejščego rodstva jazykov Severnoj Evrazii (problemy fonetičeskich sootvetstvij)" in 1964.
15. Baxter, W.H., Manaster Ramer, A.: Beyond lumping and splitting: Probabilistic issues in historical linguistics. In Renfrew, C., McMahon, A., Trask, L., eds.: Time depth in historical linguistics. Papers in the prehistory of languages. The McDonald Institute for Archaeological Research, Cambridge (2000) 167–188
16. Mortarino, C.: An improved statistical test for historical linguistics. Statistical Methods and Applications **18**(2) (2009) 193–204

17. Turchin, P., Peiros, I., Gell-Mann, M.: Analyzing genetic connections between languages by matching consonant classes. Journal of Language Relationship **1** (2010) 117–126
18. Brown, C.H., Holman, E.W., Wichmann, S., Velupillai, V.: Automated classification of the world's languages: a description of the method and preliminary results. STUF, Berlin **61**(4) (2008) 285–308
19. Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J.: Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics **25**(11) (Jun 2009) 1422–1423
20. Thompson, J.D., Higgins, D.G., Gibson, T.J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22**(22) (November 1994) 4673–4680
21. Sokal, R.R., Michener, C.D.: A statistical method for evaluating systematic relationships. University of Kansas Scientific Bulletin **28** (1958) 1409–1438
22. Saitou, N., Nei, M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4**(4) (1987) 406–425
23. Knight, R., Maxwell, P., Birmingham, A., Carnes, J., Caporaso, J.G., Easton, B., Eaton, M., Hamady, M., Lindsay, H., Liu, Z., Lozupone, C., McDonald, D., Robeson, M., Sammut, R., Smit, S., Wakefield, M., Widmann, J., Wikman, S., Wilson, S., Ying, H., Huttley, G.: Pycogent: a toolkit for making sense from sequence. Genome Biology **8**(8) (2007) R171
24. Prokić, J., Wieling, M., Nerbonne, J.: Multiple sequence alignments in linguistics. In: LaTeCH-SHELT&R '09: Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education (LaTeCH - SHELT&R 2009), Morristown, NJ, USA, Association for Computational Linguistics (2009) 18–25

Syntax-based Discourse Segmentation of Dutch Text

Nynke van der Vliet
n.h.van.der.vliet@rug.nl

University of Groningen

Abstract. An essential step in discourse parsing is the identification of suitable Elementary Discourse Units (EDUs). For English and German written text several automatic discourse segmentation approaches have been developed. For Dutch, far less research is available. This paper presents an approach to automatically decompose Dutch written text into EDUs, based on syntactic information. We describe discourse segmentation guidelines for Dutch, the discourse segmentation algorithm, and the results of applying this algorithm to an annotated corpus.

1 Introduction

Discourse relations or coherence relations are used to analyze text organization in terms of relations between text parts. These text parts can be propositions about states of affairs or speech acts. The discourse relations in a text together build up a hierarchical, connected structure of a text, in which every part of the text has a role or function to play with respect to other parts of the text. The smallest units are called Elementary Discourse Units (EDUs).

Large corpora annotated with discourse relations have been developed mainly for English (Carlson and Marcu (2001)) and German (Stede (2004)). The aim of the project Modelling Textual Organisation[1] is to develop a corpus of Dutch texts, annotated with discourse relations and lexical cohesion. For the discourse annotation we use Rhetorical Structure Theory (RST) (Mann and Thompson (1988)). Part of the project is to explore the possibilities for discourse parsing of Dutch text.

An essential step in discourse parsing is the identification of suitable EDUs. Various definitions of EDUs exist, ranging from a fine-grained segmentation to segmentation at sentence level. In classical RST, clauses are considered to be EDUs, except for clausal subjects and complements as well as restrictive relative clauses (Mann and Thompson (1988)). For the annotation of the RST discourse tree bank, Carlson and Marcu (2001) use a fine-grained segmentation in which they also treat complements of attribution verbs and phrases that begin with a strong discourse marker (e.g. *because of, in spite of, according to*) as separate EDUs. Relative clauses, nominal postmodifiers, or clauses that break up other

[1] http://www.let.rug.nl/mto/

legitimate EDUs are treated as embedded discourse units. Based on this, Lüngen et al. (2006) developed segmentation guidelines for German text, but in contrast to Carlson and Marcu (2001) exclude restrictive relative clauses, conditional clauses and proportional clauses (clauses combined by comparative connectives). Grabski and Stede (2006) suggest to also include prepositional phrases as EDUs. Tofiloski et al. (2009) adhere more closely to the original RST proposals (Mann and Thompson (1988)) and segment coordinated clauses, adjunct clauses and non-restrictive relative clauses (marked by a comma).

For Dutch, as far as we know, such an elaborate investigation of what counts as an EDU has not yet been done. RST annotations of Dutch text have used the segmentation of the original RST proposals (Mann and Thompson (1988)) (Abelen et al. (1993)) or taken consecutive clauses (Den Ouden et al. (1998)) or whole sentences (Timmerman (2007)) as EDUs.

We present discourse segmentation principles for automatic segmentation of Dutch text. We have implemented the segmentation principles in a rule-based discourse segmenter that uses syntactic information to insert segment boundaries. For the segmentation of English and German text, both rule-based (Lüngen et al. (2006), Tofiloski et al. (2009), Corston-Oliver (1998), Le Thanh et al. (2004)) and machine learning (Soricut and Marcu (2003), Sporleder and Lapata (2005), Subba and Di Eugenio (2007)) approaches to automatic discourse segmentation have been shown to be successful. For Dutch, there is insufficient annotated data to apply machine learning techniques. However, Tofiloski et al. (2009) show for their English data that a rule-based approach works as well as or better than machine learning.

2 Segmentation Principles

The definition of an elementary discourse unit is guided by the question whether a discourse relation could hold between the unit and another segment. Our segmentation is fairly coarse, separating independent and subordinate clauses as well as other complete utterances (fragments) as elementary units. Like Tofiloski et al. (2009), we treat clauses (1) and sentences (2), coordinated clauses (1 and 3) and non-restrictive relative clauses (separated by a comma) (4) as EDUs.

(1) [Je kunt meebouwen met een gift,][maar je kunt ook letterlijk meebouwen.]
[You can build with us using a donation,][but you can also build with us literally.]

(2) [Elke donatie is waardevol!]
[Each donation is valuable!]

(3) [Cavine kreeg aidsremmers][en dat maakte een levensgroot verschil.]
[Cavine got aids medication][and that made a huge difference.]

(4) [Dit gat wordt veroorzaakt door een van de maantjes van Saturnus, Mimas,][die de ringen verstoort.]
[This gap is caused by one of the moons of Saturnus, Mimas,][that disturbs the rings.]

Embedded non-restrictive relative clauses (5) and fragments functioning as complete utterances (signalled by a fullstop) (6) are also treated as EDUs.

(5) [Echter gedurende de nacht, [die op Mercurius maanden lang kan duren,] daalt de temperatuur tot zo'n -185 graden Celsius.]
[However during the night, [which can last for months on Mercury,] the temperature drops to about -185 degrees Celsius.]

(6) [Namens de dieren bedankt.]
[Thanks on behalf of the animals.]

In contrast to Tofiloski et al. (2009), we also include coordinated elliptical clauses (7) as EDUs, because the two clauses that share a verb can be seen as two separate predicates functioning in a discourse relation.

(7) [De planeet draait in 58.6 dagen om haar as] [en in 88.0 dagen om de zon.]
[The planet rotates around its axis in 58.6 days][and around the sun in 88.0 days.]

Restrictive relative clauses, subject and object clauses, and complement clauses are not treated as separate EDUs. In this we differ from Carlson and Marcu (2001), who treat restrictive relative clauses as embedded EDUs and separate the complements of attributive and cognitive verbs.

3 Implementation

The Dutch Discourse Segmenter(DDS) identifies EDU boundaries in a text. First, Alpino, a parser for Dutch text (Van Noord et al. (2006)), is used for sentence tokenization and building a syntactic tree for each sentence. After that, the syntactic trees are used as input for the actual segmenter to identify EDU boundaries.

The output of Alpino is a dependency tree, represented in XML (see Fig. 1 for a tree example). The discourse segmenter is implemented as an Xquery script that takes Alpino XML dependency trees from a text as input and outputs the positions in the text where a segment boundary should be inserted (see Fig. 2).

The script uses syntactic information and punctuation to insert EDU boundaries within sentences. It distinguishes main clauses and complementizer phrases as separate EDUs and uses the Alpino tags for this:

– Main clauses are identified by the syntactic category tag *@cat="smain"*. This tag is used for declarative sentences with a verb in the second place. All text parts with this tag are identified as EDUs. However, there can be other EDUs inside a main clause. After identifying a main clause, the algorithm searches through the nodes of the main clause for text spans that form an EDU in itself (e.g. embedded EDUs). If such a text construction is found and the remaining part of the main clause also forms an EDU, then the main clause is divided into two separate EDUs.

Fig. 1. Alpino tree example from the EDU ”en dat hun reis helemaal niet heerlijk is?” (*and that their trip is not delightful at all?*)

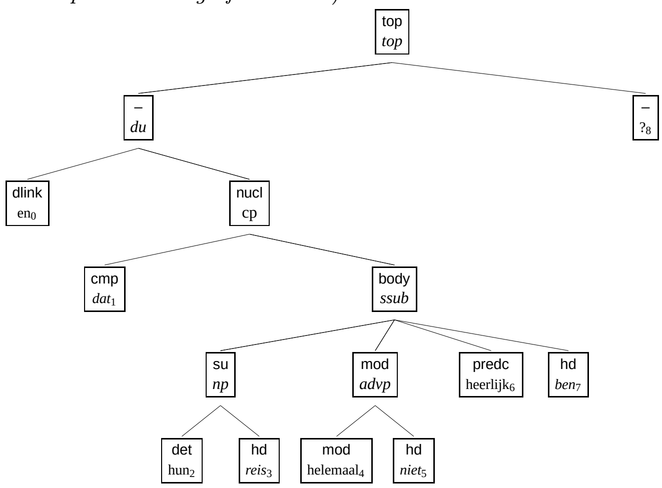


Fig. 2. Output Xquery script for example 8

(8) [De zon is zo dicht bij de aarde][dat we het oppervlak in detail kunnen bestuderen ,][wat bij de meeste andere sterren onmogelijk is .](*The sun is so close to the earth / that we can study the surface in detail, / which is not possible for most other stars*).

```
Output:
<edu sentence="8" begin="0" end="8"/>
<edu sentence="8" begin="8" end="17"/>
<edu sentence="8" begin="17" end="26"/>
```

- The tag *@cat=”sv1”* is used for sentences with the verb in the first place. Examples are yes/no questions and sentences in imperative mood. Text spans with this tag are identified as EDUs, but just as main clauses can be divided into smaller EDUs if parts of the sentence are an EDU in itself.
- Complementizer phrases are recognized using the Alpino tag *@cat=cp*, that is used for phrases that begin with a subordinate conjunction. The clauses that begin with a complementizer (e.g. *omdat*), marked by the part-of-speech tag *@pos=”comp”* in Alpino, and have a subordinate clause as body (*@cat=”ssub”* and *@rel=”body”*) are identified as EDU.
- Infinitival clauses are only identified as EDUs if they do not function as a complement. The tag *@cat=”oti”* is used for *om te* (*to*) - clauses. Only if an *om te*-clause is classified as a modifier with the tag *@rel=”mod”* it is taken as a separate EDU by the segmentation script.

– Relative clauses are identified in Alpino with the tag *@cat="rel"*. Relative clauses that do not contain an NP-antecedent are marked with the tag *@cat="whrel"*. Our segmentation only takes non-restrictive relative clauses as EDUs, but using the Alpino labels there is no guaranteed way to identify them. According to the Dutch punctuation rules, non-restrictive relative clauses should be preceded by commas. The segmenter uses this punctuation rule and inserts EDU boundaries only in cases where a relative clause is preceded by a comma.

4 Data and Evaluation

4.1 Data

Our data consists of human-annotated texts from two domains: Encyclopedia texts about astronomical objects and fundraising letters from various non-profit organizations. For the development of the segmenter we used a training set of 10 texts: 5 encyclopedia texts and 5 fundraising letters. Our test set consists of 30 texts. There are 15 encyclopedia texts, which vary in length between 14 and 33 sentences and have an average sentence length of 17.84 words. The remaining 15 texts are fundraising letters, which vary in length between 16 and 30 sentences and have an average sentence length of 12.92 words.

The texts were segmented by two trained annotators following the segmentation principles established in the project. Inter-annotator agreement using Kappa shows a high level of agreement for both genres: 0.97 for the encyclopedia texts and 0.99 for the fundraising letters.

4.2 Evaluation

For evaluation we use standard precision, recall and F-scores. Recall is the number of correctly inserted segment boundaries divided by the total number of segment boundaries in the gold standard segmentation. Precision is the number of correctly inserted segment boundaries divided by the total number of segment boundaries given by the system. For the F-scores we use F1 score (F= 2 * precision * recall/(precision + recall)). Note that our evaluation does not consider sentence boundaries as segment boundaries, because the task to segment text into sentences is done by Alpino and not by the discourse segmentation algorithm. Including sentence boundaries would inflate the scores.

5 Results

Table 1 lists the segmentation results for the encyclopedia texts and fundraising letters. Although the nature and structure of these two types of text is quite different, the results show a reasonable agreement with the manual annotated texts for both domains. The results of Tofiloski et al. (2009), the segmentation for English that we follow the most closely, are slightly better, but are produced

Table 1. Segmentation results for Encyclopedia texts and Fundraising Letters

Text type	*Boundaries*	*Recall*	*Precision*	*F-score*
Encyclopedia	164	0.67	0.73	0.70
Fundraising	92	0.74	0.76	0.75

using both syntax-based and lexical rules, whereas the DDS is purely based on syntax and punctuation.

The DDS makes use of the syntactic parser Alpino to identify EDU boundaries. However, the automatically generated Alpino dependency trees can contain parse errors, which could influence the results of the segmentation algorithm. To measure to what extent segmentation errors could be retraced to parse errors, we manually corrected the Alpino parse trees for ten texts and computed precision, recall and F-scores. As shown in table 2, some segmentation errors are due to errors in the syntactic parse trees produced automatically by Alpino. Using the manually corrected gold standard parses leads to an increase of almost 10% in the F-scores.

Table 2. Segmentation results beased on automatic parsed and manually corrected (gs) parse

		automatic parse			**gs parse**		
Text type	*Boundaries*	*Recall*	*Precision*	*F-score*	*Recall*	*Precision*	*F-score*
Encyclopedia	67	0.68	0.75	0.71	0.73	0.91	0.81
Fundraising	35	0.77	0.73	0.75	0.83	0.74	0.78
Total	102	0.72	0.75	0.73	0.75	0.92	0.82

The segmenter has the most difficulties with non-restrictive relative clauses, list structures, conjunctions and elliptical clauses. As described above, the segmenter only inserts segment boundaries when non-restrictive relative clauses are preceded by a comma, so it does not recognize other non-restrictive relative clauses as in (9).

(9) [Er kunnen prachtige lussen en uitsteeksels gevormd worden][die we protuberansen noemen.]
[Beautiful loops and protrusions can be created,][which we call protuberanses.]

In our manual segmentation we segment lists of fragments that could be seen as a series of complete utterances (see example 10) into separate EDUs, but the DDS cannot recognize these independent fragments.

(10) [Even geen dokters,][even geen onderzoek],[even geen behandeling][en even geen teleurstelling.]

[For a moment no doctors,][for a moment no research,][for a moment no treatment,][and for a moment no disappointment.]

Examples of conjoined subordinated clauses (example 11) and elliptical clauses (example 12) that the DDS cannot capture are:

(11) [We hebben een netwerk van vrijwilligers,][die deze gebieden goed kennen] [en precies weten wat nodig is.]
[We have a network of volunteers,][who know these areas well][and know exactly what is needed.]

(12) [Met een bijdrage van 15 euro geef je bijvoorbeeld al een raam][en voor 25 euro een buitendeur].
[With a gift of 15 euro you already give for example a window][and for 25 euro a door.]

Alpino does not attempt to resolve the ellipsis in conjunction-reduction constructions as in (12) and thus does not recognize the reduced second conjunct as an S-node.

6 Conclusion

We presented discourse segmentation guidelines for Dutch text and implemented these principles in a rule-based discourse segmenter that uses syntactic information and punctuation to insert EDU boundaries in a text. Application of the segmenter to encyclopedia texts and fundraising letters leads to reasonable F-scores for both genres. Our next step will be to improve the segmenter by adding lexical rules, e.g. for cue phrases like *maar* (but), *zoals* (like), *bijvoorbeeld* (for example) and punctuation marks such as : and -. After that, we will use the segmenter as the first step in discourse parsing of Dutch text and start working on the automatic labeling of RST relations between EDUs.

7 Acknowledgement

This work has been supported by grant 360-70-282 of the Netherlands Organization for Scientific Research (NWO) as part of the NWO-funded program Modelling discourse organization (http://www.let.rug.nl/mto/).

Bibliography

Abelen, E., Redeker, G., and Thompson, S. (1993). The rhetorical structure of US-American and Dutch fund-raising letters. *Text*, 13(3):323–350.

Carlson, L. and Marcu, D. (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545.*

Corston-Oliver, S. (1998). Beyond string matching and cue phrases: Improving efficiency and coverage in discourse analysis. In *The AAAI Spring Symposium on Intelligent Text Summarization*, pages 9–15.

Den Ouden, J., Van Wijk, C., Terken, J., and Noordman, L. (1998). Reliability of discourse structure annotation. *IPO Annual Progress Report*, 33:129–138.

Grabski, M. and Stede, M. (2006). Bei: Intraclausal Coherence Relations Illustrated With a German Preposition. *Discourse Processes*, 41(2):195–219.

Le Thanh, H., Abeysinghe, G., and Huyck, C. (2004). Automated discourse segmentation by syntactic information and cue phrases. In *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA 2004), Innsbruck, Austria.*

Lüngen, H., Puskàs, C., Bärenfänger, M., Hilbert, M., and Lobin, H. (2006). Discourse segmentation of German written text. In *Proceedings of the 5th International Conference on Natural Language Processing (FinTAL 2006).*

Mann, W. and Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Soricut, R. and Marcu, D. (2003). Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the HLT/NAACL*, pages 228–235.

Sporleder, C. and Lapata, M. (2005). Discourse chunking and its application to sentence compression. In *Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP)*, pages 257–264.

Stede, M. (2004). The Potsdam commentary corpus. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, pages 96–102. Association for Computational Linguistics.

Subba, R. and Di Eugenio, B. (2007). Automatic Discourse Segmentation using Neural Networks. In *Proc. of the 11th Workshop on the Semantics and Pragmatics of Dialogue*, pages 189–190.

Timmerman, S. (2007). Automatic recognition of structural relations in Dutch text. *MA thesis, University of Twente.*

Tofiloski, M., Brooke, J., and Taboada, M. (2009). A Syntactic and Lexical-Based Discourse Segmenter. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 77–80. Association for Computational Linguistics.

Van Noord, G. et al. (2006). At last parsing is now operational. In *Verbum ex machina: actes de la 13e conférence sur le traitement automatique des langues naturelles (TALN 2006): Leuven, 10-13 avril 2006*, page 20.

Learning Positional Probabilities: An Automatic System for Ordering Adjectives

Zoë Bogart
University of Malta
University of Groningen

Abstract

The goal of this research is to build a system that, given two or more English adjectives, will order them in a way that is intuitively correct to native speakers. Our system combines a statistical approach with a learning algorithm to assign every word learned from a training corpus a weight according to its preference to appear as the left-most word in a set. The weights are then used to order new sets of adjectives. Results are compared against both a corpus and against human orderings, which were obtained through an online survey.

1 Introduction

The problem of ordering multiple prenominal adjectives crops up in the area of Natural Language Generation, and it is not a trivial one. The problem centers around the fact that certain orderings of adjectives seem more natural to native speakers than others. An example of such a natural ordering occurs in (a) below:

(a) The old green tin can

By contrast, the ordering in (b) would likely be rejected by most native English speakers as sounding somehow 'off':

(b) The tin green old can

Programs designed to generate natural-sounding language must thus be able to put adjectives into the correct order to generate a phrase like (a) and not one like (b).

Unfortunately, the question of why some orderings are preferred to others is a question that has not been satisfactorily answered by linguists, though many rules have been offered to account for at least some of the ordering preferences. These rules cover almost all domains of linguistic study, from phonological to syntactic to semantic to pragmatic. A computational linguist working on the problem of adjective-ordering might find these rules an attractive starting point from which to build a system, and in fact, Wulff (2003) built a system which uses Linear Discriminant Analysis to order adjectives by taking different rules/heuristics from all of the above domains into account. When tested against adjective groups in the British National Corpus (BNC), this system achieved a classification accuracy of 78% and an accuracy of 73.5% for previously unanalyzed adjective strings. Though Wulff admits that these results are far from ideal, she claims that they are merely indications of the fact that there are additional factors at play in determining adjective order. Though this is almost certainly true, Wulff's approach is already quite complex, and one could probably keep throwing more heuristics into the mix with no guarantee of significantly improving accuracy scores. More importantly, other systems which take a different angle and rely instead on statistical information to order adjectives have achieved greater accuracy.

For example, Malouf (2000) achieves an accuracy of 89.73% on ordering BNC adjective pairs using nothing but information about the probabilities with which adjectives appear in different positions. Mitchell (2009) also uses an approach that looks at positional preferences, but her technique involves dividing adjectives into classes based on the positions they tend to inhabit relative to a following noun phrase. This technique has the advantage that it can order more than just two adjectives (which Malouf's technique cannot), but it cannot order adjectives that fall into the same class, nor can it order adjective groups that include adjectives that did not appear in the training data. Thus, while Mitchell reports a precision score of 89.63%, her system only has 74.14% recall, giving it an overall accuracy of only 66.45%.

The approach we develop here is based somewhat on Malouf's positional probabilities technique, but it adds ideas from Mitchell's class-based ordering and, like her technique, can order more than just two adjectives. Unlike Mitchell's approach, our system can order adjective groups where some of the adjectives did not appear in the training data, and, because it does not rely on the use of adjective classes, it does not run into the problem of being unable to order adjectives of the same class.

Beyond these considerations, our approach differs most notably from previous ones in that it is informed by findings from the fields of Psycholinguistics, Cognitive Neuroscience, and Machine Learning. The basis for our taking such a broad scope in building the system lies at the heart of the problem itself, which is to order adjectives in a way humans find natural. Because this problem is so closely tied up with human intuition, we believe that by using information about child language acquisition and human learning patterns to aid in our design, we can better create a system that will perform as desired. Additionally, we use human data to evaluate our system — a step that, to our knowledge, has not been taken before in this particular domain. Though we also evaluate our system against a corpus, so as to compare it to other automatic adjective-ordering systems, we give evidence that a corpus evaluation makes little sense for this task and, in this particular case, a human evaluation may be a better means of adequately gauging performance.

2 Method

We combine a simple statistical approach with an iterative learning algorithm to assign each adjective a weight that reflects its preference for occurring as the leftmost word in a set of adjectives. The weights were obtained through training on the set of $262,876$ ordered adjective pairs from the BNC used in Malouf, 2000.[1]

The basic mechanism of the algorithm for assigning weights is as follows:

1. **Initial training:** The first 100 adjective pairs are examined, and each new adjective is added to the vocabulary and assigned an initial weight of 0. The weight of each adjective in each pair is then adjusted by a value of $\frac{1}{n+2}$, where n is the number of pairs seen so far in training. This value is added to the current weight of the leftmost adjective and subtracted from the current weight of the rightmost adjective.

2. **Pair generation:** Using the adjectives in the vocabulary and their weights, the system randomly generates some number m of adjective pairs such that in each generated pair, the word with higher weight appears on the left.

3. **Training with error adjustment:** For the next 100 adjective pairs in the training data, step 1 is repeated, but in addition, each pair encountered in the training set is compared to the set of pairs generated in step 2, and further weight adjustments are made if the new pair is a confirmation or disconfirmation of any of the generated pairs (these adjustments are described in more detail below).

[1]Data are publicly available at http://www-rohan.sdsu.edu/~malouf/pubs.html

4. Step 2 is repeated for the new vocabulary and new weights, then step 3 again, and so on until all the pairs in the training set have been seen.

A diagram of the process appears in Figure 1.

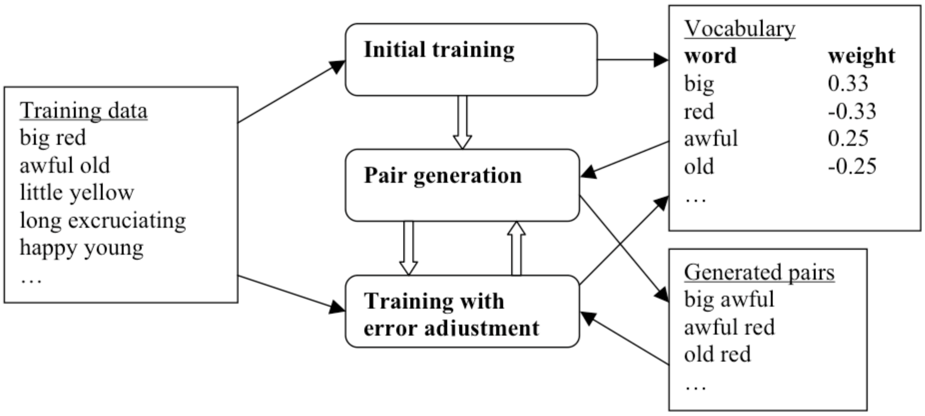


Figure 1: Process for Assigning Weights

We note that the use of weights which are adjusted during training is a technique quite similar to that used in neural networks. The learning algorithm in step 3 adds a twist to this technique by readjusting weights in a way that could be compared (on a very abstract level) with the way humans reevaluate learned material in the light of new data that confirms or disconfirms previously held beliefs. In our system's 'learning', weights are adjusted by the factor $\frac{c}{n+2}$ where c is one of 4 constant values as shown in the summary of adjustments in Table 1.

This additional adjustment of weights has the effect of confirming 'correct' weights and negating or correcting 'incorrect' weights. Values for SCONF, SNEG, BCONF, and BNEG (which stand for 'small confirmation', 'small negation', 'big confirmation', and 'big negation' respectively) were determined through a combination of trial-and-error testing to find ratios that gave higher accuracy and through use of a self-adjuster that iteratively updated the values according to the number of errors (negations) the system made in the most recent generation of test pairs in comparison with the number of errors it had made in the previous generation of test pairs. All parameter testing was done on a development set of sentences taken from

Training pair compared to generated pair	**Adjustment to left adjective's weight**	**Adjustment to right adjective's weight**
Exact match	$w + \frac{BCONF}{n+2}$ *	$w - \frac{BCONF}{n+2}$
Exact opposite	$w + \frac{BNEG}{n+2}$	$w - \frac{BNEG}{n+2}$
Left adjective matches	$w + \frac{SCONF}{n+2}$	none
Left adjective opposite	$w + \frac{SNEG}{n+2}$	none
Right adjective matches	none	$w - \frac{SCONF}{n+2}$
Right adjective opposite	none	$w - \frac{SNEG}{n+2}$

* w = adjective's current weight, n = number of pairs previously seen in training

Table 1: Weight Adjustments in Learning Algorithm

the ukWaC British English web corpus.[2] Final values for SCONF, SNEG, BCONF, and BNEG were 1, 20, 10 and 25 respectively.

In steps 1 and 3, the use of n in the denominator of the weight adjustment factors has the effect of adjusting weights by a greater magnitude for words encountered earlier in the training data. The more pairs that are seen, the smaller the adjustments to the weights become. This mimics patterns of human learning, where the first examples of a new domain learned tend to make a deeper impression than examples seen later on.[3] The use of the constant value 2 in the denominator is simply a corrective measure to avoid assigning an extremely large weight to the first leftmost word and an extremely small weight (large negative weight) to the first rightmost word.

The number m of test pairs to generate on each iteration was found to be an important factor in the performance of the system. Performance generally improved with increasing values for m, up to a certain point (about 5000 in our experiments), at which point increasing the number of test pairs generated did not result in any further improvements in performance. A chart showing the effect of generating different numbers of test pairs on overall accuracy appears in Figure 2 (logarithmic scale is used for the numbers of pairs generated).

[2] Data accessed through the Sketch Engine website at http://www.sketchengine.co.uk/

[3] We note that such an approach builds in a greater sensitivity to the training data. It might thus be desirable to use a training set consisting of more basic words for the first part of training and another training set which includes all words in the corpus for the second part of training. The use of basic or salient adjectives in the first training set would reflect patterns observed in child language acquisition (Clark 1993, Nelson 1973).

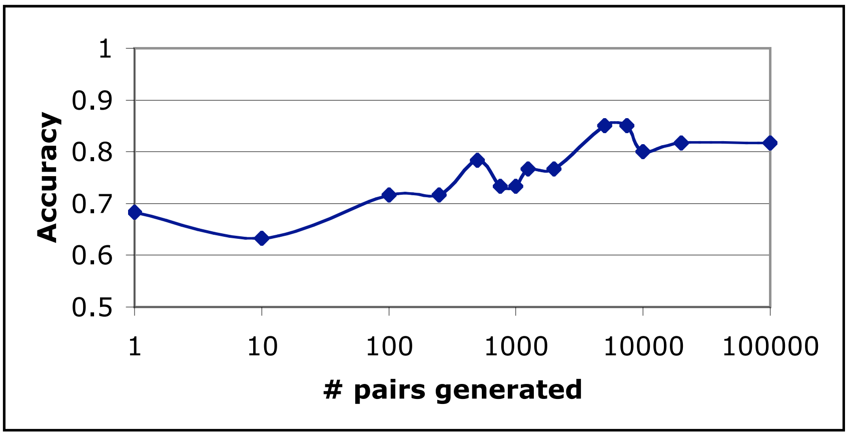


Figure 2: Pairs Generated vs. Accuracy

3 Results

Though it has been standard practice to evaluate systems like ours against corpora of human speech and text, we believe that the particular task of ordering adjectives is not well-suited for such an evaluation measure in the way that a task like parsing or part-of-speech tagging is. In those tasks, there is almost always a 'correct' solution, whereas in adjective ordering, there are often adjective pairs or groups for which multiple orderings are acceptable. For example the adjectives *new* and *classy* could be ordered either way before a modified noun to give a phrase that sounds natural to native speakers of English.[4] In these cases, evaluation of an adjective-ordering system against a corpus might penalize the system for producing an ordering which differs from that found in the corpus but which sounds completely natural to humans. For this reason we evaluated our system both against a corpus and against human judgments obtained via an online survey.

[4]Interestingly, a Google search for the phrases "classy new restaurant" and "new classy restaurant" turns up far more results for the former phrase than for the latter (12, 200 vs. 1, 080), but a search for the phrases "classy new hotel" and "new classy hotel" turns up far more results for the latter (7, 530 vs. 33, 700). These extremely informal results suggest the following noun phrase may play an important role in determining adjective order.

3.1 Corpus Evaluation

For the corpus evaluation, we performed 10-fold cross-validation on the BNC adjective pairs. Overall accuracy by this measure was 81.09% with precision of 81.63% and recall of 90.72%.[5]

3.2 Human Evaluation

Clearly, one cannot measure human 'accuracy' in the same way as one can measure the accuracy of an automatic system. To begin with, we thought it likely that there would be a certain amount of variation in human judgments, and so we obtained judgments from a sample of 60 subjects, all of whom were self-reported native English speakers. For each adjective group ordered, the order selected by the majority of subjects was used to give an indicator of the 'correct' human result. These orders were then tested for accuracy against the corpus, as was the automatic system, and results were compared. The human judgments were also compared more directly to the orderings given by the system through rank correlations based on the percentage of disagreement among humans and the differences between weights for the system. All of these methods are described in more detail below.

In order to compare the results of the automatic system against human judgments, a test set of 60 adjective sets was selected from randomly chosen sentences in the ukWaC British English web corpus, and subjects were asked to put the adjectives in these sentences into an order. Each survey question contained the noun or noun phrase from one of the ukWaC sentences preceded by a number of blanks corresponding to the number of adjectives in the sentence. The adjectives to be filled into the blanks were displayed below the question in random order, and subjects had to select one adjective to be the first word, one to be the second, etc. While almost all the sentences only contained two adjectives, two of the sentences contained three adjectives. An example of a question in the survey format is shown in Figure 3.

Subjects were invited to take the online survey via emails, Facebook™ posts, and the following websites that post links to psychological studies:[6]

[5]For calculating precision and recall, we counted pairs as being unclassified if either of the two adjectives had never been seen in training; however, the system is able to order such pairs if one of the other adjectives has been seen in training, so these pairs are included in the calculations for overall accuracy.

[6]The author is indebted to John H. Krantz of the website "Psychological Research on the Net" (http://psych.hanover.edu/Research/exponnet.html) and Ulf-Dietrich Reips of "The Web Experiment List" (http://genpsylab-wexlist.unizh.ch/) for their agreement to

______	______ authority
word one	**word two**

	word one	**word two**
local	◎	◎
relevant	◎	◎

Figure 3: Sample Survey Question

http://psych.hanover.edu/Research/exponnet.html
http://genpsylab-wexlist.unizh.ch/
Data were collected over a period of two weeks in January 2010.

3.3 Human Results

Individuals were scored for 'accuracy' by comparing their orderings to the corpus orderings, and by this measure, individual scores ranged from as low as 55% to as high as 95%. Mean accuracy was 84.6% with a standard deviation of 7.4%. When the majority ordering for each question was used to calculate overall accuracy, the human results gave 91.7% accuracy. Besides differing notably from the individual accuracy scores (only 11 of the 60 subjects had scores at or above 91.7%), these numbers mask vast discrepancies in inter-subject agreement which could reasonably be interpreted to reflect the "naturalness" of certain orderings, vis-à-vis others. Inter-subject agreement was thus measured with Fleiss' kappa[7] and found to be 0.31. Though no standardized interpretation of kappa values exists, Landis and Koch (1977) offer a rather ad hoc scale in which values between 0.21 and 0.40 are interpreted as indicating an agreement strength of "fair", which falls below "moderate" agreement (0.41–0.60) and above "slight" agreement (0.00–0.20). This kappa value of ∼0.30 does seem to indicate that there is often not a canonical correct adjective order where human judgments are concerned.

3.4 System Results

When tested on the set of 60 ukWaC sentences, the system achieved an accuracy of 85% (51/60) with precision of 84.21% and recall of 95%.

post the survey used in the human evaluation section on their respective websites.

[7]This measure was only applied to the 58 sentences containing adjective pairs, as it would have been more difficult to apply to the sentences containing three adjectives.

4 Analysis

Overall accuracy for the BNC adjective groups is higher than that reported by Wulff or Mitchell but lower than that achieved by Malouf. Though the system's accuracy is lower than that reported by Malouf, it maintains the advantage of being able to order groups of more than just two adjectives. Beyond this, the question of whether evaluation against a corpus is really appropriate for this task remains an issue. Indeed, results of the human-elicited judgments suggest that such an evaluation may not be appropriate—at least if the goal is to order adjectives in as natural and human-like a way as possible! The relatively low inter-subject agreement indicates that adjective order is not always clear-cut and so a purely corpus-based evaluation for an automatic adjective-ordering system may not accurately reflect the extent to which the system is capable of ordering adjectives in a way that seems natural and acceptable to humans.

Unfortunately, the inadequacy of corpus-based measures for evaluation does not guarantee that a comparison against human orderings will be any more helpful. The variation in human orderings makes it difficult to determine whether a given automatically generated ordering is acceptable or not, just as this variation makes it difficult to point to a corpus as being a gold standard for certain orderings. It is in confronting this issue that the use of weights in our system could prove useful. As a weight expresses a given adjective's preference for appearing closer to or farther from the following noun phrase, one could reasonably hypothesize that cases where multiple adjectives receive very similar weights are cases in which the system is less 'certain' about the ordering. If the system is doing a good job of mimicking human ordering preferences, then the cases where the system is less certain should be the same as the cases for which there was low inter-subject agreement between humans. To test our system by this means, we used the difference between weights as a measure of uncertainty for the system, and we then compared the system's uncertainty to the humans' uncertainty. Results are discussed below.

4.1 Comparing Uncertainty

We used the Pearson product-moment correlation coefficient to compare certainty rankings for humans with rankings for the system (we could not use the simpler Spearman's rank correlation coefficient as many pairs shared ranks for human judgments). The Pearson product-moment correlation coefficient was 0.4124, with a t-value of 3.388, showing the rankings to be significantly correlated with 99.9% certainty for the 1-sided t-test and with 99.8% certainty for the two-tailed t-test.

Though the cases where the majority of humans ordered adjective pairs in a way that did not match the corpus were all cases for which certainty was low, the same pattern did not hold for the automatic system. This could be due to extremely large weights for certain words, e.g. 'new', which had a weight of nearly 48 (most weights fell between the range of -1 and 1). Such a weight would result in a very high certainty score and would also cause new to be placed on the left in almost every case, which could lead to errors as it would not allow the weights of the other words in the set to be taken into account.

5 Conclusions

In this research, we have attempted to build a system that correctly orders a given group of adjectives, where 'correct' is interpreted to mean an ordering that is acceptable to humans. We used a traditional corpus-based evaluation in order to be able to compare our system to previously built systems, and we also used a human evaluation to compare our system against human orderings. The corpus evaluation showed that, while not performing as well as the state-of-the-art systems, our system, which is based solely on probabilistic information about adjective position, outperforms systems that rely on more complex linguistic information. As the statistical information about adjective positions is much easier to discover and implement computationally, these results indicate that working with this method may be preferable for the adjective-ordering task.

Our use of a learning algorithm added a twist to traditional training methods in Natural Language Processing tasks, as we attempted to use knowledge about human learning processes to inform our own system's means of 'learning' to order adjectives. This learning algorithm turned out be an important factor in the system's performance, as shown by tests that demonstrated increasing ordering accuracy for increasing numbers of test pairs generated in learning. Further, the improved performance of the system when weight adjustments for wrongly-ordered generated pairs are several times larger than weight adjustments for correctly-ordered test pairs provides an interesting parallel to work in psycholinguistics and neuroscience (Wills et al., den Ouden et al.) indicating the greater impact of error-making, as opposed to making correct predictions, in human learning.

Though we used a corpus-based evaluation in order to compare our system with previously built systems, the results of the human portion of our evaluation provide strong evidence that a corpus-based evaluation measure will not give an accurate picture of a system's ability to order adjectives in a way that is acceptable to humans. The study showed that there are many sets of adjectives for which human agreement on ordering is low, and in these situations it may be better to count any ordering produced by the system as correct. Unfortunately, it is much more difficult to evaluate a system against judgments elicited from humans, as it is impractical to obtain such judgments for any significantly large number of adjective orderings. Designing a means of evaluation that can measure a system's performance against human intuitions for a large number of adjective sets remains a problem for future research.

Finally, there remain some specific issues with the implementation of our system, namely its overly deterministic nature (i.e., it never takes the following noun or other context into account and so will always order any two adjectives in the same way), and its sensitivity to training data, which can result in overly large or small weights for specific words. Both problems could likely be resolved by modifying the program to take some semantic and/or pragmatic information into account. We plan to work on these modifications in future work, and we remain enthusiastic about the possiblities for probabilistic learning systems of this nature. Additionally, we hope that our results will inspire others working in the field of Computational Linguistics to incorporate more of the findings we are gaining from the disciplines of Psycholinguistics and Cognitive Neuroscience in their work to produce systems whose output is useful to humans.

References

[1] Baroni, M.; Bernardini, S.; Ferraresi, A. and Zanchetta, E. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation* 43(3): 209–226.

[2] Burnard, L. 1995. Users reference guide for the British National Corpus, version 1.0. Technical report, Oxford University Computing Services.

[3] Clark, E. V. 1993. *The Lexicon in Acquisition.* Cambridge: Cambridge University Press.

[4] Clark, E. V. 1997. Strategies and the Mapping Problem in First Language Acquisition. In J. Macnamara (ed.) *Language Learning and Thought*. New York: Academic Press, 147–168.

[5] Danks, J. A. and Schwenk, M. A. 1972. Prenominal Adjective Order and Communication Context. *Journal of Verbal Learning and Verbal Behavior* 11, 183–187.

[6] Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5): 378–382.

[7] Kilgarriff, A.; Rychly, P.; Smrz, P. and Tugwell, D. 2004. The Sketch Engine. *Proceedings of the* 11^{th} *EURALEX International Congress*, 105–116, Lorient, France.

[8] Landis, J. R. and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

[9] Malouf, R. 2000. The order of prenominal adjectives in natural language generation. *Proceedings of the* 38^{th} *Annual Meeting of the Asssociation for Computational Linguistics*, 85–92, Hong Kong.

[10] Mitchell, M. 2009. Class-based ordering of Prenominal Modifiers. *Proceedings of the* 12^{th} *European Workshop on Natural Language Generation*, 50–57, Athens, Greece. Association for Computational Linguistics.

[11] Nelson, K. 1973. Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* 38 1/2, 1–135.

[12] den Ouden, H. E. M.; Friston, K. J.;, Daw, N. D.; McIntosh, A. R. and Stephan, K. E. 2008. A Dual Role for Prediction Error in Associative Learning. *Cerebral Cortex* 19, 1175–1185.

[13] Quirk, R.; Greenbaum, S.; Leech, G. and Svartvik J. 1972. *A Grammar of Contemporary English*. Harlow: Longman.

[14] Whorf, B. L. 1945. Grammatical Categories. *Language* 21, 1–11.

[15] Wills, A. J.; Lavric, A.; Croft, G. and Hodgson, T. L. 2007. Predictive learning, prediction errors and attention: Evidence from event-related potentials and eye-tracking. *Journal of Cognitive Neuroscience* 19, 843–854.

[16] Wulff, S. 2003. A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8:2, 245–282.

Posters

Using Signals to Improve Automatic Classification of Temporal Relations

Leon Derczynski and Robert Gaizauskas

University of Sheffield, Regent Court, Sheffield S1 4DP, UK

Abstract. Temporal information conveyed by language describes how the world around us changes through time. Events, durations and times are all temporal elements that can be viewed as intervals. These intervals are sometimes temporally related in text. Automatically determining the nature of such relations is a complex and unsolved problem. Some words can act as "signals" which suggest a temporal ordering between intervals. In this paper, we use these signal words to improve the accuracy of a recent approach to classification of temporal links.

1 Introduction

The ability to order events, and the ability to determine which information is valid at a given time, are important in practical NLP. Effective automation of tasks such as summarisation and question answering require information extraction methods that can interpret information about time stored in documents.

One difficult problem in temporal information extraction is the ordering of events. Although accurate event ordering has been the topic of much research [1, 10, 5, 8], work using the temporal signals present in text – for example, phrases such as *after*, *for the duration of* and *while* – has been limited, and often only yields a minimal benefit [11]. Clearly these words contain temporal ordering information that human readers can access. This paper investigates the augmentation of a recent, high-performance temporal link classifier with information about temporal signals.

Our hypothesis is that signals provide information useful to TLINK classification. We also present data on signal usage within a temporally annotated corpus, in an attempt to gauge the likelihood of their being helpful and establish an upper bound on performance. After replicating existing work as a basis for comparison, we add signal-specific features and show how they lead to an improvement in classifier performance.

In this paper, we begin by describing the temporal annotation schema we have chosen to use (TimeML [12]) and provide a definition of temporal signals in the context of this paper (Sect. 2. In Sect. 3, we describe firstly how results from a previous experiment by Mani et al. [10] are replicated, and then detail the introduction of signal information into our system. Following this in Sect. 4 we detail our results, provide analysis in Sect. 5, and conclude in Sect. 6.

2 Background

Here we will introduce the annotation used in this work, introduce problems with temporal signals, and cover some of the relevant literature.

2.1 Temporal Annotation

In order to capture temporal information well, a sophisticated annotation schema is required. We use the TimeML schema [12], which includes tags for event and time expression annotation (`<EVENT>` and `<TIMEX3>` respectively), as well as temporal relations between intervals (`<TLINK>`) and signal phrases (`<SIGNAL>`). The two largest resources of TimeML annotated text are TimeBank [13] and the AQUAINT TimeML corpus[1], which we merge to form a corpus for this work.

2.2 Temporal Links

Temporal links (or **TLINK**s) describe a temporal relation between two intervals, each of which is either an event or a time expression. Allen [2] describes a set of relation types in terms of the interval endpoints. As our work is based on TimeML-annotated data, we use the set of TimeML relations, which are similar to Allen's. Each temporal link can optionally reference a signal.

2.3 Signals

Signals in TimeML are used to indicate multiple occurrences of events (temporal quantification) and also to mark words that indicate the type of relation between two intervals. For event ordering we are only interested in this latter use of signals. "A University Grammar of English" [14] lists a subset of these words in Sect. 10.5, "Time Relaters".

For example, in the sentence *John smiled after he ate*, the word *after* specifies an event ordering. This example could be represented in TimeML as follows:

```
John <EVENT id="e1"> smiled </EVENT> <SIGNAL id="s1"> after </SIGNAL>
he <EVENT id="e2"> ate </EVENT> .
<TLINK id="l1" eventID="e1" relatedToEvent="e2"
  relType="AFTER" signalID="s1" />
```

TimeML allows us to associate text that suggests an event ordering (a signal) with a TLINK. To avoid confusion, it is worthwhile clarifying our use of the term "signal". We use **SIGNAL** in capitals for tags of this name in TimeML, and **signal/signal word/signal phrase** for a word or words in discourse that describe the temporal ordering of an event pair. Examples of the signals found in TimeBank are provided in Table 1. It is important to note that not every occurrence of text such as *after* is a temporal signal. What is not shown due to space constraints is that a temporal signal such as *after* may be used by (for example) 39 TLINKs labelled AFTER, 17 labelled BEFORE, and four labelled INCLUDES; the signal text alone does not infer a single interpretation.

[1] Available for download from http://timeml.org/site/timebank/timebank.html.

Phrase	Corpus freq.	Occurrences as signal	Likelihood of being signal
subsequently	3	3	100%
after	72	67	93%
's	10	8	80%
follows	4	3	75%
before	33	23	70%
until	36	25	69%
during	19	13	68%
as soon as	3	2	67%

Table 1. A sample of phrases most likely to be annotated as a signal when they occur in TimeBank, which occur more than once in the corpus. All corpus data in this paper was provided by the CAVaT command-line tool [6].

2.4 Previous work

When temporally ordering events, it is intuitively likely that signal information may be useful. The trend in previous automated TLINK classification work has not been to directly target signals as a primary source of ordering information, although other attributes of annotated TLINKs and EVENTs have been exploited as training features. For example, the best known automatic TimeML annotation tool (TARSQI [15]) performs no SIGNAL annotation. Lapata and Lascarides [7] worked with signals, using a restricted reference list of signal tokens instead of drawing signal text from human-annotated data. This work was only on same-sentence temporal links. Their accuracy at temporal relation classification was 70.7%. Bethard and Martin [3] included some features that described signals, where the `compl-word` feature (the signal text) was the 8^{th} strongest in their set of features for temporal relation classification. However, this work has a number of limitations. First, it only uses the signal word and a simple relation type suggestion as features. It is also restricted to verb-clause construction TLINKs. Finally, The classifier only has to choose from a set of three TLINK classes (before, overlap, after).

3 Method

To explore the question of whether signal information can be successfully exploited for TLINK classification, we proceed as follows. First we re-implement a well-known TLINK relation classifier with state-of-the-art accuracy. Then we add various signal-related features to the classifier to investigate their impact on classification performance. The approach we have replicated as closely as possible is from Mani et al. [9]. In brief, the method was as follows.

Firstly, the set of possible relation types was reduced by applying a mapping. For example, as `a` BEFORE `b` and `b` AFTER `a` describe the same ordering between events `a` and `b`, we can flip the argument order in any AFTER relation to convert it to a BEFORE relation. This simplifies training data and provides more examples per temporal relation class. Secondly, the following information from each TLINK is used as features: event class, aspect modality, tense, negation, event string for each event, as well as two boolean features indicating whether both events have the same tense or same aspect. Thirdly, we trained and evaluated the predictive accuracy of the maximum entropy classifier from Carafe[2] using 10-fold cross-validation.

Corpus	Total TLINKs	With SIGNAL	Without SIGNAL
TimeBank v1.2	6418	718 (11.2%)	5700
AQUAINT TimeML v1.0	5365	178 (3.3%)	5187
ATC (combined)	11783	896 (7.6%)	10887
ATC event-event	6234	319 (5.1%)	5915

Table 2. TLINKs and signals in our data.

TLINK data came from the union of TimeBank v1.2a and the AQUAINT TimeML corpora. As the corpus used in the previous work by Mani et al. (TimeBank v1.2a) is not publicly available, we used TimeBank v1.2. This use of a publicly-available version of TimeBank instead of a private custom version was the only change from the previous method. In this work we only examine event-event links, which make up 52.9% of all TLINKs in our corpus (See Table 2).

We will later (Sect. 3.2) add features that require data to be separated into test and training sets, with more sophistication required than that available in Carafe's maximum entropy classifier; thus, as well as performing 10-fold cross-validation (**XV**), we also split all event-event TLINKs into a training set of 4156 instances and an evaluation set of 2078 instances.

	Predictive accuracy	Baseline
Mani et al. results	61.79%	51.6%
Replicated results with our tools (10-fold XV)	60.32%	53.34%
Replicated results with our tools (train/test)	60.04%	53.34%

Table 3. Results from replicating one of MITRE's TLINK classification experiments.

3.1 Replicating Previous Work

Table 3 shows results from replicating the previous experiment on event-event TLINKs. The baseline listed is the most-common-class in the training data.

[2] Available at http://sourceforge.net/projects/carafe/.

We achieved a similar score of 60.32% accuracy compared to 61.79% in the previous work. The differences may be attributed to the non-standard corpus that they use. The TLINK distribution over a merger of TimeBank v1.2 and the AQUAINT corpus differs from that listed in the paper.

3.2 Introducing Signals to the Feature Set

To add information about signals to our training instances, we use the extra features described below; the two arguments of a TLINK are represented by **e1** and **e2**.

- **Signal phrase.** This shows the actual text that was marked up as a SIGNAL. From this, we can start to guess temporal orderings based on signal phrases. However, just using the phrase is insufficient. For example, the two sentences *Run before sleeping* and *Before sleeping, run* are temporally equivalent, in that they both specify two events in the order run-sleep, signalled by the same word *before*.
- **Textual order of e1/e2.** The textual ordering of linked events can be reversed without affecting temporal order. Thus, it is important to know the textual order of events and their signals even when we know a temporal ordering. This feature assumes that the order event-signal-event is most prevalent in text; values are either e1-e2 or e2-e1.
- **Textual order of signal and e1, signal and e2.** These features describe the textual ordering of both TLINK arguments and a related signal. It will also help us see how the arguments of TLINKs that employ a particular signal tend to be textually distributed.
- **Textual distance between e1/e2.** Sentence and token count between e1 and e2.
- **Textual distance from e1/e2 to SIGNAL.** If we allow a signal to influence the classification of a TLINK, we need to be certain of its association with the link's events. Distances are measured in tokens.
- **TLINK class given SIGNAL phrase.** Most likely TLINK classification in the training data given this signal phrase (or empty if the phrase has not been seen). Referred to as signal **hint**.Referred to as signal **hint**.

4 Results

Moving to a feature set which adds SIGNAL information, including signal-event word order/distance data, 61.46% predictive accuracy is reached. The increase is small when compared to 60.32% accuracy without this information, but TLINKs that employ a SIGNAL in are a minority in our corpus (possibly due to under-annotation). It would be interesting to see the performance difference when classifying only TLINKs that use a SIGNAL.

There are in total 11783 TLINKs in the combined corpus, of which 7.6% are annotated including a SIGNAL; for just TimeBank v1.2, the figure is higher at

Predictive accuracy	XV	Split
Baseline (most common class)	53.34%	53.34%
Without signal features	60.32%	60.04%
With basic signal features	**61.46%**	60.81%
With signal features including hint	n/a	**61.98%**

Table 4. TLINK classification with and without signal features, using both 10-fold cross validation and a one-third/two-thirds split between evaluation and training data.

11.2% (see Table 2). The proportion of signalled TLINKs in our data is lowest at 5.1%.

The results of extending the feature set over a split of signalled and unsignalled links is shown in Table 5, from a one-third/two-thirds evaluation/training split.

Predictive accuracy	Unsignalled links	Signalled links
Baseline	52.68%	64.21%
Plain features	**62.05%**	55.65%
Plain + signal features	62.05%	**69.57%**
Plain + signal features + hint	62.05%	41.72%

Table 5. Predictive accuracy from Carafe's maximum entropy classifier, using features that do or do not include signal information, over signalled and non-signalled TLINKs in ATC. The baseline is accuracy when the most-common-class is always assigned.

5 Analysis

From Table 1 we can estimate the probability that a word or word sequence can be annotated as a SIGNAL associated with a TLINK. This may be of use when annotating signals, especially in the AQUAINT TimeML corpus. In any case, given that our feature set might only be helpful to 5.1% of event-event links in the ATC corpus (Table 2), the maximum performance increase at predicting signalled links can be estimated.

Let us suppose that we have perfect signal discrimination and association. Suppose our extra features do not help TLINKs without SIGNALs, and that the increase in performance is due solely to better accuracy classification of TLINKs that use signals. Let accuracy at classifying this signalled minority be a. Given a proportion of signalled TLINKs s, and predictive accuracy of our classifier when using features that do not depend on signals P_n (from Table 4):

$$P_n(1-s) + as = 0.6146 \tag{1}$$

$$\mathbf{a = 0.8381} \tag{2}$$

Thus, we may be classifying signalled TLINKs at over 80% accuracy when using the augmented features. This indicates a significant increase in predictive accuracy for signalled event-event TLINKs from the previous accuracy of 60.32%. This is a target for classification of signal-employing TLINKs.

It is hard to determine an external upper bound for the classification of signal-employing TLINKs because inter-annotator agreement (IAA) figures are only available for TimeBank, and not at this level of detail. However, we can see from [4] that TLINK IAA reached 0.55. One would have to refer to the original annotator data and identify those TLINKs which were marked as employing a signal to determine an IAA value just for TLINKs with an associated SIGNAL. IAA for signals was 0.77.

We have hypothesised that adding features to represent signals in TLINK classification will lead to an increase in predictive accuracy. To test this, we repeat the above experiment, which compared features includign and excluding signal information. Data was divided into TLINKs that employ a signal, and those that do not. We expected to see similar prediction accuracy from both feature sets when classifying TLINKs that do not use signals. The baseline was the most common class in the dataset.

If there is no performance difference between feature sets when classifying TLINKs that *do* use signals our hypothesis is incorrect, or the features we used are bad representation. If signals are helpful, and our features capture information useful for temporal ordering, we expect a performance difference when evaluating signalled TLINKs. Results in Table 5 support our hypothesis that signals are useful, but we are performing nowhere near the maximum level suggested above. Data sparsity is a problem here, as the combined corpus only contains 319 suitable TLINKs, and both source corpora evidence of signal under-annotation. The results also suggest that the signal hint feature was not helpful; this is the same result found in [3].

Exploring the strongest feature set (basic+signals; no hint), attempting to combat the data sparsity problem, we used 10-fold XV instead of a split; results are in Table 6. This shows a distinct improvement in the predictive accuracy of signalled TLINKs using this feature set over the features in previous work.

Predictive accuracy	**Baseline**	**Plain features**	**Plain and signal features**
Unsignalled links	52.68%	**61.81%**	61.81%
Only signalled links	62.41%	60.32%	**82.19%**

Table 6. TLINK predictive accuracy using 10-fold cross validation over signalled and non-signalled TLINKs

6 Conclusion

When learning to classify signalled TLINKs, there is a significant increase in predictive accuracy when features describing signals are used. This suggests that signals are useful when it comes to providing information for classifying temporal links, and also that the features we have used to describe them are effective.

Future work is focused on improving signal and TLINK annotation. We need to explore how to discriminate whether or not a string is used as a temporal signal in text. Next, after finding a temporal signal, we need to determine which intervals it temporally connects. Finally, we can attempt to annotate a temporal link based on the signal. Once finished, we can integrate all this into existing temporal annotation tools.

References

1. D. Ahn, S.F. Adafre, and MD Rijke. Towards task-based temporal extraction and recognition. In *Dagstuhl Seminar Proceedings*, volume 5151, 2005.
2. J.F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
3. S. Bethard, J.H. Martin, and S. Klingenstein. Timelines from text: Identification of syntactic temporal relations. In *International Conference on Semantic Computing*, pages 11–18, 2007.
4. B. Boguraev, J. Pustejovsky, R. Ando, and M. Verhagen. TimeBank Evolution as a Community Resource for TimeML Parsing. *Language Resources and Evaluation*, 41(1):91–115, 2007.
5. N. Chambers and D. Jurafsky. Jointly combining implicit constraints improves temporal ordering. In *Proceedings of the EMNLP*, pages 698–706. ACL, 2008.
6. L. Derczynski and R. Gaizauskas. Analysing Temporally Annotated Corpora with CAVaT. In *LREC*, 2010.
7. M. Lapata and A. Lascarides. Inferring sentence-internal temporal relations. In *Proceedings of the NAACL*, 2004.
8. H. Llorens, E. Saquete, and B. Navarro. TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2. In *Proceedings of SemEval-2010*. ACL.
9. I. Mani, M. Verhagen, B. Wellner, C.M. Lee, and J. Pustejovsky. Machine learning of temporal relations. In *Proceedings of the 44th annual meeting of the ACL*, page 760. ACL, 2006.
10. I. Mani, B. Wellner, M. Verhagen, and J. Pustejovsky. Three approaches to learning TLINKS in TimeML. Technical report, CS-07-268, Brandeis University, 2007.
11. C. Min, M. Srikanth, and A. Fowler. LCC-TE: A hybrid approach to temporal relation identification in news text. In *Proceedings of SemEval-2007*, pages 219–222. ACL, 2007.
12. J. Pustejovsky, B. Ingria, R. Sauri, J. Castano, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, and I. Mani. The Specification Language TimeML. *The Language of Time: A Reader. Oxford University Press*, 2004.
13. J. Pustejovsky, R. Sauri, R. Gaizauskas, A. Setzer, L. Ferro, et al. The TimeBank Corpus. In *Corpus Linguistics*, pages 647–656, 2003.
14. R. Quirk and S. Greenbaum. *A University Grammar of English*. London, 1973.
15. M. Verhagen and J. Pustejovsky. Temporal Processing with the TARSQI Toolkit. In *Coling 2008: Posters and Demonstrations*, pages 189–192, 2008.

Tagger for Polish based on binary classifier

Bartosz Zaborowski

Warsaw University, Faculty of Mathematics, Informatics and Mechanics
Banacha 2, 02-097 Warsaw, Poland

Abstract. Tagging for languages with complex and rich tagsets is a computationally difficult task. To simplify matters, typical approaches assume that for each word there exists only one correct tag. In the case of Polish this assumption is wrong.
In this paper I describe a different approach to morphosyntactic annotation based on binary classifier applied at the tag level (not the word level). It performs well, as it is slightly better than other known tagging methods for Polish. What is more, it handles correctly cases in which more than one tag is correct.

1 Introduction

The typical approach to the problem of tagging of highly-inflected languages with rich tagsets assumes that for each word there is exactly one possible correct tag. It is not the best assumption in the case of Polish, since there exist linguistically justified cases where a few tags are desirable (see [Przepiórkowski, 2004] for details, fig. 1 gives an example). This tagging problem can be perceived as a statistical classification task, with classes or decisions typically corresponding to tags.

The aim of this project is to build a tagger for Polish which would deal with the above mentioned true ambiguities. Additional requirements are: appropriate resources (especially time) consumption, good tagging results and no or little human work needed. An additional useful feature would be the possibility to convert a trained model into a human-readable form of rules or decision tree for further quality improvement. The project has a proof of concept character and is rather short term, so only a prototype which satisfies the above requirements was built.

2 Overview

The typical attitude to tagging is a statistical classification with as many decision classes as there are tags in the tagset. However, if we take output obtained from the morphological analyser into consideration, we could approach this task as the classical binary classification problem. Let us suppose we have a word with several possible tags. Instead of choosing only one correct tag for this word, we can attach to each possibility one of the following classes: **correct** or **incorrect**.

```
(1)
Pamiętam        ją        pijaną.
remember.1ST  her.ACC  drunk.ACC/INS
'I remember her drunk.'

(2)
a. Pamiętam      go    pijanego.
   remember.1ST him.ACC drunk.ACC

b. Pamiętam      go     pijanym.
   remember.1ST him.ACC drunk.INS

'I remember him drunk.'
```

Fig. 1. An example of a tag ambiguity: for *pijaną* both interpretations are fully correct, regardless of context (taken from [Przepiórkowski, 2004]).

There are very many known classifiers which perform well in different circumstances, for example Support Vector Machines widely used in i.a. chemical and geographical domains and C4.5 in data mining. With regard to the size of the training data (order of millions of instances) the best algorithms and implementations should be searched among data mining applications (classification based on large training data is a very typical task in this field). In most implementations the data is expected to be in the tabular form, and each row is treated independently.

While adapting data mining classifiers to tagging two things should be taken into consideration: the choice of the right attributes (columns of table) and the conversion of the output from the morphological analyser format to the tabular format (and vice versae). Additionally, in order to reduce the *information noise*[1] in the input data for classifiers it is important to abstract from actual words in the classification process. Hence, initial, bootstrap tagging is needed before the actual classifier can be used.

The main disadvantage of this method is the evident information leak: during the conversion of the tag proposals for a word to the tabular form, the knowledge that the possible tags are attached to one specific word is lost. It is possible because of the representation of a single word as a set of independent rows in the table. Unfortunately, to the best of my knowledge, there do not exist any methods which would allow for the classifier to take into account the relationships between rows in the case when the number of rows (possible tags) differs between words and frequently reaches the value of several dozen rows.

[1] In our case mostly irregularities

3 Technical assumptions

The input data should be in the output format of the morphological analyser Morfeusz [Woliński, 2006] . It is the standard format of the IPI PAN Corpus of Polish (http://korpus.pl/ ; [Przepiórkowski, 2004]). This format uses XML files conforming a variant of EAGLES Corpus Encoding Standard.

A tagset used in the IPI PAN Corpus is structured. It encodes separately information of a grammatical class and each of the grammatical categories (consult [Przepiórkowski, 2004] for details on tagset of the IPI PAN Corpus).

4 Implementation

4.1 Attribute selection

As attributes of table the following attributes have been chosen:

- the number of possible tags for processed word
- the number of correct tag possibilities for this word
- split tags for four successive words (2 preceding words, processed word, 1 following word)

Tags have been split into the smallest pieces of information contained in them, e.g. grammatical class, gender, number, case etc. Each of them is used as a separate attribute. It is similar to a tagset conversion to a positional format of the four mentioned words simultaneously.

4.2 Classifier algorithm selection

In order to simplify the process of experimenting I have based my application on the debellor data mining library [Wojnarski, 2009] which allows to apply one of 100+ classifier algorithms from the well-known WEKA [WEKA] and Rses [Rses] data mining environments. It is slightly faster, has lower memory consumption compared to the WEKA environment and provides a nice API.

The series of experiments have shown that the best results, at small cost of time, can be achieved with the J48 classifier. It is an efficient implementation of the C4.5 tree classifier. The C4.5 is an algorithm which classifies using binary trees built using a concept of information entropy (computed on the training data). See [Quinlan, 1993] for detailed description.

4.3 Model training and tagging process

Training and tagging are performed iteratively.

1. During the first iteration both the training and the test data are initially tagged before they are converted to the tabular format. The conducted experiments have shown that the classifier achieves better results when it is trained on such corrupted data rather than when it is trained on the original, correct tags. For the initial tagging a simple unigram tagger is used.

Additionally, in order to slightly improve the results and speed up the operation the data are split according to grammatical classes after having been converted into the table. Then a separate model is trained for each class. It costs some additional memory resources, but allows easy parallelisation and, therefore, allows for the shorter computation time.

Classified test data are used as a source of test data in subsequent iterations. Tag values for each word are corrected according to those results. In cases in which the classifier marked all tag proposals as incorrect, one of tags is toggled with the help of the initial value set by the unigram tagger.

2. During the second and all the subsequent iterations the classifier model is trained on the original data which are not modified by the unigram tagger. The number of iterations may be adjusted depending on the available resources and the required quality.

4.4 Remarks

The effectiveness of the conversion from the IPI PAN Corpus format to the tabular format and unigram tagger have been slightly optimized in the course of their implementation, but the quality, not the efficiency was the aim of the project.

The presented approach allows the model to be obtained in a compact, human-readable form of the decision tree. However, with respect to the adopted assumptions and the aim of the project the conversion of the model to this form has not been implemented. It would require additional work as well as modification of the debellor library, which at the moment does not allow to export trained models.

The use of the standard, universal tabular data format (WEKA's .arff) enables further experiments/processing by virtually any available data mining or machine learning tools. The rich classifier library allows experimenting with other classifiers, including meta-classifiers (compositions or boosters of simpler classifier algorithms).

5 Results

Tagger development and evaluation were conducted using data from the IPI PAN Corpus of Polish (http://korpus.pl/ ; [Przepiórkowski, 2004]). The corpus consists of over 880,000 manually annotated words. The applied tagset contains over 1,000 tags used in practice (4,000+ theoretically possible combinations of parts of tags).

The proposed tagging method was developed on the basis of the 9/10 of the data, while the final evaluation was done on the remaining 1/10 (every tenth paragraph, starting from the first one, was evaluated in this phase). In order to obtain results comparable with other algorithms a full 10-fold cross-validation on the whole corpus was done.

Experiments have shown that the best results (in acceptable time) have been obtained when the J48 was used as a classifier. The quality of results improves with each iteration and normalizes after about 5 iterations. A few configurations of metaclassifiers (including the MultiBoostAB and the RotationForest, both with the J48 as a subclassifier) achieved better results than the J48 alone when they were trained on small parts of the corpus. Unfortunately, when applied to the whole large training data, they get (very) similar quality to the J48. The only difference is a lower number of iterations needed (however, since single iteration is much slower, the whole training lasts longer).

The experiments with different window sizes (different number of surrounding words used in construction of the table) have proved that the asymmetric window 2-1-1 is the most suitable for Polish. Bigger windows (3-1-1, 2-1-2, etc.) made the data more chaotic and made it difficult for classifier to distinguish accidental configurations from true grammatical rules. Smaller window concealed some relationships between words from the classifier. It can be seen as a slight improvement from the trigram taggers, which generally performs well in highly inflected Slavic languages (e.g. [Dębowski, 2004], [Hajic et al., 2001]).

5.1 Efficiency estimation

(results obtained on a computer with Athlon X2 2.3GHz processor and 8GB of RAM memory)

10-fold cross-validation of the whole corpus:	**time consumed**
1 iteration, J48 classifier	approx. 3.5h
5 iterations, J48 classifier	approx. 7.5h

6 Evaluation

The data source has been described in section 5 (Results). The employed evaluation measures have been defined and discussed in
[Karwańska, Przepiórkowski, 2009] .

6.1 10-fold cross-validation of whole corpus

The results of the 10-fold cross-validation of the whole corpus are presented in tables 1 and 2.

6.2 Discussion

Correctness of 88.4% places this method of tagging among the best known taggers of Polish. According to [Karwańska, Przepiórkowski, 2009] , two other taggers (Dębowski Tagger and TaKIPI) using the same testing methodology on the

measure	result
correctness	0.8775
weak correctness	0.9061
precision	0.8832
recall	0.8965
F-measure	0.8898

Table 1. Results for 1 iteration, whole corpus

measure	result
correctness	0.8838
weak correctness	0.9186
precision	0.8832
recall	0.8991
F-measure	0.8910

Table 2. Results for 5 iterations, whole corpus

same corpus achieved respectively: 87.4% and 86.6%. The weak correctness and the F-measure also suggest that this is a better approach to tagging.

Similarities in results among different classifiers considered to be the best suggest that the obtained results are near to the actual amount of information contained in the (preprocessed in such way) corpus. It seems that the bottleneck of the whole tagger is the unigram tagging (it certainly causes some information to be lost): when the ideal bootstrap tagger is simulated by using the original manual annotation the result of tagging by the classifier is approx. 4% better. The classifier significantly improves the quality of the unigram bootstrap (from less than 70% to nearly 90%), consequently, it should reach higher results with a better bootstrap.

7 Summary

The presented approach enables to build an effective tagger for Polish which uses admissible amounts of resources and is relatively fast. With small additional effort, thanks to the form of the classifier, the trained model can be converted to the human-readable form of decision rules (using conditions from the decision tree) and presented to linguists for further enhancement.

References

[Karwańska, Przepiórkowski, 2009] Danuta Karwańska and Adam Przepiórkowski. 2009. *On the Evaluation of Two Polish Taggers.* The proceedings of Practical Applications in Language and Computers (PALC 2009).

[Przepiórkowski, 2004] Adam Przepiórkowski. 2004. *The IPI PAN corpus: Preliminary version.* Institute of Computer Science, Plish Academy of Sciences.

[Woliński, 2006] Marcin Woliński. 2006. *Morfeusz - a practical tool for the morphological analysis of Polish.* Intelligent information processing and web mining.

[Wojnarski, 2009] Marcin Wojnarski 2009. *Debellor - open source framework for scalable data mining and machine learning.* http://debellor.org

[Quinlan, 1993] Ross Quinlan 1993. *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers

[WEKA] Weka Machine Learning Project. http://www.cs.waikato.ac.nz/˜ml/

[Rses] Rough Set Exploration System. http://logic.mimuw.edu.pl/˜rses/

[Dębowski, 2004] Łukasz Dębowski. 2004. *Trigram morphosyntactic tagger for Polish.* Proceedings of the International IIS:IIPWM'04 Conference.

[Hajic et al., 2001] Jan Hajic, Pavel Krbec, Pavel Kveton, Karel Oliva, Vladimír Petkevič. 2001. *Serial Combination of Rules and Statistics: A Case Study in Czech Tagging.* Proceedings of ACL Conference.

Ontology-based retrieval of bio-medical information based on microarray text corpora

Kim A. Hansen, Christian Theil Have and Sine Zambach
Research group PLIS: Programming, Logic and Intelligent Systems
Department of Communication, Business and Information Technologies
Roskilde University, P.O.Box 260, DK-4000 Roskilde, Denmark
{kallanh, cth, sz}@ruc.dk

Abstract. Microarray technology is often used in gene expression experiments. Information retrieval in the context of microarrays has mainly been concerned with the analysis of the numeric data produced; however, the experiments are often annotated with textual metadata. Although biomedical resources are exponentially growing, the text corpora are sparse and inconsistent in spite of attempts to standardize the format. Ordinary keyword search may in some cases be insufficient to find relevant information and the potential benefit of using a semantic approach in this context has only been investigated to a limited degree. We explore the possibilities of retrieving biomedical information from microarrays in Gene Expression Omnibus (GEO), of which we have indexed a sample semantically, as a first step towards ontology based searches. Through an example we argue that it is possible to improve the retrieval of biological information.

Key words: semantic annotation, ontologies, domain analysis

1 Introduction

Ontology-based information retrieval and extraction of semantic information in text corpora are techniques used in connection with text searches. These techniques can help to utilize biological domain knowledge in connection with the retrieval of gene expression experiments, which is a an area where the need for shared information is increasingly important. Gene expression profiles are kept on so called microarrays, and are used for data analysis [1]. Though information exchange can be difficult due to the lack of standardization there does exist a metadata guideline that outlines the Minimum Information About a Microarray Experiment (MIAME) [2].

The National Center for Biotechnology Information database, Gene Expression Omnibus (GEO), is a public functional genomics data repository supporting MIAME-compliant data submissions[3], where it is possible to retrieve and download microarray data. Retrieval of data is mainly based on a keyword based text search, which is not always capable of finding all data of interest. Instead, the use of a bio-ontological oriented approach could be beneficial in such cases. However, an ontological based text search in connection with microarray experiments has only been investigated to a limited degree.

This paper will focus on an experimental approach to a bio-ontological oriented search. By demonstrating this technique through an example, we want to show that it is possible to retrieve relevant information that will not be found using ordinary keyword search. Also, this might lead to a more intuitive approach to search for information in the microarrays. The biomedical data sources is based on metadata from the GEO database, which is imported into our local database. The imported metadata is indexed semantically in order to enable ontological inference. The method used to index the text corpus is similar to that in [4], which is closely connected to the SIABO project [5] and addresses problems about accessing the conceptual content of biomedical texts. Similarly, the microarray-oriented MGED Ontology [3] uses the same approach to fetch data from GEO as we do. They provide a framework which can be used by developers whose environments facilitates the use of ontologies in microarray metadata. It has not been convenient to use in our case, since it is difficult to extend with our method of semantic language processing [5].

2 Methods and Results

Gene expression experiments published in the GEO database are to be described in accordance with the MIAME standard [2]. Despite standardization efforts the amount and quality of the entered information varies a lot and searches in databases can therefore be quite a challenging task. In order to demonstrate the possibility to improve the retrieval of microarray information we analyze a sample GEO experiment using the logic programming language Prolog.

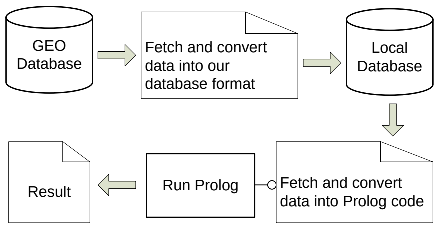


Fig. 1. Illustration of the data flow from fetching the data from GEO to generating Prolog code and returning results.

A GEO experiment is registered with a unique experiment id in the database and has links to microarrays, each registered with a unique microarray id. An experiment is conducted on a specific platform and even reuse of platforms can occur. These platforms are also described and each has a unique platform id and can be cross-referenced. To support our ontology-based search methodology, a process fetches the data from the GEO database. The data is converted into a Prolog format and saved in a local database. In this process *entities* and *properties* of those entities needs to be extracted from the textual descriptions associated with the experiments.

To illustrate how this method works, we make use of a human stem cells experiment (GSE6015) which has been fetched from GEO. The experiment is recorded according to the MIAME standard and is relatively well-described in all aspects. It would be expected to find such an experiment through ordinary keyword search. Ordinary keyword search would be expected to find such an experiment, but unfortunately this is not the case if we do a keyword search for "adult stem cell", in which case GSE6015 will not be among the results. Embryonic fibroblast cells, which are mentioned in the experiment text, are in fact adult stem cells. Thus, from a user perspective, this GSE might be a relevant result when searching for adult stem cells. Consider the fragment of an ontology below, which models the adult stem cell relationship:

```
isa(stem_cell,cell).
isa(adult(stem_cell),stem_cell).
isa(embryonic(fibroblast_cell),adult(stem_cell)).
```

Entities and properties are identified in the textual descriptions of the experiment and extracted as Prolog facts. For instance, in the sentence "Embryonic fibroplast cell" from the GSE6015 experiment, the compound noun results in an `entity` fact and the adjective in a `property` fact:

```
entity('GSE6015',1,fibroblast_cell). property('GSE6015',1,embryonic).
```

Using a simple Prolog rule it is possible to compose more complex entities from the entities and properties of the textual descriptions:

```
composed_entity(G,Id,Combined) :-
    entity(G,Id,Name),
    property(G,Id,Property),
    Combined =.. [ Property, Name ].
```

This rule finds an entity (e.g. `fibroplast_cell`) and associated property (e.g. `embryonic`) and combines them using the Prolog `=..` operator to form a composed entity term, `embronic(fibroplast_cell)`, suitable for matching the ontology. Consider for instance a sample query for "adult stem cells":

```
query(G):- isa(X,adult(stem_cell)),(entity(G,_,X);composed_entity(G,_,X)).
```

Running the `query(G)` goal will then produce the result, `G='GSE6015'`. The ontology is used to specialize the more general query "adult stem cell" to the more specific text "embryonic fibroblast cell" actually occurring in the textual description of the experiment. This example illustrates just one form of query expansion. For a thorough treatment of ontological query expansion see [6].

3 Discussion

We presented a sketch of methodology for performing ontology based search in the GEO database and demonstrated how such a search may be capable of retrieving information that eludes ordinary keyword based search. Our method is in an early stage where the ideas are still being refined and there are a number of challenges to be solved.

We are investigating methods for the automation of extraction of entities and properties from the text, which present challenges such as ambiguity, inconsistent naming and anaphora. The matching of a query to a document relies on assumptions about the implied meaning of extracted relations and in the ontology. This approach is not semantically well-founded, but pragmatically useful since it enables users to retrieve the otherwise elucidating relevant experiments. However, traditional information retrieval evaluation measures are difficult to apply due to the lack of a golden standard. Instead, we plan to do a qualitative evaluation and continuous improvement in cooperation with biologists.

An important issue of the functionality of semantic search in microarray-corpora is that the textual, although usually in line with MIAME, is of differing qualities. Some, like the case of GSE6015, have even filled out most of the blanks with a lot of information, whereas others, as in GSE1310 (another embryonic stem cell study), is lacking a lot of information. It is written according to the MIAME standard, although very sparsely so. Sparse annotations is a complication, inherent to much of the microarray-corpora, where an ontology based approach may be advantageous since it can provide a means to bridge the query to the experiment through ontological inference.

The design of the ontology is a demanding task and even with a large effort it may not be elaborate enough to cover all relevant cases. As an alternative it may be interesting to explore a vector-space model [7].

Since it is naive to believe that the annotations will be complete in the future, we find that ontology-based search for microarray data can be useful and is an area that needs more investigating.

References

1. Churchill, G.A.: Fundamentals of experimental design for cdna microarrays. Nat Genet (Dec 2002)
2. Brazma A. *et al.*: Minimum information about a microarray experiment (miame)-toward standards for microarray data. Nat Genet (Dec 2001)
3. Whetzel, P. L. *et al.*: The MGED Ontology: a resource for semantics-based description of microarray experiments). Bioinformatics (Apr 2006)
4. Andreasen, T., Bulskov, H. In: On Deriving Data Summarization through Ontologies to Meet User Preferences). Roskilde University (Feb 2009) 71–75
5. Andreasen, T. *et al.*: Siabo: Semantic information access through biomedical ontologies. In: KEOD. (2009)
6. Bhogal, J., Macfarlane, A., Smith, P.: A review of ontology based query expansion. Inf. Process. Manage. **43**(4) (2007) 866–886
7. Castells, P., Fernandez, M., Vallet, D.: An adaptation of the vector-space model for ontology-based information retrieval. IEEE Trans. on Knowl. and Data Eng. **19**(2) (2007) 261–272

Qualia and Property extraction from Italian Prepositional Phrases

Fabio Celli

CLIC-CIMeC, University of Trento

Abstract. This paper addresses the problem of meaning composition in Italian Prepositional Phrases. Exploiting prepositional phrase categories provided by Italian traditional grammar I manually annotated 1000 Prepositional Phrase instances and then I mapped the categories into Pustejovsky's qualia structures and Wu-Barsalou's property classes. The results in a classification task are F1 = 0.731 with prepositional phrase categories, F1 = 0.797 with qualia structures and F1 = 0.886 with property classes. Property classes showed to be good in distinghuising between entity and situational properties of nouns simply using prepositions as features.

1 Introduction and Related work

There has been an effort in recent years towards the exploitation of prepositions in semantic information extraction (see for example O'Hara and Wiebe 2008 [3]). A huge problem in this area is preposition ambiguity, that makes hard to use prepositions as features in computational linguistics. Girju [2] used Prepositional Phrases, translated in different languages, in a semantic relation extraction task and she demonstrated that her strategy helps in rising the performance of the classifier. In order to improve the possibility of comparison between prepositions in different languages I explore here the semantics of Italian Prepositional Phrases using two different theoretical frameworks: Pustejovsky's qualia structures ([6]) and Wu and Barsalou's [10] classes of generated properties. Qualia are well-known and powerful descriptions for lexical semantics and many tried to extract them automatically, for example Cimiano and Wenderoth ([1]) among others. For instance qualia structures are: **Formal role** (shape, position..); **Constitutive role** (material, parts..); **Telic role** (purpose, function..) and **Agentive role** (creator, causal chain..). The classes of generated properties I chose are the five higher level categories from Wu and Barsalou's proposal. To the best of my knowledge, nobody tried to use them for a classification task with Italian Prepositional Phrases before. The five classes I mentioned are: **Taxonomical Properties** (synonym, category, superordinate..); **Entity Properties** (associations, components, quality..); **Situational Properties** (location, manner, participant..); **Introspective Properties** (evaluation, cognitive operation, negation..) and **Miscellaneous Properties** (meta-comment, repetition). In the next section there is the description of the experiments and then in section 3 a discussion follows, section 4 is dedicated to conclusions.

2 Experiments

Corpus selection and annotation I sampled 1000 Italian Prepositional Phrases from CORISsmall, a reduced version of CORIS, a 100M-word, balanced corpus of written Italian. I first run a supersense tagger (Picca et Al 2008 [4]) on the dataset in order to obtain semantic classes for nouns in the prepositional phrases (they are: animal, person, plant, body, group, attribute, quantity, shape, possession, substance, artifact, food, object, location, time, act, process, phenomenon, event, relation, cognition, communication, feeling, motive, state, emotion). Then I manually annotated each prepositional phrase with the categories provided by Italian traditional grammar (see Prepositional Phrase categories in table 1), that are a mixture of semantic relations (predication, matter, location ..) and semantic roles (agent, beneficiary ..). In the annotation the observed agreement (Scott 1955, [8]) between two native speakers, one with and one without a background in theoretical linguistics, is $\pi = 0.8$.

Experiments I run all experiments in Weka (Witten-frank 2000 [9]) using decision trees (Quinlan 1993 [7]) and support vector machines with polynomial kernel (Platt 1998 [5]) as algorithms, a 66% training-33% testing split and prepositions and noun supersenses as features. In experiment 1 I put Prepositional Phrase classes as target of the classification and results are F1 = 0.722 with decision trees and F1 = 0.731 with support vectors. Although this result is good I think that sparseness cuts down the performance of the classifier both with decision trees and support vector machines, since there are many classes, some of which obtained very good performances (eg. specification, accompainment), but most of them have very few instances, thus yielding poor performances (eg. quality, mean, manner, purpose, distribution..). This result has been taken as the baseline for experiments 2 and 3, reported below. In order to reduce sparseness I mapped the prepositional phrase categories into Pustejovsky's Qualia structures and classes of generated Properties by Wu and Barsalou described above. The agreement on the mapping has been calculated between two italian native speakers with a background in linguistics. It is $\pi = 0.904$ from Prepositional Phrase types to Wu-Barsalou's categories and $\pi = 0.619$ from Prepositional Phrase types to Pustejovsky's qualia. The mapping is reported in detail in table 1.

Then I run experiment 2 with qualia and experiment 3 with property classes as targets of the classification respectively. In both the experiments the settings are the same described above. For the qualia classification task a class "NA" has been added in order to account for missing data (see table 1). Results are shown in table 2.

3 Discussion

Both Wu and Barsalou's classes and Pustejovsky's qualia outperform the baseline reported in experiment 1. Wu and Barsalou's properties yield exactly the same performance both with trees and support vector machines. Looking at the

PP category	Properties	Qualia	Semantics
specification	Entity	NA	x Of y
quality	Entity	constitutive	x HasQuality y
topic	Situation	NA	x About y
limitation	Situation	NA	x LimitedTo y
relation	Situation	NA	x InRelationWith y
time	Situation	formal	x AtTime y
mean	Situation	telic	x ByMeansOf y
manner	Situation	constitutive	x MannerOf y
matter	Entity	constitutive	x MadeOf y
quantity	Entity	formal	x HasQuantity y
comparison	Introspective	formal	x ComparedTo y
purpose	Situation	telic	x HasPurpose y
beneficiary	Situation	telic	x To y
location	Situation	formal	x LocationAt y
predication	Category	NA	x IsA y
accompainment	Situation	NA	x With y
accusation	Introspective	agentive	x GuiltyOf y
distribution	Situation	NA	x DividedIn y
denomination	Category	constitutive	x Named y
source	Situation	agentive	x From y
agent	Situation	agentive	x CauseOf y

Table 1. mappings.

results of experiments 2 and 3 (table 2) we can notice that qualia and properties catch different aspects of Prepositional Phrase semantics. On the one hand Qualia show a good classification performance only for the "NA" class, which is by far the most populous one, and for the formal role, that is easily separable from the "NA" class, unlike the telic and consitutive roles. The agentive role seems to have too few instances (22) for extracting good rules. On the other hand Wu-Barsalou's classes allow us to distinguish between entity and situational properties very well. The rules generated by the decision trees for qualia and properties show that qualia generate a large tree considering prepositions first and then noun supersenses, while property classes use only prepositions as features for the classification task revealing that *di* (of) is associated to entity properties, while all the other monosyllabic prepositions, namely *a* (to), *da* (from), *in* (in), *con* (with), *su* (on), *per* (for) and *tra/fra* (within), are associated to situational properties. We note that also support vectors make use only of prepositions for the classifiaction: this explains why the performance is the same both with decision trees and support vector machines.

4 Conclusions

In this paper I showed that it is possible to classify automatically situational and entity properties in Italian Prepositional Phrases directly from raw text

classes	**F1** (supprot vectors)	**F1** (decision trees)
formal	0.735	0.795
constitutive	0.294	0.296
agentive	0.333	0
telic	0.27	0.256
NA	0.905	0.906
avg	**0.796**	**0.797**
category	0	0
entity	0.932	0.932
situation	0.785	0.785
introspective	0	0
miscellaneous	0	0
avg	**0.886**	**0.886**

Table 2. Results of experiments 2 and 3.

by exploiting just prepositions. In the future it would be interesting to repeat the experiments with prepositional phrases in other languages in order to test whether or not prepositions introduce specific noun properties as they seem to do in Italian.

References

1. Cimiano, P., Wenderoth, J., 2005 Learning Qualia Structures from the Web, *In: Proceedings of the ACL Workshop on Deep Lexical Acquisition*, pp. 28-37.
2. Girju, R. 2009. The syntax and semantics of prepositions in the task of automatic interpretation of nominal phrases and compounds: A cross-linguistic study. *in Computational Linguistics 35(2) - Special Issue on Prepositions in Applications*, A. Villavicencio, V. Kordoni, and T. Baldwin (eds.)
3. O'Hara, T., Wiebe, J. 2008. Exploiting Semantic Role Resources for Preposition Disambiguation *In Computational Linguistics*, 35(2), pp. 151-184.
4. Picca, D., A. M. Gliozzo, and M. Ciaramita 2008. Supersense Tagger for Italian. In *In Proceedings of LREC.*
5. Platt, J. 1998. Machines using Sequential Minimal Optimization. In Schoelkopf, B., Burges, C., Smola, A. (ed), *Advances in Kernel Methods, Support Vector Learning.*
6. Pustejovsky, J. 1991. The generative lexicon. In *Computational Linguistics* 17(4), pp.209-441.
7. Quinlan, R. 1993. C4.5: Programs for Machine Learning. *Morgan Kaufmann Publishers*, San Mateo, CA.
8. Scott, W. A. 1955. Reliability of content analysis: The case of nominal scale coding. In *Public Opinion Quarterly* 19(3) pp.321-325.
9. Witten, I, H., Frank, E. 2000 Data *Mining. Practical Machine Learning Tools and Techniques with Java implementations.* Morgan and Kaufman, San Francisco, CA.
10. Wu, L.L, Barsalou, L.W. 2009. Perceptual simulation in conceptual combination: Evidence from property generation. In *Acta Psychologica*, 132, 173-189.

4

Reviewers

Adam Lopez (University of Edinbourg)

Adam Wyner (University of Liverpool)

Alessandro Lenci (University of Pisa)

Andrei Popescu-Belis (Idiap Research Institute)

Andrew Trotman (University of Otago)

Andy Way (Dublin City University)

Angelo Dalli (University of Malta)

Anna Feldman (Montclair state university)

Anna Kupsc (University of Bordeaux, Polish Academy of Sciences)

Anne Abeille (University of Paris 7)

Anotnis Bikakis (Foundation for Research and Technology - Hellas)

Barbora Hladka (Charles University)

Bruno Mery (University of Bordeaux, INRIA, LaBRI)

Bryan Renne (University of Groningen)

Caroline Sporleder (Saarland University)

Chiara Ghidini (University of Trento)

Chris Kennedy (University of Chicago)

Daniel Marcu (University of Southern California)

Daniele Mundici (University of Florence)

Dario Colazzo (University of Paris South)

David Devault (University of Southern California)

David Ripley (Institut Jean Nicod at the ENS in Paris)

4. REVIEWERS

David Vallet (Autonomous University of Madrid)
Diego Calvanese (Free University of Bozen-Bolzano)
Dirk Walther (Technical University of Madrid)
E.Allyn Smith (Ohio State University)
Emiliano Lorini (IRIT)
Eric Pacuit (Tilburg University)
Floris Roelofsen (Umass Amherst)
Francesca Poggiolesi (Free University Brussels)
Frank van Eynde (University of Leuven)
Frank Veltman (ILLC University of Amsterdam)
Gemma Boleda (University of Catalunya)
Gerhard Brewka (University of Leipzig)
Gerhard Jaeger (University of T übingen)
Greg Kobele (University of Chicago)
Guillaume Aucher (University of Luxembourg)
Hans Guesgen (Massey University)
Hans-Martin Gaertner (Center of Linguistics, Berlin)
Hans-Ulrich Krieger (Saarland University DFKI)
Harry Bunt (University of Tilburg)
Inderjeet Mani (The MITRE Corporation, Bedford, MA)
Ingo Gl ockner (Fern University in Hagen)
Ivana Kruijff-Korbayova (Saarland University)
Jan Reimann (University of California, Berkeley)
Jan van Eijck (Uil-OTS Utrecht University)
Jens Ulrik Hansen (Roskilde University)
Jirka Hana (Charles University)
Johan Bos (University of Rome)
John Nerbonne (University of Groningen)
Kevin Duh (University of Washington)
Kristoffer Arnsfelt Hansen (Aarhus University)
Larry Moss (Indiana University)
Lionel Clment (University of Bordeaux, LaBRI)
Makoto Kanazawa (National Institute of Informatics Tokyo)

Maurice Margenstern (University of Metz)
Melvin Fitting (City University New York)
Michael Covington (University of Georgia)
Morten Rhiger (Roskilde University)
Olga Borik Authonomous University of Barcelona
Paul Dekker (University of Amsterdam)
Paul Egre (CNRS, Institut Jean-Nicod, Paris)
Peter Pagin (The Swedish Collegium for Advanced Studies in Social Science)
Petr Hajek (Institute of Computer Science AS CR)
Philipp Cimiano (University of Bielefeld)
Raffaella Bernardi (Free University of Bozen-Bolzano)
Reinhard Muskens (Tilburg University)
Roman Kontchakov (Birkbeck College London)
Ron Arstein (University of Southern California)
Roxana Girju (University of Illinois at Urbana-Champaign)
Scott Piao (Lancaster University)
Sebastian Sequoiah-Grayson (University of Leuven)
Stefan Evert (University of Osnabrck)
Sujata Ghosh (University of Groningen)
Tadeusz Litak (University of Leicester)
Thomas F. Gordon (FOKUS Berlin/University of Potsdam)
Tim van de Cruy (University of Groningen)
Timo Baumann (University of Potsdam)
Tomohiro Hoshi (Stanford University)
Trevor Bench-Capon (University of Liverpool)
Viktor Pekar (University of Wolverhampton)
Violaine Prince (Montpellier University 2)
Yiannis Moschovakis (University of California, University of Athens)
Yves Bestgen (Belgian national fund for Scientific Research)
Yvette Yannick Mathieu (University of Paris 7)