Preface

These proceedings contain the papers presented at the Student Session of the 27th European Summer School in Logic, Language and Information (ESSLLI 2015), taking place at the Universitat Pompeu Fabra in Barcelona from August 3rd to 14th, 2015. This is a truly special edition of the ESSLLI Student Session as it is the twentieth anniversary of the Student Session. Its enduring success comes from the fact that it offers an excellent venue for students to present their work on a diverse range of topics at the crossroads of logic, language and information and to receive feedback on their work from renowned experts in their respective fields. We are delighted that the Student Session has this year attracted submissions from almost all continents and has thereby become a veritable international event whose scope reaches far beyond Europe. As in previous years, we have received many high-quality submissions which made it very hard for us and the reviewers to make our acceptance decisions. We received 77 submissions in total, 46 of which were submitted for oral presentation, and 31 of which were submitted for poster presentation. At the Student Session, 16 of these submissions were presented as talks and 8 submissions were presented in the form of a poster. This makes for an acceptance rate of 35% for talks and 26% for posters and an overall acceptance rate of 31%.

We would like to thank each of the co-chairs, as well as the area experts, for all their invaluable help in the reviewing process and organization of the Student Session. We would also like to thank the ESSLLI Organizing Committee in Barcelona, for organizing the entire summer school, and catering to all our needs. They have been wonderful hosts to the Student Session and we certainly enjoyed working together with them. We are also greatly indebted to Ronald de Haan, the chair of the Student Session 2014, as he let us use many of the materials from the previous year. We would of course also like to take this opportunity to thank our main sponsor, the Springer Verlag. Springer have provided the prizes for the best poster and talk in the past and have continued their generous support this year, as well.

Naturally, an event like the Student Session would not be possible if it wasn’t for the people who put a lot of effort into their academic work and decided to submit it to us. Our thanks go to all those who submitted papers. We encourage them to keep doing such excellent work and hope that we have provided an appropriate platform for the presentation for their ideas.

August 2015

Miriam Kaeshammer & Philip Schulz
Chairs of the ESSLLI 2015 Student Session
Organization

Program Committee

Chairs
Miriam Kaeshammer  Universität Düsseldorf
Philip Schulz  Universiteit van Amsterdam

Logic & Computation co-chairs
Giovanni Cina  Universiteit van Amsterdam
Zeynep Saribatur  Technische Universität Wien

Logic & Language co-chairs
Marisa Köllner  Universität Tübingen
Veronika Richtakikova  Universitat Pompeu Fabra Barcelona

Language & Computation co-chairs
Angeliki Lazaridou  University of Trento
Maria Nadejde  University of Edinburgh

Area experts

Logic & Computation
Thomas Icard  Stanford University

Logic & Language
Yusuke Kubota  University of Tsukuba/Ohio State University

Language & Computation
Shravan Vasishth  University of Potsdam
Table of Contents

Logic & Computation

<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractional Isomorphisms of Kripke Structures</td>
<td>Julian Bitterlich</td>
<td>1</td>
</tr>
<tr>
<td>Performing conditional strategies in strategic STIT theory</td>
<td>Hein Duijf</td>
<td>13</td>
</tr>
<tr>
<td>Sufficient (In)completeness for Definitions in Hierarchic Superposition</td>
<td>Adrián Rebola-Pardo</td>
<td>25</td>
</tr>
<tr>
<td>Implicit Conflicts in Abstract Argumentation</td>
<td>Christof Spanring</td>
<td>37</td>
</tr>
</tbody>
</table>

Logic & Language

<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>How flexible are categorical models of meaning?</td>
<td>Antonin Delpeuch</td>
<td>45</td>
</tr>
<tr>
<td>Minimizing Scope Ambiguity Hypothesis</td>
<td>Milica Denić</td>
<td>54</td>
</tr>
<tr>
<td>Negative Coordination in French</td>
<td>Aurore González</td>
<td>66</td>
</tr>
<tr>
<td>Experimental investigations of probability expressions</td>
<td>Michele Herbstritt</td>
<td>77</td>
</tr>
<tr>
<td>On the Semantics and Pragmatic Underpinning of Essentially Plural Nouns</td>
<td>Yuki Ito</td>
<td>89</td>
</tr>
<tr>
<td>A Family of Exclusives in Hebrew</td>
<td>Dina Orenstein</td>
<td>96</td>
</tr>
<tr>
<td>Changes in the Distributional Patterns of the Spanish Copulas: the Influence of Contextual Modulation</td>
<td>Sara Sánchez-Alonso</td>
<td>107</td>
</tr>
<tr>
<td>A Formal Semantics of the Final Rise</td>
<td>Julian J. Schlöder</td>
<td>117</td>
</tr>
<tr>
<td>Latin Long-Distance Reflexives in Messenger Reports</td>
<td>Per Erik Solberg</td>
<td>128</td>
</tr>
</tbody>
</table>
Ex Post “Cheap Talk”: Value of Information and Value of Signals  
Liping Tang  

Chisholm’s Paradox and Dynamic Kratzer Semantics for Deontic ‘Ought’s  
Yuna Won  

Durative Phrases in Constructions Containing Double Identical Verbs in Mandarin Chinese  
Xiaoqian Zhang  

Language & Computation  
A Study of Semi-Supervised Spanish Verb Sense Disambiguation  
Cristian Cardellino  

Turkish Children’s Acquisition of Nouns and Verbs: A Developmental Study  
Fatma Iskin  

Distributional Gastronomics  
Jonathan Oberlander, Laura Bostan  

Gender Identification for Adversarial Writing  
Yunita Sari  

An Event-Based Ontology for Sign Language Phonology  
Ayça Muge Sevinç  

Understanding belief bias by measuring prior beliefs for a Bayesian model of syllogistic reasoning  
Michael Henry Tessler  

Automatic alignment using Pair Hidden Markov Models  
Johannes Wahle
Fractional Isomorphisms of Kripke Structures

Julian Bitterlich

TU-Darmstadt

bitterlich@mathematik.tu-darmstadt.de

Abstract. We generalise fractional isomorphisms from graphs to Kripke structures and also define a variant of the colour refinement algorithm for Kripke structures. We show that the discriminating power of these two notions is captured by universal, two-way counting-bisimulation equivalence. This forms also the basis for a new perspective on the relationship between Sherali-Adams relaxations of graph isomorphism and bijective $k$-pebble game equivalences. Our analysis here draws on the natural translation between Ehrenfeucht-Fra"issé style back&forth systems and bisimulations in the corresponding game graphs whose nodes are the observable configurations in the Ehrenfeucht-Fra"issé game.

1 Introduction

Two graphs $A$ and $B$ with finite sets of vertices $A$ and $B$ and adjacency matrices $A \in \{0,1\}^{A,A}$ and $B \in \{0,1\}^{B,B}$ are isomorphic if, and only if, the system

$$\text{ISO}(A,B) : \begin{align*}
A X &= X B \\
1 &= \sum_{a' \in A} X_{a'b} = \sum_{b' \in B} X_{ab'} \text{ for all } a \in A, b \in B \\
0 &\leq X_{ab} \text{ for all } a \in A, b \in B
\end{align*}$$

has an integer solution $X \in \mathbb{Z}^{A,B}$. The linearisation of this integer programming problem leads to the notion of fractional isomorphisms: Two simple graphs $A$ and $B$ are fractionally isomorphic if $\text{ISO}(A,B)$ has a real solution $X \in \mathbb{R}^{A,B}$.

Tinhofer [12] and Ramana, Scheinermann and Ullmann [9] showed that two graphs are fractionally isomorphic if, and only if, they are indistinguishable by the colour refinement algorithm which, by Immerman and Lander [6], has the same discriminating power as $C^2$, the 2-variable fragment of first-order logic with counting quantifiers.

Does this equivalence of fractional isomorphisms and $C^2$-equivalence still hold true if we look at fractional isomorphisms of graphs without any restriction for the edges, i.e., directed graphs? A straightforward generalisation using an unaltered version of $\text{ISO}(A,B)$, does not result in a symmetric relation (cf. Fig. 1a). To remedy this asymmetry it is natural to add the equation $BX^t = X^tA$ to
the equations defining a fractional isomorphism but even then fractional isomorphisms are too weak to capture $C^2$-equivalence (cf. Fig. 1b)). In the first part of this paper we show that universal, two-way modal logic with counting, $ML_{\forall,-,C}^\nu$, exactly corresponds to fractional isomorphisms on graphs. We even generalise this to Kripke structures and get the equivalence of fractional isomorphisms, indistinguishably by a suitable variant of the colour refinement algorithm and $ML_{\forall,-,C}^\nu$-equivalence.

The colour refinement algorithm can be generalised to the $k$-dimensional Weisfeiler-Lehman algorithm ($k$-WL) for each $k \in \mathbb{N}$. Immerman and Lander [6] showed that this corresponds to $C^k$-equivalence. Fractional isomorphisms can also be generalised by more sophisticated linearisation methods. Applying the Sherali-Adams relaxation [11] technique to ISO results in a family of linear programs $Iso^k$. Atserias and Maneva [1] and Malkin [7] related these relaxations to $k$-WL in an interleaving way, i.e., for $k \geq 2$, the solvability of $Iso^k$ implies indistinguishability by $k$-WL and indistinguishability by $k$-WL implies the solvability of $Iso^{k-1}$. Grohe and Otto [4] showed that this interleaving is strict. However, saying that the solvability corresponds “more or less” to $C^k$ we get the following analogy with the previous discussion about Kripke structures:

<table>
<thead>
<tr>
<th>Kripke Structures</th>
<th>Directed Graphs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ML_{\forall,-,C}^\nu$</td>
<td>$C^k$</td>
</tr>
<tr>
<td>colour refinement algorithm</td>
<td>$k$-WL</td>
</tr>
<tr>
<td>fractional isomorphisms</td>
<td>$\approx$ solvability of $Iso^k$</td>
</tr>
</tbody>
</table>

The second part of this paper concerns the analysis of this parallelism. We uncover these parallels as manifestations of a general translation of first-order logic to modal logic as described in [8]. The main idea of this translation is that we can see the well-known Ehrenfeucht-Fraissé games as modal games on the transition systems of observable configurations. Using this point of view we can easily generalise the colour refinement algorithm which results in the $k$-WL, thus establishing the connection of $C^k$ and $k$-WL as another manifestation of the connection of $ML_{\forall,-,C}^\nu$ and the colour refinement algorithm. The same approach works for fractional isomorphisms and their generalisations which are tightly bound (in the fashion of the results of Malkin and Atserias and Maneva) to $Iso^k$. Thus we establish a conceptual different view of the result of Atserias, Malkin and Maneva.

1 In [1] this equivalence is wrongly acclaimed to be true.
2 Note that our definition of fractional isomorphism in Definition 6 differs from the definition given here, in this introduction, but Corollary 1 shows that both versions have the same discriminating power.
3 Actually what we call the $k$-dimensional Weisfeiler-Lehman algorithm they called $(k+1)$-dimensional Weisfeiler-Lehman algorithm.
2 Back&Forth Systems

We are not going to introduce the logics $C^k$ and $ML$ but only their corresponding b&f systems. For introductions see [2,3,10].

**Definition 1.** Let $A$ and $B$ be Kripke structures over the modal signature $\tau = \{(R_\alpha)_{\alpha \in \Gamma}, (P_\theta)_{\theta \in I}\}$ on the domains $A$ and $B$. A non-empty relation $Z \subseteq A \times B$ is a counting-bisimulation between $A$ and $B$, $Z : A \sim_C B$, if $Z$ only relates propositionally equivalent elements, i.e., $a \in P^A_\theta$ iff $b \in P^B_\theta$ for all $(a,b) \in Z$ and proposition $\theta \in I$, and $Z$ satisfies the counting b&f conditions w.r.t. $R_\alpha$ for every modality $\alpha \in \Gamma$:

- **(counting forth)** for any $(a,b) \in Z$ and distinct $a_1, \ldots, a_n \in A$ s.t. $(a,a_i) \in R^A_\alpha$, there are distinct $b_1, \ldots, b_n \in B$ s.t. $(a_i,b_i) \in Z$ and $(b_i) \in R^B_\alpha$.
- **(counting back)** for any $(a,b) \in Z$ and distinct $b_1, \ldots, b_n \in B$ s.t. $(b,b_i) \in Z$ and $(a,a_i) \in R^A_\alpha$, there are distinct $a_1, \ldots, a_n \in A$ s.t. $(a_i,b_i) \in Z$ and $(a,a_i) \in R^A_\alpha$.

$Z$ is global, $Z : A \sim_{\sim_C} B$ if it also satisfies the counting b&f condition w.r.t. the global accessibilities $A \times A$ and $B \times B$. $Z$ is two-way, $Z : A \sim_{\sim_C} B$ if it also satisfies the counting b&f conditions w.r.t. the inverses of $R_\alpha$.

Over the class of finite Kripke structures the counting b&f conditions for $Z$ w.r.t. a relation $R_\alpha$ are equivalent to

$$(\dagger) \text{ for all } (a,b) \in Z \text{ there is a matching in } Z \text{ between } N(a) \text{ and } N(b).$$

Here $N(a)$ and $N(b)$ are the neighbourhoods of $a$ respectively $b$ w.r.t. $R_\alpha$, i.e., $N(a) := \{a' : (a,a') \in R^A_\alpha\}$ and $N(b) := \{b' : (b,b') \in R^B_\alpha\}$. The condition $(\dagger)$ used on the global accessibilities $A \times A$ and $B \times B$ implies that a counting-bisimulation between finite Kripke structures is global if, and only if, it contains a matching between the domains of those structures.

Let $\sigma$ be a purely relational signature and $\mathfrak{A}$ and $\mathfrak{B}$ two $\sigma$-structures on domains $A$ and $B$. $\text{Part}_k(\mathfrak{A}, \mathfrak{B})$ is the set of those pairs of tuples $(a,b) \in A^k \times B^k$ such that $\{(a_i,b_i) : i \in [k]\}$ is a partial isomorphism; we write $a \leftrightarrow b$ for tuples $(a,b)$ in $\text{Part}_k(\mathfrak{A}, \mathfrak{B})$ and $a_i \leftrightarrow b_i$ for $(a_1,\ldots,a_{i-1},a,a_{i+1},\ldots,a_k)$.

**Definition 2.** Let $\mathfrak{A}$ and $\mathfrak{B}$ be two $\sigma$-structures. A non-empty $Z \subseteq \text{Part}_k(\mathfrak{A}, \mathfrak{B})$ is a counting $k$-pebble b&f system between $\mathfrak{A}$ and $\mathfrak{B}$ if for all $a \leftrightarrow b \in Z$, $i \in [k]$.
(counting forth) for any distinct $a'_1, \ldots, a'_n \in A$, there are distinct $b'_1, \ldots, b'_n \in B$ s.t. $a' \mapsto b'_j \in Z$ for all $j \in [n]$. (counting back) for any distinct $b'_1, \ldots, b'_n \in B$, there are distinct $a'_1, \ldots, a'_n \in A$ s.t. $a'_i \mapsto b'_j \in Z$ for all $j \in [n]$.

**Definition 3.** Let $A$ be a $\sigma$-structure on the domain $A$ and $k \in \mathbb{N}$. The $k$-cube $A$ of $A$ is a Kripke structure defined as follows: its signature is $\tau^k(\sigma)$ that consists of binary relations $R_i$ for $i \in [k]$ and unary predicates $P_\theta$ for atomic $\sigma$-types $\theta$ in the $k$-variable fragment of first-order logic. The domain of $A$ is $A^k$ and the interpretations of the relations in $\tau$ are $P_\theta^A := \{a \in A^k : a_i = \theta\}$ and $R_i^A := \{(a,a') : a_j = a'_j \text{ for } j \neq i\}$.

Note that $\tau^k(\sigma)$ is finite if $\sigma$ is finite and that two tuples $a, a' \in A^k$ have the same atomic type if, and only if, $(a, a') \in \text{Part}_k(A, B)$. Furthermore, the underlying frame of a $k$-cube is an multi-$\mathcal{SO}$ frame of diameter $k$, i.e., all accessibilities are equivalence relations and any two elements in the domain are linked by a path of length at most $k$. The following lemma is most important for us to lift results on modal logic to first-order logic.

**Lemma 1.** Let $A$ and $B$ be two finite $\sigma$-structures and $A$ and $B$ their $k$-cubes. Then any $Z \subseteq A^k \times B^k$ is a counting $k$-pebble $b\ell f$ system between $A$ and $B$ if, and only if, it is a (global, two-way) counting-bisimulation between $A$ and $B$.

**Proof.** $Z \subseteq \text{Part}_k(A, B)$ is equivalent to all $(a, b) \in Z$ being atom-equivalent in $A$ and $B$. The Ehrenfeucht-Fra"{i}ssé style counting forth-condition over $A$ and $B$ is equivalent to the bisimulation forth-condition over $A$ and $B$. Similarly for the counting back-conditions. It remains to show that any counting-bisimulation $Z$ between $A$ and $B$ is two-way and global. $Z$ is two-way, since all accessibilities are symmetric. For globality of $Z$ take any $(a, b) \in Z$. Then there is matching in $Z$ between the sets $A_\ell := \{a' \in A^k : a'_i = a_i \text{ for } i > \ell\}$ and $B_\ell := \{b' \in B^k : b'_i = b_i \text{ for } i > \ell\}$ for every $\ell \in [k]$. For $\ell = 0$ this is clear. A matching for $A_{\ell+1}$ and $B_{\ell+1}$ can be obtained from a matching of a $A_\ell$ and $B_\ell$ by applying $(\ell)$ to $R^A_{\ell+1}$ and $R^B_{\ell+1}$ for every matched pair of $A_\ell \times B_\ell$. Note that the $R^A_{\ell+1}$ neighbourhoods for any distinct $a'_1, a'_2 \in A_\ell$, $b'_1, b'_2 \in B_\ell$ are disjoint. \hfill $\Box$

### 3 The Colour Refinement Algorithm

We introduce the the colour refinement algorithm and the notion of balanced, stable partitions. In this section we only consider finite graphs. A graph is a Kripke structure over the signature $\tau_G = \{R\}$. For a graph $A$, a node $a \in A$ and subset $p \subseteq A$ denote the outdegree of a w.r.t. $p$ by $\nu(a,p) := |\{(a,a') \in R^A : a' \in p\}|$ and the indegree of a w.r.t. $p$ by $\nu(p,a) := |\{(a',a) \in R^A : a' \in p\}|$.

**Definition 4.** Let $A = (A, R^A)$ and $B = (B, R^B)$ be graphs. A common partition $P$ is a partition of $A \cup B$. Furthermore, $P$ is balanced if $|p_A| = |p_B|$ for all $p \in P$ and $P$ is stable if, for all $a, b$ in the same class\(^4\) and classes $p \in P$ we have $\nu(a,p) = \nu(b,p)$ and $\nu(p,a) = \nu(p,b)$.

\(^4\) We call the elements of a partition class.
3.1 Description of the Colour Refinement Algorithm

Given two graphs \( A = (A, R_A) \) and \( B = (B, R_B) \), the colour refinement algorithm computes a common partition by the following iterative procedure: the algorithm starts with the trivial partition with one class. Then, in each step, the partition \( P \) of the previous step is refined to \( P' \) via

\[
\begin{align*}
(a, b) \text{ are in the same class w.r.t. } P' & \text{ iff they are in the same class w.r.t. } P \\
& \text{ and } \nu(a, p) = \nu(b, p) \text{ and } \nu(p, a) = \nu(p, b) \text{ for all } p \in P.
\end{align*}
\]

In particular, after the first refinement, two vertices are in the same class iff they have the same indegree and outdegree. The algorithm stops if no further refinement is achieved; this happens after at most \(|A| + |B|\) many rounds. We say that the colour refinement algorithm does not distinguish \( A \) and \( B \) if the output partition is balanced.

3.2 Refinement vs. Bisimulations

The color refinement algorithm has the same distinguishing power as global, two-way counting-bisimulations.

**Proposition 1.** Two finite graphs are not distinguished by the colour refinement algorithm if, and only if, they are globally, two-way counting-bisimilar.

We use the balanced, stable partition as a mediator between the colour refinement algorithm and global, two-way counting-bisimulations. The proposition follows readily from Lemma 2 and Lemma 3.

**Lemma 2.** Two graphs are indistinguishable by the colour refinement algorithm if, and only if, there is a balanced, stable partition.

**Proof.** For the forward direction observe that the partition produced by the colour refinement algorithm is always stable. For the converse direction let \( P \) be a stable partition. Then it is also easy to prove that \( P \) is a refinement of all the partitions in the inductive definition of the output partition of the colour refinement algorithm. So, if \( P \) is also balanced, the output of the colour refinement algorithm is so too. \( \square \)

**Lemma 3.** Two graphs are globally, two-way, counting-bisimilar if, and only if, there is a balanced, stable partition.

**Proof.** Let \( A \) and \( B \) be graphs. Let \( P \) be a balanced, stable partition. Then it is easy to see that \( Z := \{p_A \times p_B : p \in P\} \) is a global, two-way counting-bisimulation. For the converse direction let \( Z \) be a global, two-way counting-bisimulation between \( A \) and \( B \). Consider the graph \( G = (A \cup B, \{(a, b) : (a, b) \in Z \text{ or } (b, a) \in Z\}) \). Again it is an easy exercise to show that the connected components of \( G \) form a balanced, stable partition. \( \square \)
4 Fractional Isomorphisms

As in the previous section we consider finite graphs. For a graph \( A = (A, R_A) \) we define its adjacency matrix \( A \in \{0, 1\}^{A,A} \) via \( A_{aa'} = 1 \iff (a, a') \in R_A \).

**Definition 5.** A matrix \( X \in \mathbb{R}^{A,B} \) is doubly stochastic if each entry of \( X \) is non-negative and the sum over each row as well as over each column of \( X \) is 1, i.e., \( \sum_{a' \in A} X_{ab} = \sum_{b' \in B} X_{a'b'} = 1 \) for all \( a \in A \) and \( b \in B \).

**Definition 6.** Let \( A = (A, R_A) \) and \( B = (B, R_B) \) be two graphs. \( X \in \mathbb{R}^{A,B} \) is a fractional isomorphism from \( A \) to \( B \) if \( X \in \mathbb{R}^{A,B} \) is doubly stochastic and \( A X \geq X B \) and \( B X^t \geq X^t A \), where the inequality has to be seen component-wise.

**Proposition 2.** Two finite graphs are fractionally isomorphic if, and only if, they admit a balanced, stable partition.

We get the following corollary to the proof.

**Corollary 1.** Let \( A \) and \( B \) be graphs on the domains \( A \) and \( B \). Then there is a doubly stochastic matrix \( X \in \mathbb{R}^{A,B} \) s.t. \( AX \geq XB \) and \( BX^t \geq X^t A \) if, and only if, there is doubly stochastic matrix \( X' \in \mathbb{R}^{A,B} \) s.t. \( AX' = XB \) and \( BX'^t = X'^t A \).

The rest of this section is devoted to the proof of Proposition 2.

**From Balanced, Stable Partitions to Fractional Isomorphisms.** Let \( A \) and \( B \) be graphs and \( P \) a balanced, stable partition. Define the matrix \( X \in \mathbb{R}^{A,B} \) by

\[
X_{ab} := \begin{cases} 
2/|p|, & \text{if there is a } p \in P \text{ with } a, b \in p \\
0, & \text{else.}
\end{cases}
\]

We leave it to the reader to show that \( X \) is doubly stochastic and that \( AX = XB \) as well as \( BX^t = X^t A \). Thus, in particular, \( X \) is a fractional isomorphism from \( A \) to \( B \). Observe that this also enables us to argue for the non-trivial direction of Corollary 1. By Proposition 2 we obtain from a fractional isomorphism from \( A \) to \( B \) a balanced, stable partition which in turn can be used to obtain the desired \( X' \) with \( AX' = XB \) and \( BX'^t = X'^t A \).

**From Fractional Isomorphisms to Balanced, Stable Partitions.** For a matrix \( X \in \mathbb{R}^{A,B} \) and subsets \( p_A \subseteq A, p_B \subseteq B \) we denote by \( X_{p_A p_B} \in \mathbb{R}^{p_A, p_B} \) the restriction of \( X \) to those subsets.

With any doubly stochastic matrix \( X \in \mathbb{R}^{A,B} \) we associate the undirected, bipartite graph \( G(X) := (A \cup B, \{(a, b) : X_{ab} \neq 0 \text{ or } X_{ba} \neq 0\}) \). The connected components of \( G(X) \) form a partition on \( A \cup B \), the partition induced by \( X \).

**Definition 7.** A doubly stochastic matrix \( X \in \mathbb{R}^{A,B} \) is irreducible if \( G(X) \) is connected.
Observation 1. Let $X$ be a fractional isomorphism from the graph $A$ to the graph $B$. Then the partition $P$ induced by $X$ is such that $X_{pA,pB}$ is an irreducible doubly stochastic matrix for all $p \in P$.

Lemma 4. Let $X \in \mathbb{R}^{A,B}$ be a doubly stochastic matrix. Then $|A| = |B|$.

Proof. $|A| = \sum_{a \in A} \sum_{b \in B} X_{ab} = \sum_{b \in B} \sum_{a \in A} X_{ab} = |B|$.

Lemma 5. Let $X \in \mathbb{R}^{A,B}$ be an irreducible, doubly stochastic matrix and $v \in \mathbb{R}^A$, $w \in \mathbb{R}^B$ vectors. If $v \geq Xw$ and $w \geq X^tv$, then all entries of $v$ and $w$ are equal, i.e., $v_a = v_b$ for all $a \in A, b \in B$. The same conclusion holds true if $v \leq Xw$ and $w \leq X^tv$.

Proof. We consider the graph $G(X)$ with a valuation $\ell : A \cup B \to \mathbb{R}$ on its nodes defined by $\ell(a) := v_a$ and $\ell(b) := w_b$ for $a \in A$ and $b \in B$. Let $a \in A \cup B$ s.t. $\ell(a)$ is minimal. If $a \in A$, then

$$\ell(a) \geq \sum_{b \in N(a)} \ell(b)X_{ab} \iff \sum_{b \in N(a)} (\ell(a) - \ell(b))X_{ab} \geq 0.$$ 

Thus $\ell(a) = \ell(b)$ since $X_{ab} > 0$ for $b \in N(a)$. Similarly, if $a \in B$ but using $w = X^tv$. So $\ell$ is constant over $G(X)$, since the graph is connected.

For the other implication apply the Lemma to $\hat{v} = -v$ and $\hat{w} = -w$.

Now we can show that the partition $P$ induced by a fractional isomorphism $X$ is balanced and stable. Balancedness follows from Observation 1 and Lemma 4. In order to establish stability of $P$ we show that $\nu(a,p') = \nu(b,p')$ for every $p' \in P$ and $a \in A, b \in B$ that are in the same class w.r.t. $P$. Towards this, for any given $p, p' \in P$, consider the vectors $v \in \mathbb{R}^A$ and $w \in \mathbb{R}^B$ given by

$$v_a := \nu(a, p') \quad \text{and} \quad w_b := \nu(b, p').$$

In order to apply Lemma 5, which yields the desired result, we have to show that $v \geq X_{pA,pB}w$ and $w \geq (X_{pA,pB})^tv$. So let $a \in p_A$, then

$$v_a = \sum_{a' \in p'_A} \mathbb{k}_{a,a'} = \sum_{a' \in p'_A} \sum_{b' \in p'_B} X_{a,a'} = \sum_{b' \in p'_B} \sum_{a' \in p'_A} \mathbb{k}_{a,a'}X_{a,a'} = \sum_{b' \in p'_B} (A^X)_{ab'}$$

$$\geq \sum_{b' \in p'_B} (X^B)_{ab'} = \sum_{b' \in p'_B} \sum_{b \in p_B} \mathbb{B}_{bb'}X_{ab} = \sum_{b \in p_B} X_{ab} \sum_{b' \in p'_B} \mathbb{B}_{bb'} = (X_{pA,pB}w)_a,$$

where we have used that $X$ solves $\text{Simu}[A,B]$. Similarly we get $w \geq (X_{pA,pB})^tv$. Thus, by Lemma 5, we have $\nu(a, p') = \nu(b, p')$ for all $a, b \in p$. We establish $\nu(p', a) = \nu(p', b)$ similarly. Thus $P$ is stable.

5 Generalisations to Kripke Structures

In the following we generalise the colour refinement algorithm and fractional isomorphisms to Kripke structures. As usual, all structures we consider are finite.
As a consequence of Corollary 1 we chose to define fractional isomorphisms for Kripke structures, in contrast to Definition 6, via matrix equality rather than inequality.

Let \( A \) be a Kripke \( \tau \)-structure with \( \tau = \{ (R_\alpha)_{\alpha \in \Gamma}, (P_\theta)_{\theta \in I} \} \). For each modality \( \alpha \in \Gamma \) define \( \nu_\alpha \) and the adjacency matrix \( A_\alpha \in \{0, 1\}^A \) in the straightforward ways: \( \nu_\alpha(a, p) := |\{ (a, a') \in R_\alpha^A : a' \in p \}| \) and \( (A_\alpha)_{a a'} = 1 :\iff (a, a') \in R_\alpha^A \).

We also define the characteristic vector \( a_\theta \in \{0, 1\}^A \) for each colour \( \theta \in I \) via \( (a_\theta)_a = 1 :\iff a \in P_\theta^A \).

**Definition 8.** The colour refinement algorithm for Kripke structures operates like the colour refinement algorithm for graphs, described in Sec. 3.1 with the following changes: the initial partition is given by propositional equivalence and the refinement condition (\( * \)) is

\[
\{ a, b \text{ are in the same class w.r.t. } P \} \text{ iff they are in the same class w.r.t. } P \text{ if and only if } \nu_\alpha(a, p) = \nu_\alpha(b, p), \nu_\alpha(p, a) = \nu_\alpha(p, b) \text{ for all } p \in P \text{ and } \alpha \in \Gamma.
\]

**Definition 9.** Let \( A \) and \( B \) be two Kripke \( \tau \)-structures. A doubly stochastic matrix \( X \in \mathbb{R}^A \times \mathbb{B}^A \) is a fractional isomorphism from \( A \) to \( B \) if

\[
\begin{align*}
A_\alpha X &= X B_\alpha & &\text{for all } \alpha \in \Gamma \\
\mathbb{B}_\alpha X^t &= X^t A_\alpha & &\text{for all } \theta \in I
\end{align*}
\]

**Lemma 6.** Two finite Kripke structures are globally, two-way counting-bisimilar if, and only if, they are indistinguishable by the colour refinement algorithm if, and only if, they are fractionally isomorphic.

**Proof.** We only prove the equivalence of fractional isomorphisms and global, two-way counting-bisimulations. The equivalence between the discriminating power of global, two-way counting-bisimulations and the colour refinement algorithm is proven analogously.

First, we treat the case without propositions. Then \( Z \) is a global, two-way counting-bisimulation between \( A \) and \( B \) iff \( Z : \mathcal{A}|_\alpha \sim_{\forall \gamma, - C} \mathcal{B}|_\alpha \) for all \( \alpha \)-reducts of \( A \) and \( B \). By the proof of Lemma 3, whose constructions do not depend directly on the underlying graph structure, this is equivalent to the existence of a single partition that is balanced and stable for all \( \alpha \)-reducts of \( A \) and \( B \) for all \( \alpha \in \Gamma \).

Finally, by the construction of Proposition 2, which again does not depend on the graph structure of \( \mathcal{A}|_\alpha \) and \( \mathcal{B}|_\alpha \), this is equivalent to the existence of a doubly stochastic matrix \( X \) s.t. \( A_\alpha X = X B_\alpha \) as well as \( \mathbb{B}_\alpha X^t = X^t A_\alpha \) for all \( \alpha \in \Gamma \).

We can treat propositions as special cases of modalities in the following way: pass from a Kripke \( \tau \)-structure \( A \) to a Kripke \( \tau' \)-structure \( A' = \{ A, (R_\alpha)_{\alpha \in \Gamma \cup J} \} \) where \( \tau' = \{ (R_\alpha)_{\alpha \in \Gamma \cup J} \} \) consists only of binary relation symbols. Their interpretations in \( A' \) are \( R_\alpha^{A'} := R_\alpha^A \), for \( \alpha \in \Gamma \), and \( R_\alpha^{A'} := P_\alpha^A \times A \), for \( \alpha \in I \).

Observe that \( Z : \mathcal{A} \sim_{\forall \gamma, - C} \mathcal{B} \) iff \( Z : \mathcal{A'} \sim_{\forall \gamma, - C} \mathcal{B'} \). Furthermore it is easy to show that \( X \) is a fractional isomorphism between \( A \) and \( B \) iff it is a fractional isomorphism between \( A' \) and \( B' \). \( \square \)
6 Applications to Cube Structures

In this section we apply the notion of fractional isomorphisms of Kripke structures to arbitrary \( \sigma \)-structures by passing to their cubes (see Definition 3). As before, all structures considered are assumed to be finite.

**Definition 10.** Two \( \sigma \)-structures are \( k \)-fractionally isomorphic if their \( k \)-cubes are fractionally isomorphic.

We want to discuss the relationship of \( k \)-WL and the colour refinement algorithm on \( k \)-cubes. For this we first state \( k \)-WL as given in [5]: given two structures \( A \) and \( B \) the \( k \)-WL iteratively computes a colouring of \( A^k \cup B^k \). Initially, two tuples \( a, b \in A^k \cup B^k \) are in the same class if \((a, b) \in \text{Part}_k(A, B)\). In each round the partition \( P \) is refined by distinguishing tuples \( a, b \) if there is an \( i \in [k] \) and \( p \in P \) s.t. \( a \) and \( b \) have a different number of \( i \)-neighbours in \( p \), where an \( i \)-neighbour is a tuple which differs at most in the \( i \)-th component. It is easy to see that this describes the run of the colour refinement algorithm on the \( k \)-cubes of \( A \) and \( B \).

**Observation 2.** The \( k \)-dim Weisfeiler-Lehman algorithm for two \( \sigma \)-structures is given by the colour refinement algorithm on their \( k \)-cubes.

Together with Lemma 6 and Lemma 1 this yields.

**Proposition 3.** Two finite \( \sigma \)-structures are counting \( k \)-pebble equivalent if, and only if, they are indistinguishable by the \( k \)-dim Weisfeiler-Lehman algorithm if, and only if, they are \( k \)-fractionally isomorphic.

Note that \( 2 \)-fractional isomorphisms are different from fractional isomorphisms on the class of graphs (cf. Fig. 1b). However by [9] and [12] these two notions agree on the class of simple graphs.

We can write down the equations involved in the definition of \( k \)-fractional isomorphism in a neater format: \( \text{Fiso}^k(A, B) \) is a linear program in the variables \((X_{ab})_{(a,b) \in A^k \times B^k} \) given by

\[
\begin{align*}
\text{Fiso}^k(A, B) : & \quad X_{ab} \geq 0 \text{ for all } a \in A^k, b \in B^k \quad \text{\{Positivity Cond.\}} \\
& \quad \sum_{a' \in A^k} X_{a'b} = \sum_{b \in B^k} X_{ab'} = 1 \\
& \quad \sum_{a' \in A} X_{a'b} = \sum_{b' \in B} X_{ab'} \\
& \quad \text{for all } i \in [k], (a, b) \in A^k \times B^k \quad \text{\{Continuity Cond.\}} \\
& \quad X_{ab} = 0 \text{ if } (a, b) \notin \text{Part}_k(A, B) \quad \text{\{Compatibility Cond.\}}
\end{align*}
\]

Here we have reformulated the conditions on the propositions in the style of Lemma 5 which results in the compatibility condition.

**Observation 3.** Two \( \sigma \)-structures \( A \) and \( B \) are \( k \)-fractionally isomorphic if the system \( \text{Fiso}^k(A, B) \) has a solution.

9
Sherali-Adams Relaxation of the Graph Isomorphism Problem. Applying the Sherali-Adams relaxation technique [11] to Iso results in a sequence of linear programs $Iso^k$ which by the works of Atserias, Malkin and Maneva [1,7] have a tight connection with the counting $k$-pebble game.

**Proposition 4.** Let $k \geq 2$. Then for any two finite graphs $A$ and $B$ the solvability of $Iso^k(A,B)$ implies counting $k$-pebble equivalence of $A$ and $B$ and counting $k$-pebble equivalence of $A$ and $B$ implies the solvability of $Iso^{k-1}(A,B)$.

Our treatment demonstrates this proposition as a manifestation of the modal view of first-order logic according to Lemma 1. Basically the same approach via cube structures is used in [4], the journal version of [5], but without any emphasis on modal logic.

Finally, we want to remark that, although the solvability of the naive linearisation of Iso is not captured by any logic over the class of directed graphs, as pointed out in the introduction, our analysis shows that for the result of Atserias, Malkin and Maneva it does not matter whether or not we restrict ourselves to simple graphs or we use the symmetrized version of Iso, i.e., adding $BX = X^t A$ to the equations of Iso$(A,B)$. In all of these cases the discriminating power of the $k$-level Sherali-Adams relaxation lies between $(k+1)$-pebble and $k$-pebble equivalence. However, keep in mind that the interleaving in Proposition 4 is strict in all of these cases by [4].

We devote the rest of this section to the proof of proposition 4. First we state the equations of $Iso^k$ according to [1] and introduce $Iso^{k-}$ a weakened form of $Iso^k$. Then we show that the solvability of $Iso^k$ is sandwiched between the solvability of $Iso^{k-}$ and $Iso^{(k+1)-}$ and finally that $Iso^{k-}$ admits a solution iff $Fiso^k$ does. By Proposition 3 and Observation 3 this implies Proposition 4.

The system $Iso^k(A,B)$ uses variables $X_p$ indexed by sets $p \subseteq A \times B$ of size up to $k$. We use as abbreviations $X_{ab}$ for $X_{\{(a,b)\}}$ and $X_{p,ab}$ for $X_{p,\{(a,b)\}}$.

\begin{align*}
Iso^k(A,B) : & \quad X_p \geq 1 \text{ for all } |p| \leq k \end{align*}

\begin{align*}
& \quad X_\emptyset = 1, \quad \begin{array}{l}
\begin{aligned}
X_p = & \sum_{a' \in A} X_{p,a'b} = \sum_{b' \in B} X_{p,ab'} \\
& \text{for all } |p| < k, a \in A, b \in B
\end{aligned}
\end{array} \quad \text{Positivity Condition} \\
& \quad \begin{array}{l}
\begin{aligned}
\sum_{a' \in A} \kappa_{aa'} X_{p,a'b} = & \sum_{b' \in B} \mathbb{B}_{b'b} X_{p,ab'} \\
& \text{for } |p| < k, a \in A, b \in B
\end{aligned}
\end{array} \quad \text{Continuity Condition} \\
& \quad \begin{array}{l}
\begin{aligned}
\sum_{a' \in A} X_{p,a'b} = & \sum_{b' \in B} X_{p,ab'} \\
& \text{for } |p| < k, a \in A, b \in B
\end{aligned}
\end{array} \quad \text{Compatibility Condition}
\end{align*}

The linear program $Iso^{k-}$ is obtained from $Iso^k$ by replacing the compatibility condition with

\begin{align*}
& \quad \text{Weak Compatibility Condition} \\
& \quad X_p = 0 \text{ if } p \text{ is no part. Isomorphism for all } |p| \leq k
\end{align*}
Lemma 7. Let $k \geq 2$. If $\text{Iso}^k$ has a solution then $\text{Iso}^{k-1}$ has a solution and if $\text{Iso}^{k-1}$ has a solution then $\text{Iso}^{k-2}$ has a solution.

Proof. Assume that $X$ is a solution of $\text{Iso}^k(A, B)$. By induction over the continuity conditions we have $X_q \geq X_p$ for all $q \leq p$. Furthermore $X_p > 0$ implies that $p$ is a partial bijection. For example, if $p$ is s.t. $X_p > 0$ and $(a, b), (a, b^*) \in p$ with $b \neq b^*$ then $X_{ab} = \sum_{b' \in B} X_{ab, ab'} \geq X_{ab} + X_{ab, ab^*}$ contradicts that $X_{ab, ab^*} \geq X_p > 0$. Finally we show that $X_p > 0$ implies that $p$ is a partial Isomorphism. For example let $a, a^*$ in the range of $p$ s.t. $(a, a^*) \in R^A$ and $b, b^*$ their images under $p$, then $A_{aa^*}X_{ab, a^*b} \leq \sum_{a' \in A} A_{aa'} X_{ab, a'b} = \sum_{b' \in B} B_{b'b} X_{ab, ab'} = B_{b^b} X_{ab}$ implies that also $(b, b^*) \in R^B$, since $0 < X_p \leq X_{ab, a^*b^*} \leq X_{ab}$.

Now let $X$ be a solution of $\text{Iso}^{k-1}(A, B)$. Then by truncating the solution to the variables $X_p$ with $|p| \leq k-1$ we get a solution of $\text{Iso}^{k-1}(A, B)$: let $|p| < k-1$ and $a \in A$ and $b \in B$, then

$$\sum_{a' \in A} A_{aa'} X_{p, a'b} = \sum_{a' \in A} A_{aa'} \sum_{b' \in B} X_{p, a'b} = \sum_{b' \in B} B_{b'b} \sum_{a' \in A} A_{aa'} X_{p, a'b}$$

where the third equality is due to the fact that if $X_{p, a'b} > 0$ then the set $\{(a', b), (a, b^*)\}$ forms an partial isomorphism.

Note that $\text{Iso}^{k-1}$ can be applied to arbitrary $\sigma$-structures, not only to graphs.

Lemma 8. Let $\mathfrak{A}, \mathfrak{B}$ be two $\sigma$-structures. Then $\text{Iso}^{k-1}(\mathfrak{A}, \mathfrak{B})$ has a solution if, and only if, $\text{Fiso}^k(\mathfrak{A}, \mathfrak{B})$ has a solution.

Proof. Any solution $\hat{X}$ of $\text{Iso}^{k-1}(\mathfrak{A}, \mathfrak{B})$ can easily be transformed into a solution of $\text{Fiso}^k(\mathfrak{A}, \mathfrak{B})$ by putting $X_{ab} := \hat{X}_p$ for $p := \{(a_i, b_i) : i \in [k]\}$.

For the converse direction we first introduce the notion of a permutation invariant solution of $\text{Fiso}^k$. Let $\pi$ be a permutation of $[k]$ and $a$ a $k$-tuple. The tuple resulting of applying $\pi$ to the index of $a$ is denoted by $a^{\pi}$, i.e., the $i$-th component of $a^{\pi}$ is $a_{\pi(i)}$. Then a permutation invariant solution of $\text{Fiso}(\mathfrak{A}, \mathfrak{B})$ is a solution s.t. $X_{ab} = X_{a^{\pi}b^{\pi}}$ for all permutations $\pi \in \text{Sym}(k)$. Let $X$ be a solution of $\text{Fiso}^k(\mathfrak{A}, \mathfrak{B})$, then

$$(X)_{ab} := \frac{1}{k!} \sum_{\pi \in \text{Sym}(k)} X_{a^{\pi}b^{\pi}}$$

is also a solution of $\text{Fiso}^k(\mathfrak{A}, \mathfrak{B})$ and clearly permutation invariant.

So now let $X$ be a permutation invariant solution of $\text{Fiso}^k(\mathfrak{A}, \mathfrak{B})$. We construct a solution of $\text{Iso}^k(\mathfrak{A}, \mathfrak{B})$ by putting

$$\hat{X}_p := 1 \text{ and } \hat{X}_p := X_{ab} \text{ for } p = \{(a_1, b_1), \ldots, (a_k, b_k)\}.$$
First we show that this definition is well-defined: let $p \subseteq A \times B$ be of size up to $k$ and $p = \{(q_1,p_1),\ldots,(q_\ell,p_\ell)\}$ with no repetitions of entries. Let $a' = (q_1,\ldots,q_\ell,q_\ell,\ldots,q_1) \in A^k$ and $b' = (p_1,\ldots,p_\ell,p_\ell,\ldots,p_1) \in B^k$. Then $X_{ab} = X_{a'b'}$ for all $a \in A^k, b \in B^k$ such that $p = \{(a_1,b_1),\ldots,(a_k,b_k)\}$: since $X$ is permutation invariant we can assume that $a_i = q_i, b_i = p_i$ for $i \leq \ell$. The remaining entries $i > \ell$ in $a$ and $b$ can be changed to $q_\ell$ and $p_\ell$ via $X_{a'\omega b''} = \sum_{a' \in A} X_{a'\omega b''} = \sum_{b'' \in B} X_{a'\omega b''} = X_{ab}$. It is clear that $\hat{X}$ satisfies the positivity and weak compatibility conditions. For the continuity condition observe that
\[
\sum_{b' \in B} X_{ab'} = \sum_{b' \in B} X_{(a_1,\ldots,a_1)(b',\ldots,b')} = \sum_{b \in B^k} X_{(a_1,\ldots,a_1)b} = 1 = \hat{X}_\emptyset
\]
and for $p \neq \emptyset$ let $a$ and $b$ such that $a_1 = a_2, b_1 = b_2$ and $p = \{(a_i,b_i) : i \in [k]\}$ then
\[
\sum_{a' \in A} \hat{X}_{a',a'b'} = \sum_{a' \in A} X_{a'\omega b'} = \sum_{a' \in A'} X_{a'\omega b'_{\omega}} = X_{ab} = \hat{X}_p.
\]
\[\square\]

References
Performing conditional strategies in strategic STIT theory

Hein Duijf

Department of Philosophy and Religious Studies, Universiteit Utrecht

Abstract. We introduce a formalization of conditional strategies in strategic STIT theory. This will turn out to have unexpected consequences, in particular it turns out that performing a strategy conditional on \( c \) is equivalent to performing that strategy conditional on a logically weaker condition. Hence it will turn out that performance of a strategy conditional on \( c \) can already commit you to performing that strategy if \( c \) is not the case. We will argue in favour of our formalization, this result and some further consequences. Our investigation points to a misunderstanding that the conditions in the conditional strategies are moment-determinate.

1 Introduction

In this paper we present a formal investigation on conditional strategies. These are strategies that include some conditional actions. We often describe the strategies we are performing in this way, for instance ‘if it is hot outside, I eat lunch in the park’ or ‘I will buy you a present if you behave nicely’. These sentences describe my current strategy by mentioning how certain actions are triggered, this is at least a partial description of my current strategy. Take the example of ‘if it is hot outside, I eat lunch in the park’. I may even utter this sentence on a snow day,\(^1\) which implies that this action is not triggered, i.e. I will not eat lunch in the park today.

Furthermore, we want to examine these conditional strategies in a multi-agent setting. The main advantage of this is that it enables us to make a start in the study of interaction and collective action.\(^2\) For instance in a coordination game I have to coordinate my actions with yours, that is to say that I want to perform a strategy conditional on your current strategy. This makes the current exposition applicable to these central concepts in interaction and collective agency.

\(^*\) I gratefully acknowledge financial support from the ERC-2013-CoG project REINS, nr. 616512.

\(^1\) As I was told, a snow day in Canada means a day off from school because of arctic weather conditions.

\(^2\) Contrast collective action with mere cumulative action, which is represented in STIT by a coalition \( A \) in the agency operator \( [A \cdot \text{STIT}] \). The formal study of collective action is still in its infancy.
There has been a lot of work on formalizations of strategies and strategic reasoning. The seminal work on ATL [1] adds strategies to temporal logic. However, the syntax is confined since the main operator only considers the existence of strategies to ensure certain properties. This work is extended in [14] to Counterfactual ATL (CATL). The essential difference is a syntactical extension of the language with an operator $C_i(\sigma, \varphi)$ with the intended meaning of 'suppose agent $i$ chooses the strategy denoted by $\sigma$; then $\varphi$ holds'. In [5] and [8] a first-order extension of temporal logic (both called Strategy Logic) is proposed in which one can quantify over the strategies. In [5] they introduce separate quantifiers for each agent, while in [8] the quantification is over general strategies but an operator $(a,x)\varphi$ is added to say that ‘bind agent $a$ to the strategy associated by $x$’. All of these frameworks focus on the existence of strategies and lack talk about the structure of the strategies involved. In this paper we aim to provide a formalization of performing conditional strategies, so we have to be able to point out structural properties of strategies.

In [15] the semantics are similar to that of ATL, except strategies are defined as “rough plans” in contrast to the complete, deterministic strategies that are considered in ATL. So on the semantic side they propose a more liberal view on strategies. One of the main themes they discuss is the representation of strategies by formulas of the language. In the current paper we aim to represent the performance of conditional strategies by formulas of some language. Since their language is very elementary it is not rich enough for our purposes.

It is not surprising that Dynamic Logics have been proposed for modelling strategies and the structure of the strategies (see [12] and [13]). However, these logics seem only appropriate for finite extensive form games and it is not clear how the methods generalize to infinite games.

The logic proposed in [11] evaluates formulas at game-strategy pairs thereby combining aspects of Game Logic (see the original work [9] and the overview [10]) and strategic reasoning. A multi-sorted language is proposed to emphasize the structure of the strategy. The logic includes two ‘types’ of conditional strategies $[\psi \rightarrow a]^i$ and $\pi \rightarrow \sigma$, which are interpreted as ‘player $i$ chooses move $a$ whenever $p$ holds’ and ‘player $i$ sticks to the specification given by $\sigma$ if on the history of play, all moves made by $i$ conform to $\pi$’ respectively. The $\psi$ in the first formula is of a special nature, namely it is a boolean combination of propositional letters, and the second formula represents that the other players have acted conform $\pi$. We, however, do not want to commit ourselves to these restrictions on conditions. For instance, it is not obvious whether a sentence like ‘I will buy you a present if you behave nicely’ can be represented in that logic. Indeed, it could be that at the very moment I buy you a present you misbehave. The proposed logic does not allow my current actions to be conditional on your current actions.

In order to adequately formalize conditional strategies we will use the framework G.STRAT (first introduced in [3]), which is an extension of basic STIT frameworks to a strategic and multi-agent setting. As is usual in STIT frameworks, we are able to talk about the structure of a strategy by explicating which (temporal) properties it ensures. In addition, this formalism allows us to talk
about the dynamics of the world, in particular it allows us to say that an agent *is performing* an action. In consequence, we can use this to say that when a certain condition holds the agent *performs* a certain action, which means that the action is triggered by the condition.

The paper is organized as follows. In Section 2 we introduce the logical framework **G.STRAT**. Subsequently, in Section 3 we will present our formalization of conditional strategies, treat several examples to challenge our intuitions, derive some formal results and argue in favour of our formalization. The proofs of the formal results can be found in the Appendix A. We will conclude in Section 4 with some discussion, a summary of the results and key contributions of this paper. If the reader does not want to delve deep into formalisms, it is advisable to start from Section 3 and only look up formal aspects in Section 2 when needed.

## 2 Strategic stit theory: **G.STRAT**

Below we present the formal syntax and semantics of **G.STRAT**, a logic that was first presented in [3] (though our presentation is slightly different) and extends the classical **STIT** framework (cf. [2] and [7]) to strategic actions. The **G** in the acronym **G.STRAT** stands for ‘Group’ and **STRAT** stands for ‘strategic’.

**Definition 1** (Syntax). Well-formed formulas of the language $\mathcal{L}_{G.STRAT}$ are defined by:

$$\varphi := p | \neg \varphi | \varphi \land \varphi | [Asstit] \varphi | \Box \varphi | X \varphi | G \varphi$$

The $p$ are elements from a countable infinite set of propositional symbols $\mathcal{P}$, and $A$ is a subset of a finite set of agents $\text{Ags}$. We will often use $i$ to refer to an arbitrary agent in this set. We use $\varphi, \psi, \ldots$ to represent arbitrary well-formed formulas. We use the standard propositional abbreviations, the standard notation for the duals of modal boxes (i.e., diamonds).

We now go on to define the semantic structures for **G.STRAT**. Here we take the viewpoint that strategies are sets of histories obeying certain structural properties. This is nothing more than a convenient shift of viewpoint and does not result in a fundamentally new type of strategies. If we define strategies as mappings from states / moments to choices, we have to define ‘compliance’ of a history to a strategy as a secondary concept. By defining strategies as sets of histories with a certain structure there is no need anymore to define a notion of ‘compliance’.

We introduce a semantics for this language using standard techniques in modal logic. The units of evaluation are ‘profiles’. A profile records the dynamic aspects of a system of agents. The profiles of our semantics take a moment, a history and a strategy profile (a list of strategies, one for each agent in the system)$^3$ as components. So, the formulas of **G.STRAT** are evaluated against

---

$^3$ Beware that we have ‘profiles’ and ‘strategy profiles’ in our terminology. A strategy profile is a part of a profile.
tuples $\langle m, h, \alpha_1, \alpha_2, \ldots, \alpha_n \rangle$, where $m$ is a moment, $h$ a history, and $\alpha_1, \alpha_2, \ldots, \alpha_n$ a strategy profile. Then, the truth of formulas is evaluated against the background of a current moment, a current history, and a current strategy-profile.

If, under this semantics, we want to consider truths that do not depend on dynamic aspects as represented by the histories and the strategies, we can use the historical necessity operator $\square$. In particular, if $\square \varphi$ holds, $\varphi$ can be said to hold ‘statically’. In $\text{stl}$-theory, one says that $\varphi$ is ‘moment determinate’. We also say that $\varphi$ is ‘settled’\(^4\), which refers to the fact that it is completely independent of any action currently taken by any agent in the system.

**Definition 2** (Strategic frames). A frame is a tuple $F = \langle M, H, \{\text{Strat}(i) \mid i \in \text{Ags}\}, \{R_A \mid A \subseteq \text{Ags}\} \rangle$ such that:

1. $M$ is a non-empty set of moments, which are denoted by $m, m', \ldots$
2. $H$ is a non-empty set of ‘backwards bundled’ histories. A history $h \in H$ is a sequence $\ldots m, m', m'' \ldots$ of mutually different elements from $M$. For $m$ appearing strictly before $m'$ on the history $h$ we write $m <_h m'$. To denote that $m'$ succeeds $m$ on $h$ we use a successor function $\text{Succ}$ and write $m' = \text{Succ}(m,h)$. A history $h \in H$ and its successor function $\text{Succ}$ are isomorphic to $(\mathbb{Z}, +1)$. Furthermore, let $H_n = \{h \mid h \in H$ and $n$ on $h\}$. The following constraint on the set $H$ ensures a deterministic past:
   a. for all $h \in H$, if $m = \text{succ}(n, h)$ then $H_m \subseteq H_n$
3. $\text{Strat}(i)$ yields for each $i \in \text{Ags}$ a non-empty set of strategies. Strategies are non-empty sets of histories. A strategy profile\(^5\) is a list of strategies $\alpha_1, \alpha_2, \ldots, \alpha_n$, where $\{1, 2, \ldots, n\} = \text{Ags}$ and $\alpha_i \in \text{Strat}(i)$ for any $i$. For strategy profiles we will use the vector notation $\vec{\alpha}$ when we need to be more concise. We introduce the notion of strategies for agent $i$ at $m$, denoted by $\text{Strat}_m(i)$, as the set of strategies that are compliant with $m$, formally $\{\alpha \mid \alpha \in \text{Strat}(i) \land \alpha \cap H_m \neq \emptyset\}$.\(^6\) Strategies obey the constraints a., b., c., d. and e. below.
   a. for all $m \in M$ and every agent $i$, there is a strategy at $m$
   b. for all $m \in M$ and for any $i \in \text{Ags}$, if $\alpha_i$ is a strategy for agent $i$ at $m$, and if for a history $h' \in H$ through $m$ it holds that for all $x$ on $h'$ there is a $h'' \in \alpha_i$ such that $x$ is on $h''$, then $h' \in \alpha_i$

\(^4\) Settledness does not necessarily mean that a property is always true in the future (as often thought). Settledness may also apply to the condition that $\varphi$ occurs ‘some’ time in the future, now, or indeed any condition expressible as linear time temporal formula. So, settledness is a universal quantification over the branching dimension of time, and not over the dimension of duration.

\(^5\) In the game forms of game theory strategy profiles are referred to by means of names associated with the choices of agents in states. Here we abstract from names of choices.

\(^6\) It makes sense to talk about strategies at a moment. Since we represent strategies by the set of histories that are compliant with that strategy, a strategy will most likely not be compliant with every moment. Note that this notion is redundant when we view strategies as functions from moments to choices, like is usually done in for instance $\text{ATL}$ (see [1]), since in that case we would have $\text{Strat}_m(i) = \text{Strat}(i)$. 

16
c. for all \( m \in M \) and for any \( i \in Ags \), if \( \alpha_i \) is a strategy for agent \( i \) at \( m \), and there is another strategy \( \beta_i \) for agent \( i \) at \( n = \text{Succ}(m, h) \), then there is a strategy \( \gamma_i \) for agent \( i \) at \( m \) such that \( \gamma_i = (\alpha_i \setminus H_n) \cup (\beta_i \cap H_n) \)

d. for all \( m \in M \) for all strategy profiles at \( m \), i.e. lists of strategies \( \alpha_1, \ldots, \alpha_n \) at \( m \), we have that \( \bigcap_{i \in Ags} \alpha_i \) contains at least one history through \( m \)

e. for all \( m \in M \) for all strategy profiles at \( m \), i.e. lists of strategies \( \alpha_1, \ldots, \alpha_n \) at \( m \), we have that \( \bigcap_{i \in Ags} \alpha_i \) contains at most one history through \( m \)

4. Profiles are exactly the tuples \( \langle m, h, \alpha_1, \alpha_2, \ldots, \alpha_n \rangle \) such that \( m \) belongs to \( h \) and for all \( i \in Ags \), \( h \in \alpha_i \). These will be the units of evaluation (possible worlds) of our modal semantics.

5. The relations \( R_A \) are ‘effectivity’ equivalence classes over profiles such that \( \langle m, h, \alpha_1, \alpha_2, \ldots, \alpha_n \rangle R_A \langle m', h', \beta_1, \beta_2, \ldots, \beta_n \rangle \) if and only if \( m = m' \), and for all \( i \in A, \alpha_i = \beta_i \)

Item 1 gives the basic elements of the frames: the set of moments \( M \).

Item 2 defines histories to be linearly ordered sets of moments. It also defines a bundling structure for the set \( H \) of all histories: histories that come together when going in the past direction will stay together in that direction. This implies that in the future direction, once bundles have separated they will never come together again.

Item 3 defines strategies. Strategies are sets of histories with certain properties. A strategy profile is a choice of strategy for each agent in the system. Condition 3(a) ensures that for each moment, for each agent, its strategy at that moment is defined by some bundle of histories. Condition 3(b) ensures that our bundled histories semantics behaves as a tree semantics on aspects relevant for the comparison with \( ATL \) (see [1]). It says that if an agent indefinitely complies to a strategy then he is performing that strategy. In particular, this property ensures that \( G.\text{STRAT} \) obeys \( ATL \)'s induction axiom. Condition 3(c) ensures that if at a certain moment an agent has two different strategies in his repertoire, recombinations of these strategies where first one strategy is followed and later on (next) the other, are always also in the repertoire of that agent at that moment. Condition 3(d) ensures that we can always recombine strategies of individual agents contributing to a group strategy into a new joint strategy. This implements the \( stit \)-requirement of independence of agency (no agent can choose in such a way that some other agent is deprived of one of its choices). Condition 3(e) ensures that there is exactly one history complying to all strategies of a strategy profile at a moment. This reflects the idea that a choice of strategy for each agent in the system completely determines the entire future.

Item 4 introduces profiles, which are the units of evaluation for our modal semantics. Profiles are tuples consisting of a moment, a history and a strategy for each agent.

Item 5 defines \( R_A \) to be a relation reaching all profiles that only deviate from the current profile in the sense that agents not among \( A \) perform a choice different from the current one. This reflects the basic idea of agency saying that acting or choosing is ensuring a condition irrespective of what other agents do or choose.
Now we are ready to define the formal semantics of the language \( L_{G.STRAT} \). The semantics is multi-dimensional, and the truth conditions are quite standard. First we define models based on the frames of the previous definition.

**Definition 3** (Strategic models). A frame \( \mathcal{F} = \langle M, H, \{Strat(i) \mid i \in Ags\}, \{R_A \mid A \subseteq Ags\} \rangle \) is extended to a model by adding a valuation \( \pi \) of atomic propositions:

- \( \pi \) is a valuation function assigning to each atomic proposition the set of profiles (see Definition 2) in which they are true.

**Definition 4** (Truth, validity, logic). Truth \( \mathcal{M}, \langle m, h, \vec{a} \rangle \models \varphi \), of a \( G.STRAT \)-formula \( \varphi \) in a profile \( \langle m, h, \vec{a} \rangle \) of a model \( \mathcal{M} = \langle M, H, \{Strat(i) \mid i \in Ags\}, \{R_A \mid A \subseteq Ags\}, \pi \rangle \) is defined as (suppressing the model denotation ‘\( M \)’):

\[
\begin{align*}
\langle m, h, \vec{a} \rangle \models p & \iff \langle m, h, \vec{a} \rangle \in \pi(p) \\
\langle m, h, \vec{a} \rangle \models \neg \varphi & \iff \not \langle m, h, \vec{a} \rangle \models \varphi \\
\langle m, h, \vec{a} \rangle \models \varphi \land \psi & \iff \langle m, h, \vec{a} \rangle \models \varphi \text{ and } \langle m, h, \vec{a} \rangle \models \psi \\
\langle m, h, \vec{a} \rangle \models X\varphi & \iff \text{for } m' = \text{Succ}(m, h) \text{ it holds that } \langle m', h, \vec{a} \rangle \models \varphi \\
\langle m, h, \vec{a} \rangle \models G\varphi & \iff \text{for all } m' \text{ such that } m \leq h, m' \text{ it holds that } \langle m', h, \vec{a} \rangle \models \varphi \\
\langle m, h, \vec{a} \rangle \models [A \text{ sstit }]\varphi & \iff \text{for all } h', \vec{b} \text{ such that } \langle m, h, \vec{a} \rangle \models R_A \langle m', h', \vec{b} \rangle \text{ it holds that } \langle m, h', \vec{b} \rangle \models \varphi \\
\end{align*}
\]

Validity on a \( G.STRAT \)-model \( \mathcal{M} \) is defined as truth in all profiles of the \( G.STRAT \)-model. General validity of a formula \( \varphi \) is defined as validity on all possible \( G.STRAT \)-models. The logic \( G.STRAT \) is the subset of all general validities of \( L_{G.STRAT} \) over the class of \( G.STRAT \)-models.\(^7\)

### 3 Conditional strategies

In order to adequately formalize conditional strategies we will use the \( G.STRAT \) framework as presented in Section 2. We briefly present key elements and aspects

\(^7\) Little is known about the validities of this logic, for instance there is no axiomatization nor completeness result for this logic. To guide some of the formal intuitions of the reader we mention some validities without proof or conceptual motivation: \( \square \) and \( [i \text{ sstit }] \) are \( S5 \)-modalities, the \( [i \text{ sstit }] \)-operator is monotone in its agency argument, i.e. for \( A \subseteq B \) we have \( [A \text{ sstit }] p \rightarrow [B \text{ sstit }] p \), the temporal part is the standard discrete linear temporal logic containing \( X \) and \( G \), some interaction principles are \( \equiv [i \text{ sstit }] \square p \rightarrow \square p, i [i \text{ sstit }] X p \rightarrow X [i \text{ sstit }] p, i [i \text{ sstit }] [i \text{ sstit }] p \rightarrow [i \text{ sstit }] [i \text{ sstit }] p, [i \text{ sstit }] p \rightarrow [i \text{ sstit }] [i \text{ sstit }] p \) (see Lemma 10 in Appendix A), and \( \Diamond [A \text{ sstit }] p \land \Diamond [B \text{ sstit }] q \rightarrow \Diamond [A \cup B \text{ sstit }] p \land q \) for disjoint coalitions \( A \) and \( B \) (independence of agency).
of G.STRAT that are needed in order to understand our formalism. G.STRAT is a strategic STIT logic that evaluates formulas against tuples \( \langle m, h, \vec{\alpha} \rangle \) that are referred to as profiles. The elements of these profiles are a moment \( m \), a history \( h \), and a strategy profile \( \vec{\alpha} \) specifying a strategy \( \alpha_i \) for each agent \( i \) in the language.

The central agency operator is the modality \([i \text{ sstit}] \varphi\) which stands for ‘agent \( i \) strategically sees to it that \( \varphi \)’. Relative to a profile \( \langle m, h, \vec{\alpha} \rangle \) the modality \([i \text{ sstit}] \varphi\) is interpreted as ‘agent \( i \) is in the process of executing \( \alpha_i \) thereby ensuring the (temporal) condition \( \varphi \)’. In addition, the logic includes temporal modalities \( X \varphi \) and \( G \varphi \) which are interpreted, relative to a profile \( \langle m, h, \vec{\alpha} \rangle \), as ‘\( \varphi \) holds in the next moment after \( m \) on the current history \( h \)’ and ‘\( \varphi \) holds in every moment after \( m \) on the current history \( h \)’ respectively. Finally, the logic includes a historical necessity operator \( \Box \varphi \) which is interpreted, relative to a profile \( \langle m, h, \vec{\alpha} \rangle \), as ‘\( \varphi \) holds on any profile at \( m \)’ or more informally ‘\( \varphi \) is settled at \( m \)’ which says that the truth does not depend on dynamic aspects as represented by the histories and the strategies. We now return to the main story concerning conditional strategies and our formalization thereof.

People often make conditional statements about their future choices: ‘if it is hot outside, I eat lunch in the park’, ‘I will buy you a present if you behave nicely’. It is crucial to view these phrases as denoting aspects of the strategy I am currently performing.

**Definition 5. (Conditional strategy)** We say that relative to a profile \( \langle m, h, \vec{\alpha} \rangle \) in a model \( M \) an agent \( i \) performs a strategy ensuring \( \varphi \) conditional on \( c \) if and only if the following truth condition is satisfied:

\[
M, \langle m, h, \vec{\alpha} \rangle \models [i \text{ sstit}] (c \rightarrow [i \text{ sstit}] \varphi). \tag{9}
\]

The above formula means that agent \( i \), along history \( h \), performs strategy \( \alpha_i \) that ensures that for all histories in the strategy the conditional is ensured. Informally, this says that agent \( i \) is currently performing a strategy that ensures that in case the condition holds he performs a strategy ensuring \( \varphi \).

The nested \([i \text{ sstit}]\) operator might look confusing, but this makes perfect sense. In order to argue in favour we break the formula down: the conditional strategy is formalized as \( c \rightarrow [i \text{ sstit}] \varphi \), but one should not forget that here we focus on an agent performing such a strategy. This is formalized by the second \([i \text{ sstit}]\) operator and the \( G \) operator.\(^{10}\)

---

\(^{8}\) One could argue that these phrases announce a commitment that I currently have. I believe the current framework can be useful for this interpretation so it might be useful in modeling commitments.

\(^{9}\) In [4] Broersen proposes to formalize such a conditional strategy by using the formula \([i \text{ sstit}] G(c \rightarrow [i \text{ sstit}] X \varphi)\). We think that this is just a minor difference. Since our current research and findings also apply to that proposal, we will not discuss the differences.

\(^{10}\) Note that it might make sense not to require an agent to ensure that henceforth some temporal condition holds, but that the agent ensures this temporal condition until a certain deadline. In light of Definition 5 this might be similarly formalized.
The possibility to nest strategic action modalities is a crucial feature of G.STRAT. This allows us to express many interesting properties of strategies and consequently of abilities. For instance, the formula $\Diamond [i \sstit [j \sstit] \varphi]$ says that agent $i$ has the ability to ensure that agent $j$ performs a strategy ensuring $\varphi$, or the formula $[i \sstit]([i \sstit] \varphi \land X[i \sstit] \psi)$ says that agent $i$ currently performs a strategy that ensures that he is performing a strategy ensuring $\varphi$ now and immediately after that a strategy ensuring $\psi$.

Note that in case an agent is performing a strategy ensuring $\varphi$ conditional on $c$, he might never perform a strategy ensuring $\varphi$. This is obvious since the condition may never be satisfied (for instance if $c = \bot$). It also shows that this conditional strategy of the agent might be interleaved with other strategies. The only guarantee that we have when the agent performs this conditional strategy is that he ensures that the conditional is met. The current formal setting enables us to investigate these conditional strategies in further detail.

**Proposition 6.** Let $M$ be a G.STRAT model and $\langle m, h, \vec{a} \rangle$ be a profile in $M$. Then relative to $\langle m, h, \vec{a} \rangle$ agent $i$ performs a strategy ensuring $\varphi$ conditional on $c$ if and only if

$$M, \langle m, h, \vec{a} \rangle \models [i \sstit]G((i \sstit)c \rightarrow [i \sstit] \varphi),$$

that is, agent $i$ performs a strategy ensuring $\varphi$ conditional on $\langle i \sstit \rangle c$.

This means that performing a strategy ensuring $\varphi$ conditional on $c$ is equivalent to performing a strategy ensuring $\varphi$ conditional on $\langle i \sstit \rangle c$. The latter condition is satisfied exactly when my strategy does not ensure $\neg c$, i.e. if there is a history $h'$ in my strategy $\alpha_i$ that validates $c$. Note that we have $\models c \rightarrow \langle i \sstit \rangle c$ and $\not\models \langle i \sstit \rangle c \rightarrow c$, so the above proposition says that performing a strategy ensuring $\varphi$ conditional on $c$ is equivalent to performing a strategy ensuring $\varphi$ conditional on the logically weaker condition $\langle i \sstit \rangle c$.

The following example is devised to show that this property might be very counterintuitive. However, we will subsequently locate where the intuitive mismatch comes from and argue in favour of our formalism by way of another example. But first we will treat an example revealing some unexpected results of our formalization:

Consider the strategy to kill Jones gently conditional on Smith killing Jones.\(^{11}\) We consider an agent $i$ and let Smith be agent $j$, and we write $K$ for killing Jones gently. According to Definition 5 we formalize the performance by agent $i$ of this conditional strategy as $[i \sstit]G([j \sstit]K \rightarrow [i \sstit]K_G)$, which is equivalent to $[i \sstit]G((i \sstit)[j \sstit]K \rightarrow [i \sstit]K_G)$ by Proposition 6. Note that allowing Smith to kill Jones is logically equivalent to Smith having the ability to kill Jones, i.e. $\models [i \sstit][j \sstit]K \leftrightarrow$

\(^{11}\) As the attentive reader may already notice, this example is inspired by the Forrester paradox (see [6]).
\( \Diamond [j sstit]K \). Hence we derive that performing this conditional strategy is equivalent to \( [i sstit]G(\Diamond [j sstit]K \rightarrow [i sstit]K_G) \). In words, performing the strategy to kill Jones gently conditional on Smith killing Jones is logically equivalent to performing the strategy to kill Jones gently conditional on Smith having the ability to kill Jones. Note that Smith’s ability to kill Jones does not imply that he actually kills Jones. So it feels like this does not meet our intuitions.

The discussion of the previous paragraph amounts to:

**Proposition 7.** Let \( M \) be a \( \text{G.STRAT} \) model and \( \langle m, h, \vec{\alpha} \rangle \) be a profile in \( M \). We see that relative to \( \langle m, h, \vec{\alpha} \rangle \) agent \( i \) performs a strategy ensuring \( \varphi \) conditional on \( j \) performing a strategy ensuring \( \psi \) if and only if

\[
M, \langle m, h, \vec{\alpha} \rangle \models [i {sstit}G(\Diamond [j {sstit}]K \rightarrow [i {sstit}]K_G)],
\]

that is, agent \( i \) performs a strategy ensuring \( \varphi \) conditional on the (settled) condition that \( j \) is able to perform a strategy ensuring \( \psi \).

What does this result and the motivating example teach us? I think we should draw the conclusion that conditional strategies are quite hard to perform. At the very least the performance of such a conditional strategy might incline you to perform some actions while the condition is not met. This happens especially when the conditions are not settled.

One may see this result as providing evidence against our formalization of performing a conditional strategy, therefore I now want to give an example in favor of our formal result:

Consider the example of the gentle killing of Jones conditional on Smith killing Jones. Suppose Smith has the ability to kill Jones by lethal injection. However, if he does it will take an hour before Jones actually dies. In this scenario I do not know whether Smith kills Jones by lethal injection and the only way to find out is running a test which takes two hours to process. Hence there is no way for me to find out whether Smith kills Jones before Jones dies. If I do not kill Jones gently before he dies, I could have failed to ensure the condition in the conditional strategy (since Smith could have injected Jones with a lethal dose). We conclude that in order to ensure that the conditional is met I have to kill Jones gently without being certain that Smith kills Jones. In fact, the mere ability of Smith implies that I have to kill Jones gently in order to ensure the conditional.

Having motivated our formalization of conditional strategies and the formal result of Proposition 7 we now try to locate the (or: a probable) source of our previous intuition that the formal result of Proposition 7 was flawed. In order to do this, we shift our attention to settled conditions.

A formula being settled means that it is determined only by the current moment, i.e. the property is moment-determinate. This means that it does not depend on the current history, nor on the strategies that the agents are currently performing. In contrast to other formalisms the current \( \text{STIT} \) framework

\[12\] This follows from the axiom of independence of agency.
allows us to differentiate between static and dynamic properties. These moment-determinate formulas correspond to static properties.

What happens if the conditions are settled?

**Proposition 8.** Let $\mathcal{M}$ be a $\text{G.STRAT}$ model and $\langle m, h, \vec{a} \rangle$ be a profile in $\mathcal{M}$. Suppose $\mathcal{M} \models c \rightarrow \Box c$. Then relative to $\langle m, h, \vec{a} \rangle$ agent $i$ performs a strategy ensuring $\varphi$ conditional on $c$ if and only if

$$\mathcal{M}, \langle m, h, \vec{a} \rangle \models [i \text{ sstt}]G(\Box c \rightarrow [i \text{ sstt}]\varphi),$$

that is, agent $i$ performs a strategy ensuring $\varphi$ conditional on $c$ being settled.

Note that we have $\models \Box c \rightarrow c$ and $\not\models c \rightarrow \Box c$. The above proposition says that, if in our model $c$ implies that it is settled, performing a strategy ensuring $\varphi$ conditional on $c$ is equivalent to performing a strategy ensuring $\varphi$ conditional on the logically stronger condition of $c$ being settled.

I believe it is a key contribution of the current enterprise that we observe that our intuitions of conditional strategies could be flawed because we tend to think of conditions as being moment-determinate.

4 Discussion

In this paper we have presented a formalization of conditional strategies. The formal setting provides us with useful tools to examine consequences and equivalences of conditional strategies. We showed that performing a strategy ensuring $\varphi$ conditional on $c$ is equivalent to performing a strategy ensuring $\varphi$ conditional on the logically weaker condition $\langle i \text{ sstt} \rangle c$. This observation has some counterintuitive instances. I argued in favour of the presented formalization and the derived results through examples. I believe this mismatch with our intuitions stems from a misunderstanding that we regard the conditions to be moment-determinate.

It is useful to review the work done in [11] in light of the current research. Recall from the introduction that the logic they propose has two operators that involve conditional strategies, namely $[\psi \rightarrow a]$ and $\pi \rightarrow \sigma$, which are interpreted as ‘player $i$ chooses move $a$ whenever $p$ holds’ and ‘player $i$ sticks to the specification given by $\sigma$ if on the history of play, all moves made by $i$ conform to $\pi$’ respectively. The $\psi$ in the first formula is of a special nature, namely it is a boolean combination of propositional letters, and the second formula represents the condition that the other players have acted conform $\pi$.

We see that the conditions they allow for conditional strategies are in fact moment-determinate. We recall that (contrary to our formal framework) they assign to each atomic proposition a set of states at which the proposition holds. Therefore, any boolean combination of atomic proposition will be moment-determinate. Furthermore, it is quite natural to assume that the history of play is settled, i.e. moment-determinate. So it intuitively follows that both types of conditional strategies proposed in [11] involve a moment-determinate condition.
We do not know whether the authors are aware of possible challenges arising when one takes a more liberal view on the conditions involved. We believe that the G.STRAT framework is an adequate formal setting to investigate conditional strategies with a clear mind. Additionally, it seems that this framework can be fruitfully applied to forms of strategic reasoning by zooming in on the structure of the strategies.

A Appendix: Proofs of propositions

We prove the propositions using the following two lemmas:

Lemma 9. For all formulas $\psi$, $\chi$ we have that $([i\ sstit](\psi \rightarrow [i\ sstit]\chi)) \leftrightarrow ([i\ sstit]\psi \rightarrow [i\ sstit]\chi)$ is valid.

Proof. $\Rightarrow$: Suppose $M, \langle m, h, \alpha \rangle \models [i\ sstit](\psi \rightarrow [i\ sstit]\chi) \land (i\ sstit)\psi$. Then for all profiles $\langle m', h', \beta \rangle$ with $\beta_i = \alpha_i$ we have $M, \langle m', h', \beta \rangle \models \psi \rightarrow [i\ sstit]\chi$. In addition, there exists a profile $\langle m', h', \beta' \rangle$ such that $M, \langle m', h', \beta' \rangle \models \psi$. Hence for this profile we have $M, \langle m, h', \beta' \rangle \models [i\ sstit]\chi$, which implies $M, \langle m, h, \alpha \rangle \models [i\ sstit]\chi$, since $\beta'_i = \alpha_i$.

$\Leftarrow$: Suppose $M, \langle m, h, \alpha \rangle \models \langle i\ sstit\rangle\psi \rightarrow [i\ sstit]\chi$. Let $\langle m', h', \beta \rangle$ be any profile such that $\beta_i = \alpha_i$ and $M, \langle m', h', \beta \rangle \models \psi$. We see that this implies $M, \langle m, h, \alpha \rangle \models \langle i\ sstit\rangle\psi$. Hence $M, \langle m, h, \alpha \rangle \models [i\ sstit]\chi$. It follows that $M, \langle m, h', \beta' \rangle \models [i\ sstit]\chi$, so we conclude that $M, \langle m, h, \alpha \rangle \models [i\ sstit]([\psi \rightarrow [i\ sstit]\chi])$.

Lemma 10. For all formulas $\psi$ we have that $[i\ sstit]G[i\ sstit]\psi \leftrightarrow [i\ sstit]G\psi$ is valid.

Proof. The left-to-right is immediate from the factivity of $[i\ sstit]$. To prove the right-to-left implication we assume $M, \langle m, h, \alpha \rangle \models [i\ sstit]G\psi$. This means that for all profiles $\langle m', h', \beta \rangle$ with $\beta_i = \alpha_i$ and $m' \geq_h m$ we have $M, \langle m', h', \beta \rangle \models \psi$. Any profile $\langle m', h'', \gamma \rangle$ with $\gamma_i = \beta_i$ also satisfies $\gamma_i = \alpha_i$, so the previous observation implies that for these we also have $M, \langle m', h'', \gamma \rangle \models \psi$. Hence $M, \langle m', h', \beta \rangle \models [i\ sstit]\psi$. This proves the claim.

Now we prove the Propositions using these lemmas:

Proposition 6 Lemma 10 implies that we have $M, \langle m, h, \alpha \rangle \models [i\ sstit]G\psi \land [i\ sstit][c \rightarrow [i\ sstit]\varphi]$. Applying Lemma 9 by noting that we can replace with equivalents within a $[i\ sstit]$- and a $G$-modality, gives us $M, \langle m, h, \alpha \rangle \models [i\ sstit]G([i\ sstit]c \rightarrow [i\ sstit]\varphi)$.

Proposition 7 We assume that $c = [j\ sstit]\psi$. This proposition now follows from the previous proposition and the fact that $[i\ sstit][j\ sstit]\psi \leftrightarrow [j\ sstit]([i\ sstit]\psi$.
Proposition 8 Under the assumption that $M \vDash c \rightarrow \Box c$ we have $M \vDash c \leftrightarrow \Box c$. Hence, $M \vDash (c \rightarrow [i \text{ stit}]\varphi) \leftrightarrow (\Box c \rightarrow [i \text{ stit}]\varphi)$. This gives us $M \vDash [i \text{ stit}]G(c \rightarrow [i \text{ stit}]\varphi) \leftrightarrow [i \text{ stit}]G(\Box c \rightarrow [i \text{ stit}]\varphi)$, which completes the proof of this proposition.

References

Sufficient (In)completeness for Definitions in Hierarchic Superposition

Adrián Rebola-Pardo

Dresden University of Technology, Germany

1 Introduction

While automated theorem proving systems for first-order logic have become reasonably powerful in the last years, many industrial applications of theorem proving require reasoning with respect to logics that are not expressible in first-order logic (“background (BG) logics”, e.g. integer or real arithmetic), combined with uninterpreted (“foreground”, FG) operators. Different approaches have been proposed to tackle refutational theorem proving on such logics.

Separately, both features are easy to handle. Decision procedures for many background logics have been developed (e.g. Fourier-Motzkin Elimination [8] or Cooper’s Algorithm [7] for linear arithmetic); uninterpreted operators are natural to first-order logic, thus covered by superposition [2].

The Hierarchic Superposition (HSP) calculus [3, 5] proposes the use of specialized provers for the background logic as off-the-shelf solvers. Background problems are first inferred from the foreground input by first-order superposition, and the background solver is then called to check their satisfiability. The HSP calculus has been implemented in the test-bed solver Beagle [6].

HSP is somewhat similar to Satisfiability Modulo Theories (SMT) [11] insofar as both make use of a background solvers as black-box decision procedures. Whereas background clauses in HSP are derived by superposition inferences, SMT uses a variant of the DPLL algorithm. Although some complete fragments for SMT exist [10], these results are limited by the heuristic nature of instantiation methods [9] used in SMT.

Refutational completeness results [5] for the HSP calculus require the background logic to be compact and the input clause set to satisfy a condition called sufficient completeness related to the ability to eliminate BG-sorted FG terms. In the context of finitely quantified theorem proving [4], Finite Domain Transformations (FDT) are generated from a clause set by adding definitions of the form $f(x) \approx \alpha$ where $\alpha$ is a BG term. FDTs were stated to introduce sufficient completeness into finitely quantified clause sets, but no proof was provided for this claim. Were this true, refutational completeness of HSP for FDTs would follow, thus allowing a refutationally complete search procedure for finitely quantified problems.

In this work, we show that this claim is actually false: sufficient completeness is not achieved for general FDTs. However, this does not mean that HSP is not refutationally complete for FDTs, since sufficient completeness is only a sufficient
condition for refutational completeness. We find a fragment, that of FG-ground, BSFG-safe problems, that contains all FDTs generated from clause sets without FG-sorted variables. Furthermore, we prove that HSP is refutationally complete for this fragment, thus showing the procedure presented in [4] to be refutationally complete for finitely quantified theorem proving, as long as FG-sorted variables do not occur in the input clause set.

This work is divided as follows. Section 2 briefly introduces the HSP calculus. Section 3 discusses the relation between sufficient completeness and Finite Domain Transformations, with a counterexample to the aforementioned statement. In Section 4 we introduce the FG-ground, BSFG-safe fragment, and discuss how it is preserved by operations performed by HSP solvers. Section 5 shows how can refutational completeness be recovered for this fragment. Section 6 outlines the main results.

2 Hierarchic Superposition

Hierarchic theorem proving works in the framework of many-sorted first-order logic with equality under a signature \( \Sigma \) given by a countably infinite set of operators together with their sort. Terms over the signature \( \Sigma \) are defined as usual. An equation is a multiset of terms \( \{s,t\} \), whereas a disequation is a multiset of terms \( \{s,s,t,t\} \). We write equations and disequations as \( s \approx t \) and \( s \not\approx t \) respectively. Equations and disequations together are referred to as literals. We use the notation \( (\neg)s \approx t \) to denote a literal that may be either \( s \approx t \) or \( s \not\approx t \).

For the sake of simplicity, we do not consider predicates other than equality; every predicate \( p \) with sort \( \xi_1 \ldots \xi_n \) can be regarded as an operator with sort \( \xi_1 \ldots \xi_n \rightarrow \xi_{\text{Bool}} \) where \( \xi_{\text{Bool}} \) is a distinguished sort and every atom \( p(t_1,\ldots,t_n) \) is rewritten as the equation \( p(t_1,\ldots,t_n) \approx \top \), for a distinguished operator \( \top \) of sort \( \xi_{\text{Bool}} \).

A clause is a multiset of literals. We write the clause \( \{L_1,\ldots,L_n\} \) as \( L_1 \lor \ldots \lor L_n \), and the empty \( \Sigma \)-clause as \( \Box \).

Any of the former objects is said to be ground if no variable occurs on it. Substitutions are defined as usual. An interpretation \( I \) is defined by a family of disjoint, non-empty sets \( I_\xi \) for every sort \( \xi \), called carrier sets, together with a mapping \( f^I : I_\xi \times I_{\xi^n} \rightarrow I_\xi^0 \) for every operator \( f \) of sort \( \xi_1 \ldots \xi_n \rightarrow \xi_0 \). The interpretation \( t^I \) of a ground term \( t \) is defined recursively as usual. \( I \) is said to be term-generated if all elements in \( I^S \) are the interpretation of some ground term of sort \( \xi \).

\( I \) is said to satisfy a ground equation \( s \approx t \) if \( s^I = t^I \); similarly, \( I \) satisfies \( s \not\approx t \) if \( s^I \neq t^I \). \( I \) satisfies a clause \( C \) if every ground instance of \( C \) contains a literal satisfied by \( I \). Finally, \( I \) satisfies a set of clauses \( N \) if \( I \) satisfies every clause in \( N \). We abbreviate the expression “\( I \) satisfies \( N \)” as \( I \models N \), and similarly for clauses and literals.

A specification \( S = (\Sigma,M) \) is a pair where \( \Sigma \) is a signature and \( M \) is a class of term-generated interpretations closed under isomorphism. Consider a pair of specifications \( S_B = (\Sigma_B,M_B) \) and \( S_F = (\Sigma_F,M_F) \) such that \( \Sigma_B \) is a
subsignature of \( \Sigma_F \) (i.e. all sorts and operators of \( \Sigma_B \) exist in \( \Sigma_F \) as well). Given a \( \Sigma_B \)-interpretation \( I \) and a \( \Sigma_F \)-interpretation \( J \), we say that \( J \) is a conservative extension of \( I \) if \( I \) is the restriction of \( J \) to operators and sorts in \( \Sigma_B \). The pair \((S_B, S_F)\) is a hierarchic specification if \( M_F \) contains exactly the conservative extensions of interpretations in \( M_B \).

We refer to terms over the signature \( \Sigma_B \) as background (BG) terms; all other terms are called foreground (FG), and similarly for literals and clauses. Note that FG operators with a BG target sort are allowed; these are said to be background-sorted foreground (BSFG) operators. Similarly, interpretations in \( M_B \) or \( M_F \) are called BG models or FG models respectively. We assume BG specifications to fulfill the following properties:

- \( \Sigma_B \) includes a countably infinite set of Skolem BG-sorted constants, called parameters. We denote parameters by greek letters.
- \( \Sigma_B \) includes a set of BG-sorted constants called domain elements, such that for every \( I \in M_B \) and every pair of distinct domain elements \( d_1 \neq d_2 \), we have \( d_1 \neq d_2 \).
- \( \Sigma_B \) includes an boolean sort denoted by \( \xi_{\text{Bool}} \) and a boolean-sorted top constant \( \top \).
- \( \Sigma_B \) includes a domain predicate operator \( p_D \) of sort \( \xi \to \xi_{\text{Bool}} \) for every finite set of domain elements \( D \) and every BG sort \( \xi \), such that \( p_D^I(o) = \top^I \) if and only if \( o \in \{ d^I \mid d \in D \} \) for all BG models \( I \). For readability, we denote the equation \( p_D(t) \approx \top \) as \( t \in D \).

We say that a clause set \( N \) is satisfiable (with respect to hierarchic semantics) if there is an FG model \( J \in M_F \) with \( J \models N \). Furthermore, the hierarchic signature is said to be compact if, for every unsatisfiable set of BG clauses \( N \), a finite, unsatisfiable subset \( N' \subseteq N \) exists.

The goal of Hierarchic Superposition is to detect when a clause set is unsatisfiable. Upon an input clause set, HSP works by iteratively appending all possible HSP inferences to the input. Every time a new BG clause is inferred, an off-the-shelf BG solver is called on the set of BG clauses; unsatisfiability of the BG clauses would then prove unsatisfiability of the input. Abstraction is used to split literals into their BG and FG components by replacing a clause \( C[t] \) by \( C[x] \lor x \not\approx t \) where \( x \) is a fresh variable. Weak abstraction provides a criterion to decide which terms need to be abstracted in a clause. We denote the weakly abstracted version of a clause \( C \) by \( \text{abs}(C) \).

HSP is parameterized by a hierarchic reduction ordering, that is, a well-founded ordering \( \prec \) with the following properties:

- \( \prec \) is a total order when restricted to the set of ground terms.
- For all terms \( s, t \) with \( s \prec t \) and all simple substitutions \( \sigma \), we have \( s \sigma \prec t \sigma \).
- For all terms \( u, s, t \) with \( s \prec t \), \( u[s] \prec u[t] \).
- For all ground FG terms \( s \), all ground, non-domain element, BG terms \( t \) and all domain elements \( d \), we have \( d \prec t \prec s \).

A hierarchic reduction ordering can be extended to a well-founded ordering over literals and clauses by multiset extension [1].
A full-detail description of HSP inferences is presented in [5]. Due to space constraints, we present a simplified version thereof. HSP inferences are modifications of the original superposition calculus [2]. In particular, we have inference rules such as \textbf{Positive Superposition}:

\[
\frac{l \approx r \lor C \quad s[u] \approx t \lor D}{\text{abs}((s[r] \approx t \lor C \lor D)\sigma)}
\]

where:

– Neither \( l \) nor \( u \) are BG terms.
– \( u \) is not a variable.
– \( \sigma \) is a mgu of \( l \) and \( u \) that maps BG-sorted variables to either domain elements or variables.
– \( l\sigma \nRightarrow r\sigma \)
– \( s\sigma \nRightarrow t\sigma \)
– \( (l \approx r)\sigma \) is strictly maximal in \((l \approx r \lor C)\sigma \).
– \( (s \approx t)\sigma \) is strictly maximal in \((s \approx t \lor D)\sigma \).

Calls to the BG solver are modeled by the \textbf{Close} inference rule:

\[
\text{Close} \quad C_1 \quad \ldots \quad C_n \quad \square
\]

where:

– \( C_1, \ldots, C_n \) are BG clauses.
– \( \{C_1, \ldots, C_n\} \) is unsatisfiable with respect to hierarchic semantics.

Furthermore, a simplification rule allows to replace clauses in a satisfiability-preserving way. The current implementation of HSP allows the application of several ad hoc techniques within the simplification rule, namely demodulation by unit clauses, tautology deletion, elimination of subsumed clauses, arithmetic simplification, unabstraction of domain elements and negative unit simplification [6].

HSP is endowed as well with \textit{redundancy criteria} that detect inferences which are not needed to infer the empty clause. A clause set \( N \) is \textit{saturated} if all possible HSP inferences over \( N \) are redundant. We say that HSP is \textit{refutationally complete} for a class of clause sets \( \mathcal{P} \) if for all saturated unsatisfiable clause sets \( N \in \mathcal{P} \) we have \( \square \in N \).

### 3 Sufficient Completeness and Definitions

An interpretation \( I \) is said to be \textit{BG-complete} if for every BG-sorted term \( t \) there exists a BG term \( s \) such that \( I \models s \approx t \). Define \( \text{GndTh}(\mathcal{M}_B) \) as the set of ground BG clauses satisfied by all \( I \in \mathcal{M}_B \). A clause set \( N \) is said to be \textit{sufficiently complete} if every interpretation \( I \) (not necessarily in \( \mathcal{M}_B \)) with \( I \models \text{sgi}(N) \cup \text{GndTh}(\mathcal{M}_B) \) is BG-complete. The following completeness result for HSP was proved in [5].
Theorem 1. Let $C$ be the class of weakly abstracted, sufficiently complete clause sets. If the hierarchic specification is compact, then HSP is refutationally complete for $C$.

In practice, sufficient completeness is a rather restrictive condition. Different techniques have been developed in order to introduce sufficient completeness in a problem, such as the Define rule in [5] or Finite Domain Transformations (FDT) in [4]. Such methods work by selecting every occurrence of a BSFG operator $f$ and replacing the term $f(t_1,\ldots,t_n)$ in that position by a parameter $\alpha$. Furthermore, a definition of the form $f(t_1,\ldots,t_n) \equiv \alpha$ is added to the clause set. This forces $f(t_1,\ldots,t_n)$ to be interpreted constantly over the variables occurring in $t_1,\ldots,t_n$. The method from [4] in particular reduces satisfiability of finitely quantified clauses to satisfiability of some of its FDTs. For simplicity, we introduce here a simplified version of FDTs; a definition in full detail is given in the original article.

The FDT method operates over finitely quantified clause sets. A clause $C$ is said to be finitely quantified if it is of the form $C^\dagger \lor \bigvee_{i=1}^n x_i \notin D_i$ where:

1. The $x_i$ are pairwise disjoint and comprise exactly the variables below a BSFG operator in the clause $C^\dagger$.
2. The $D_i$ are finite sets of same-sorted domain elements.
3. $C^\dagger$ does not contain further predicate domains.

Consider a set of finitely quantified clauses $N$ with variables $x_1,\ldots,x_n$ quantified over $D_1,\ldots,D_n$. A domain partition $P$ of $N$ is defined by a partition of each $D_i$ into subdomains $D_1^{i_1},\ldots,D_k^{i_i}$. We now build the Finite Domain Transformation of $N$ for the domain partition $P$. For every choice $j = (j_1,\ldots,j_n)$ where $1 \leq j_i \leq k_i$, and every clause $C = C^\dagger \lor \bigvee_{i \in I} x_i \notin D_i$, the clause set $\text{FDT}(C,j)$ is given by the following procedure:

1. Initialize $D$ as the clause $C^\dagger \lor \bigvee_{i \in I} x_i \notin D_i^{j_i}$ and $M := \emptyset$.
2. While $D$ contains some BSFG operator, choose a fresh parameter $\alpha$ and a position $\pi$ minimal among those containing a BSFG operator. Add the clause $D[\pi] \equiv \alpha \lor \bigvee_{i \in I} x_i \notin D_i^{j_i}$ to the clause set $M$, and set $D := D[\alpha]$.
3. Let $\text{FDT}(C,j) = \{D\} \cup M$.

The FDTs of $N$ are defined for every domain partition $P$ as:

$$\text{FDT}(N,P) = \bigcup_{1 \leq i \leq n} \bigcup_{C \in P} \text{FDT}(C,(j_1,\ldots,j_n))$$

Example 1. Consider the clause $C$ given by $f(x) < 0 \lor x \notin [-2..-1]$ and a domain partition given by the partition $[-2..-1] \cup [0..2] = [-2..2]$ for $x$. Then $\text{FDT}(C,P)$ contains the clauses:

$$\alpha < 0 \lor x \notin [-2..-1]$$
$$f(x) \equiv \alpha \lor x \notin [-2..-1]$$
$$\beta < 0 \lor x \notin [0..2]$$
$$f(x) \equiv \beta \lor x \notin [0..2]$$
FDTs are used in [4] to build a search procedure over the domains $D_1, \ldots, D_n$ that reduces satisfiability of $N$ to satisfiability of $\text{FDT}(N, P)$ for some domain partitions $P$. Refutational completeness of FDTs would thus yield a refutationally complete procedure for finitely quantified problems. In the original article it was stated as a theorem that FDT produces sufficiently complete clause sets, which implies refutational completeness of HSP for FDTs by Theorem 1. However, this claim is actually false.

**Example 2.** Consider a hierarchic specification using the linear integer arithmetic (LIA) BG specification, i.e. the integer logic where only addition and inequalities are allowed. Let $\xi_B$ be the integer BG sort and $\xi_F$ a freely interpreted FG sort. Consider $\xi_F$-sorted constants $a, b$ and operators $f, g$ of sort $\xi_B \to \xi_F$ and $\xi_F \to \xi_B$ respectively. Let $N = \{a \neq b, \ g(f(x)) \neq x \vee x \notin [1..2]\}$. For a domain partition $P$ taking $[1..2]$ as the only subdomain for $[1..2]$, we obtain the following clauses in $N' = \text{FDT}(N, P)$:

$$
a \neq b
\alpha \neq x \vee x \notin [1..2]
g(f(x)) \approx \alpha \vee x \notin [1..2]$$

Let $J$ be an extension of the standard model for LIA to a domain $\mathbb{Z} \cup \{\infty\}$ for the integer sort $\xi_B$, completed with a carrier set $J^{\xi_F} = \{a^*, b^*\}$ and the following interpreted operators: $\alpha^J = 0$, $a^J = a^*$, $b^J = b^*$ and:

- $f^J(x) = a^*$ for all $x \in \mathbb{Z} \cup \{\infty\}$.
- $g^J(a^*) = 0$ and $g^J(b^*) = \infty$.

Then, $J \models N'$ (in particular $J \models \text{sgi}(N')$), and $J \models \text{GndTh}(\mathcal{M}_B)$. Nevertheless, $\infty$ can only be obtained as the interpretation of a BSFG term (e.g. $g(b)$), but no BG term is interpreted to $\infty$ under $J$. Thus, $N'$ is not sufficiently complete. This contradicts the aforementioned claim.

### 4 The FG-ground, BSFG-safe fragment

The previous example shows that sufficient completeness is not achieved in general for FDTs, so refutational completeness for FDTs is not guaranteed by Theorem 1. However, we can modify the proof for this theorem to extend it to a new class of problems including the ones produced by FDT.

A term is said to be **BSFG-free** if it contains no BSFG operators; and **BSFG-headed** if it is of the form $f(t_1, \ldots, t_n)$ where $f$ is a BSFG operator and the $t_i$ are all BSFG-free. BSFG-free literals and clauses are defined analogously. A literal is a **BSFG-definition** if it is of the form $f(t_1, \ldots, t_n) \approx \alpha$, where $f(t_1, \ldots, t_n)$ is a BSFG-headed term and $\alpha$ is a parameter. A clause is said to be **BSFG-safe** if it only contains BSFG-free literals and BSFG-definitions. Furthermore, we say that a term, literal or clause is **FG-ground** if it contains no FG variable.

Our results are restricted to FG-ground clause sets. For such inputs, we show that FDTs belong to a fragment where HSP is refutationally complete. In
particular, the following result that follows straightforward from the definition of FDTs.

**Proposition 1.** Let $N$ be an FG-ground, finitely quantified clause set, and $P$ a domain partition for the finite domains in $N$. Then, $\text{FDT}(N, P)$ is FG-ground, BSFG-safe.

We obtain the following results showing that the FG-ground, BSFG-safe fragment is stable under the operations of the HSP calculus. This means that, from an input FG-ground clause set, the clause set obtained by saturation by HSP from its FDTs belong to the FG-ground, BSFG-safe fragment.

**Proposition 2.** Let $N$ be an FG-ground, BSFG-safe clause set. Then,

1. If $N'$ is obtained from $N$ by weak abstraction and HSP inferences, then $N'$ is FG-ground, BSFG-safe.
2. If $N'$ is obtained from $N$ by the simplification rule applying demodulation by unit clauses, tautology deletion, elimination of subsumed clauses, arithmetic simplification, unabstraction of domain elements or negative unit simplification, then $N'$ is FG-ground, BSFG-safe.

We have not included the proof for this result due to lack of space and interest, since it is rather long and purely syntactical. The FG-ground, BSFG-safe fragment is not stable under the general simplification rule described in [5]. Nevertheless, this rule is much more general than the actual application cases. In practice, HSP-based solvers only apply the techniques mentioned in Proposition 2.

### 5 Refutational completeness

Our approach is based on a modification of the proof of Theorem 1 found in [5]. The following definitions are introduced there for an arbitrary BG model $I \in \mathcal{M}_B$ and a saturated clause set $N$.

- $m(t) = \min_s \{s \mid s$ is a ground BG term and $|I| = t \approx s\}$
- $E'_I = \{t \rightarrow m(t) \mid t$ is a ground BG term and $t \neq m(t)\}$
- $E''_I = \{l \rightarrow r \in E'_I \mid l$ is not reducible by $E' \setminus \{l \rightarrow r\}\}$
- $E_I = \{l \approx r \mid l \rightarrow r \in E''\}$
- $D_I = \{l \neq r \mid l, r$ are ground BG terms with different $E_I$-normal forms\}$
- $A_I = \{C\sigma \in \text{sg}(N) \mid C \in N$ and $\sigma$ is a simple, $E''_I$-reduced substitution\}$
- $N_I = A_I \cup E_I \cup D_I$

Both the canonical term rewriting system $E''_I$ and the clause set $E_I \cup D_I$ force equality over ground BG terms to behave as in $I$; they can be thought of as encodings of $I$ into ground first-order clauses.

The following is an abstraction of the steps in the proof of Theorem 1 in [5].
1. Assume that $N$ does not contain $\Box$. It is shown that a BG model $I$ satisfying
the BG clauses in $N$ exists. The existence of one further term-generated
interpretation $J$ satisfying $N_I$ is then proved.

2. Since $N$ is sufficiently complete and satisfies $\text{sg}(N) \cup \text{GndTh}(\mathcal{M}_B)$, BG-
completeness of $J$ is shown.

3. Since $J$ is BG-complete and term-generated, from $J \models C_I$ follows that
$J \models N$.

4. Since $J$ is BG-complete and term generated, from $J \models E_I \cup D_I$ follows that
$J$ is a conservative extension of $I$.

5. Thus, $J$ an FG model of $N$, so $N$ is satisfiable with respect to the hierarchic
semantics.

Note that the notion of BG-completeness is not explicitly defined in the
original work, so steps 3 and 4 in the original proof seem to rely on sufficient
completeness of $N$; nevertheless, it suffices to consider that $J$ is BG-complete.
Our proof for the FG-ground, BSFG-safe fragment exploits this abstraction. In
particular, in this proof sufficient completeness is exclusively required to show
that $J$ is BG-complete in step 2.

The idea behind our proof is to modify step 2 by “rewiring” $J$ into a term-
generated, BG-complete interpretation $K$ that satisfies $N_I$. Hence, the proofs in
steps 3 and 4 are valid for $K$, showing that $K$ is an FG model of $N$. Let $J^\xi$ be
the carrier set in $J$ for every sort $\xi$. Carrier sets in $K$ are defined by:

$$K^{\xi} = \{ o \in J^{\xi} \mid \text{exists a ground BG term } t \text{ with } t^I = o \}$$

if $\xi$ is a BG sort, and $K^{\xi} = J^{\xi}$ if $\xi$ is an FG sort. This is: we drop elements
from BG carrier sets that prevent $J$ from being BG-complete. This requires some
modification on the interpretation of operators, since now they might not be
well-defined. In this case, there are no constraints on how interpreted operators
for $K$ are remapped in undefined values. In particular, we can choose an element
$q_\xi \in K^{\xi}$ for every sort $\xi$, and define for every operator $f$ with sort $\xi_1 \ldots \xi_n \rightarrow \xi_0$:

$$f^K(o_1, \ldots, o_n) = \begin{cases} f^I(o_1, \ldots, o_n) & \text{if } f^I(o_1, \ldots, o_n) \in K^{\xi_0} \\ q_\xi & \text{if } f^I(o_1, \ldots, o_n) \notin K^{\xi_0} \end{cases}$$

The next result follows trivially from the definition of $K$.

**Lemma 1.** $K$ is a BG-complete interpretation. Furthermore, if $J$ is term-
generated, then so is $K$.

**Example 3.** Consider the clause set $N'$ and the interpretation $I$ given in Example 2. The only element in the $\xi_B$ sort carrier set violating BG-completeness is
$\infty$. Thus we can drop $\infty$ from $J^{\xi_B}$ and remap $g^I(b)$ to, for example, 7. This
yields the interpretation $K$ with carrier sets $K^{\xi_B} = \mathbb{Z}$ and $K^{\xi_F} = \{ a^*, b^* \}$ and
interpreted operators: $\alpha^K = 0$, $\alpha^K = \alpha^*$, $\beta^K = \beta^*$ and

- $f^K(x) = a^*$ for all $x \in \mathbb{Z}$.
- $g^K(a^*) = 0$ and $g^K(b^*) = 7$. 

---

32
Then, $K$ is a BG-complete, term-generated interpretation that satisfies $N'$. Furthermore, $K$ is a conservative extension of the standard model of the LIA specification, which shows satisfiability of $N'$ with respect to the hierarchic semantics.

We now need to show that $K$ satisfies the conditions in $J$ that lead to show that $J \models N$ in step 3 and $J$ is a conservative extension of $I$ in step 4. That is, we prove that $K$ satisfies $E_I \cup D_I \cup C_I = N_I$. For this we need the following technical result.

**Lemma 2.** 1. If $t$ is a ground BG term, then $t^J = t^K$.
2. If $t$ is a ground BSFG-free term, then $t^J = t^K$.
3. If $C$ is a ground BSFG-safe clause, then $J \models C$ implies $K \models C$.

**Proof.** 1. We proceed by structural induction on $t$. If $t$ is a $\xi$-sorted constant, then since $t$ is a ground BG term, $t^J \in K^\xi$, so $t^J = t^K$. Assume now that $t = f(t_1, \ldots, t_n)$ where the Claim 1 holds for $t_1, \ldots, t_n$, which are ground BG. Thus, if the sort of $f$ is $\xi_1 \ldots \xi_n \rightarrow \xi_0$ we have

$$f^J(t_1^K, \ldots, t_n^K) = f^J(t_1^J, \ldots, t_n^J) = t^J \in J^{\xi_0}$$

and we conclude $t^K = t^J$. Claim 1 follows by induction.

2. Once more, we show Claim 2 by structural induction on $t$. If $t$ is a constant, then it is either an FG-sorted constant or a BG constant. In the former case, $t^K = t^J$ trivially. In the latter case, the same holds by Claim 1.

Assume now that $t = f(t_1, \ldots, t_n)$ where Claim 2 holds for $t_1, \ldots, t_n$ and $f$ is of sort $\xi_1 \ldots \xi_n \rightarrow \xi_0$. Since $t$ is BSFG-free, so are the $t_i$. If $t$ is FG-sorted, then

$$t^K = f^K(t_1^K, \ldots, t_n^K) = f^K(t_1^J, \ldots, t_n^J) = f^J(t_1^J, \ldots, t_n^J) = t^J$$

Otherwise, $t$ is BG-sorted, and since $t$ contains no BSFG operators, in particular $t$ is a BG term. Then, $t^K = t^J$ by Claim 1.

3. We show here that, for every ground BSFG-safe literal $L$, $J \models L$ implies $K \models L$; Claim 3 follows then trivially. If $L$ is BSFG-safe, then it is either BSFG-free or a BSFG definition. In the case when $L$ is BSFG-free, $L$ is of the form $L = (\neg) t \approx s$ for BSFG-free terms $t, s$. Then by Claim 2 we have $t^K = t^J$ and $s^K = s^J$, and hence $K \models L$ if and only if $J \models L$.

We now show the case when $L$ is a BSFG definition. This means that $L$ is an equation of the form $\alpha \approx t$ where $\alpha$ is a parameter and $t = f(t_1, \ldots, t_n)$ is a BSFG-headed term, so the $t_i$ are BSFG-free terms. By Claim 2, we have $t^K = f^K(t_1^J, \ldots, t_n^J)$ and $\alpha^K = \alpha^J$. Since $J \models L$, then $t^J = f^K(t_1^J, \ldots, t_n^J) = \alpha^J \in K^\xi$ where $\xi$ is the sort of $\alpha$ and $t$. We conclude

$$t^K = f^K(t_1^K, \ldots, t_n^K) = f^J(t_1^J, \ldots, t_n^J) = t^J = \alpha^J = \alpha^K$$

so $K \models L$ as we wanted. $\square$
We can now show that $K$ satisfies the same properties as $J$ did in the proof sketch above for the case when $N$ is FG-ground and BSFG-safe.

**Theorem 2.** If $N$ is an FG-ground, BSFG-safe clause set and $J \models N_I$, then $K \models N_I$.

**Proof.** We show that $N_I$ is a set of ground BSFG-safe clauses. The result then follows by Claim 3 in Lemma 2. Let $C$ be a clause in $N_I = E_I \cup D_I \cup C_I$. If $C \in E_I \cup D_I$, then $C$ is a ground BG clause; in particular, $C$ is ground BSFG-safe.

Otherwise, $C \in C_I$. Then, a simple, $E''_I$-reduced substitution $\sigma$ and a clause $C'$ exist such that $C = C' \sigma$ is ground. Let $L = (\neg)s \approx t$ be an arbitrary literal in $C$. Then, there is a literal $L' = (\neg)s' \approx t'$ in $C'$ such that $s' \sigma = s$ and $t' \sigma = t$. In particular, since $C'$ is FG-ground and BSFG-safe, so is $L'$.

- If $L'$ is a BSFG-free literal, then $s'$ is an FG-ground, BSFG-free term. Thus $s'$ contains no occurrence of BSFG operators or FG variables. Now, $s = s' \sigma$ only contains symbols occurring in either $s'$ or in $x \sigma$ for variables $x$ occurring in $s'$. However, all variables in $s'$ are BG-sorted, and since $\sigma$ is a simple substitution, the $x \sigma$ are BG terms; in particular, they are BSFG-free, therefore so is $s$. The same argument shows that $t$ is BSFG-free. Thus, $L$ is BSFG-free.

- If $L'$ is a BSFG definition, then $s'$ is a parameter and $t'$ is of the form $f(t'_1, \ldots, t'_n)$, where $f$ is a BSFG operator and the $t'_i$ are BSFG-free. Since $s'$ is ground, in particular $s = s'$. Furthermore, $t = f(t'_1 \sigma, \ldots, t'_n \sigma)$. The same argument as above shows that the $t'_i \sigma$ are BSFG-free, which in particular implies that $t$ is a BSFG-headed term, thus showing that $L$ is a BSFG definition.

Thus, $L$ is a ground BSFG-safe literal. Since the choice of $L$ in $C$ was arbitrary, $C$ is a ground BSFG-safe clause. Lemma 2 then shows that $J \models C$ implies $K \models C$. \qed

We can now rewrite the proof sketch above as follows to show refutational completeness for the FG-ground, BSFG-safe fragment:

1. Assume that $N$ does not contain $\Box$. It is shown that a BG model $I$ satisfying the BG clauses in $N$ exists. The existence of one further term-generated interpretation $J$ satisfying $N_I$ is then proved.
2. Obtain a term-generated, BG-complete interpretation $K$ from $J$ by rewiring.
3. Since $N$ is FG-ground, BSFG-safe, $K \models N_I$ follows from Theorem 2.
4. Since $K$ is BG-complete and term-generated, from $K \models C_I$ follows that $J \models N$.
5. Since $K$ is BG-complete and term generated, from $K \models E_I \cup D_I$ follows that $K$ is a conservative extension of $I$.
6. Thus, $K$ an FG model of $N$, so $N$ is satisfiable with respect to the hierarchic semantics.

We have thus proved our main result:
Theorem 3. Let $S$ be the class of weakly abstracted, FG-ground, BSFG-safe clause sets. If the hierarchic specification is compact, then HSP is refutationally complete for $S$.

In particular, by Propositions 1 and 2 we obtain the following corollary that partially fixes the problem posed by discovering that in general FDTs are not sufficiently complete. This is done by showing that they belong nevertheless to a refutationally complete fragment for HSP, namely the FG-ground, BSFG-safe fragment, provided that the original clause set is FG-ground.

Corollary 1. HSP is refutationally complete for FDTs generated from finitely quantified FG-ground clause sets.

The requirement for the original clause set $N$ to be FG-ground becomes apparent in the proof of Theorem 2. Clauses in $C_I$ are simple ground instances of the clause set $N'$ obtained by FDT and saturation by HSP. In the case when FG-sorted variables occur in $N'$, we cannot guarantee that simple ground instances of $N'$ are BSFG-safe, since an FG-sorted variable might be mapped by a simple substitution to a term containing BSFG operators. In particular, this implies that Theorem 2 might not be true for the case where FG-variables are allowed. However, we have not found a counterexample to the generalization of Theorem 3 where the condition on FG-groundness is omitted. We believe this generalization to be true, but further research is needed to attain a proof.

6 Conclusions and future work

Hierarchic Superposition is a powerful calculus for many-sorted first-order logic with background theories. Applications for Hierarchic Superposition include formal verification of software and processes, which often requires to detect unsatisfiability of quantified formulae.

Completeness results for Hierarchic Superposition guarantee refutational completeness in the sufficiently complete fragment. However, sufficient completeness is a rather complex property that is often missing in input clause sets. Several techniques based on definition introduction have been developed to circumvent this problem. In particular, an application of Hierarchic Superposition to finitely quantified theorem proving relies on iteratively solving problems where definitions are introduced for background-sorted foreground operators. The generated problems are called Finite Domain Transformations, and were claimed to be sufficiently complete.

We found a counterexample to this claim, this is, we provide problems whose Finite Domain Transformations do not lie in the sufficiently complete case. Thus, refutational completeness was not guaranteed in principle for Finite Domain Transformations. Without a proof for refutational completeness of Hierarchic Superposition for such problems, recent advances in the field would be ineffective, since they rely on this feature.

Nevertheless, such counterexamples only show that Finite Domain Transformations do not belong to a given refutationally complete fragment, but this does
not mean that Hierarchic Superposition is not refutationally complete for them. We define a new fragment, that of FG-ground BSFG-safe formulae, that contains all Finite Domain Transformations generated for inputs without foreground-sorted variables. We show this fragment to be stable under the operations applied by current implementations of Hierarchic Superposition.

Finally, we show that Hierarchic Superposition is refutationally complete for the FG-ground BSFG-safe fragment. This is, for saturated clause sets in this fragment and compact background theories, unsatisfiability detection can be reduced to detect whether the empty clause belongs to the clause set. In particular, we obtain refutational completeness for Finite Domain Transformations from clause sets without foreground-sorted variables. This represents a partial solution to the problem we found with our counterexample.

We believe that an extension of our results to the general BSFG-safe fragment is possible, since we have not been able to find a counterexample to this conjecture. The extension to allow foreground variables is a prioritary path for further research. A positive answer to this problem would yield full refutational completeness for Hierarchical Superposition in finitely quantified theorem proving through the use of Finite Domain Transformations.

References

Implicit Conflicts in
Abstract Argumentation

Christof Spanring
University of Liverpool, Department of Computer Science, Agent ART Group
TU Wien, Faculty of Informatics, Database and Artificial Intelligence Group

Abstract. Abstract argumentation refers to the art of dealing with arguments and directed attacks between arguments. The justification status of sets of arguments is defined via several well established principles, called semantics. Traditionally attacks are also called conflicts. However, as it turns out non-trivial semantics provide an implicit concept of conflict that does not coincide with and can not be expressed through attacks. This work is all about the why and how of implicit conflicts.

Keywords: abstract argumentation, implicit conflicts, argumentation semantics, expressiveness

1 Introduction and Motivation

This work deals with Dung-style abstract argumentation as first introduced in [6]. In a nutshell we use some structure \( F = (X, A) \) called argumentation framework (AF), consisting of a set of arguments \( X \) and an attack relation on these arguments \( A \subseteq X \times X \). Here arguments might refer to natural language arguments such as \( a = \text{"Death penalty is legit"} \) or \( b = \text{"God does not want us to kill"} \), whereby in this constellation naturally \( b \) attacks \( a \) but not the other way around. In graphical representations we identify nodes with arguments and directed edges with attacks.

Example 1. Consider the AF \( F = (X, A) \) as graphically represented in Figure 1. We have \( X = \{a, b, c, d\} \) and \( A = \{(b, a), (c, b), (d, c), (c, d)\} \).

![Fig. 1. Natural Language Example, Is Death Penalty Legit?](image-url)
To elaborate on a meaning of truth we use argumentation semantics (see [1] for an established selection of such). Formally an argumentation semantics is a mapping $\sigma$, assigning sets $\sigma(F) \subseteq \mathcal{P}(X)$. The intention being that for $S \in \sigma(F)$ we have that $S$ represents a collection of arguments justified by some desired principles, depending only on the attack relation and not on names or some underlying meaning of the arguments.

The sets $S \in \sigma(F)$ are called $\sigma$-extensions of $F$. For extensions we are interested in such principles as consistency (conflict- or attack-freeness), admissibility (self-defense), maximality, stability (assigning to each argument exactly one status of either accepted or defeated), range-maximality (maximal area of influence), and directionality (prefer arguments that occur earlier in directed argumentation paths).

**Example 2.** We again consider the AF $F$ from Example 1, and require for some semantics $\sigma$ to provide consistency, admissibility and maximality. Now for $a$ to be accepted we need to defeat $b$ and thus accept $c$ and also defeat $d$. On the other hand for if we accept $d$ then $c$ is defeated, $b$ is accepted and $a$ is rejected. Thus we have $\sigma(F) = \{\{a, c\}, \{b, d\}\}$.

Now observe that $a$ and $d$ do not occur together in any of these two extensions, despite none of them attacking the other. This is what we will call an implicit conflict. In this case on the abstract argumentation semantics level we are able to resolve this conflict by explicitly adding an attack $(d, a)$. However, observe that the legal questions surely do not depend on some random people’s belief. Some gods might even support death penalty.

This objection aside, on a purely abstract level, however, the question occurs whether implicit conflicts can always be expressed explicitly through modification of the attack relation without altering the given extension-sets.

Quite some effort has already been put into investigating relations between various argumentation semantics, see e.g. [2, 9, 8, 7]. In this work we discuss answer questions recently posed in [4] and answered in [10], and further investigate meaning and existence conditions of implicit conflicts.

It is easy to see that for some AFs some conflicts are implicit. For (maximal) conflict-free sets it is also easy to see that implicit conflicts occur only in relation with self-attacking arguments. In [4] an AF was presented showing that for maximal admissible and range-maximal admissible semantics and some AFs there is no semantically equivalent AF with only explicit conflicts. Hence some conflicts are implicit by their very nature and can not be explicitly represented via attacks.

Section 2 formally introduces all the required definitions. Section 3 presents results for implicit and explicit conflicts. Section 4 closes with a final discussion.

## 2 Background

Let us first recall the basic definitions as introduced in [6].
Definition 1. An Argumentation Framework (AF) is an ordered pair \( F = (X, A) \) where \( X \) is a set of arguments and \( A \subseteq X \times X \) represents the attack relation. For \( x, y \in X \) and \((x, y) \in A\) we say that \( x \) attacks \( y \) in \( F \) and write \( x \rightarrow_F y \). For \( x \in X \) and \( Y \subseteq X \) we say that \( x \rightarrow_F Y \) (or \( Y \rightarrow_F x \)) if there is some \( y \in Y \) such that \( x \rightarrow_F y \) (or \( y \rightarrow_F x \)), analogous for \( Y, Z \subseteq X \) and \( Y \rightarrow_F Z \). For \( Y \subseteq X \) we call \( Y^+_F = Y \cup \{x \in X \mid y \rightarrow_F x\} \) the range of \( Y \) in \( F \). If the referred to AF \( F \) is obvious from context we might drop the subscript \( F \) in above definitions.

Definition 2 (Semantics). A semantics \( \sigma \) is a mapping assigning to each AF \( F = (X, A) \) a collection of reasonable sets of arguments \( \sigma(F) \subseteq P(X) \), the set of \( \sigma \)-extensions of \( F \). The intention being that for \( S \in \sigma(F) \) we have that \( S \) represents a collection of arguments justified by some desired principles.

We now give a definition of commonly used argumentation semantics, that are conflict-free, admissible, naive, preferred, stable [6], stage [11], semi-stable [5] and CF2 semantics [3].

Definition 3. For AF \( F = (X, A) \) and \( S \subseteq X \) we define

- \( S \in cf(F) \) if \( \forall x, y \in S : x \not\rightarrow y \);
- \( S \in ad(F) \) if \( S \in cf(F) \) and \( \forall x \in X : x \rightarrow_S S \implies S \rightarrow x \);
- \( S \in na(F) \) if \( S \in cf(F) \) and \( \forall T \in cf(F) : S \subseteq T \implies S = T \);
- \( S \in pr(F) \) if \( S \in ad(F) \) and \( \forall T \in ad(F) : S \subseteq T \implies S = T \);
- \( S \in sb(F) \) if \( S \in cf(F) \) and \( S^+ = X \);
- \( S \in sg(F) \) if \( S \in cf(F) \) and \( \forall T \in cf(F) : S^+ \subseteq T^+ \implies S^+ = T^+ \);
- \( S \in ss(F) \) if \( S \in ad(F) \) and \( \forall T \in ad(F) : S^+ \subseteq T^+ \implies S^+ = T^+ \);
- \( S \in c2(F) \) if \( S \in na(F) \) and \( Y^+_F \cap Z = Y \oplus Z \implies S \cap Y \in c2(F) \).

Example 3. Again consider the AF \( F \) from Example 1. We have as extension-sets

\[
\begin{align*}
cf(F) &= \{\{a\}, \{b\}, \{c\}, \{d\}, \{a, c\}, \{a, d\}, \{b, d\}\}; \\
ad(F) &= \{\{c\}, \{d\}, \{a, c\}, \{b, d\}\}; \\
na(F) &= \{\{a, c\}, \{a, d\}, \{b, d\}\}; \\
pr(F) &= \{\{a, c\}, \{b, d\}\} = sb(F) = ss(F) = sg(F) = c2(F).
\end{align*}
\]

As we further on put emphasis on AFs that do or do not provide and extension-sets that allow and do not allow implicit conflicts the following definition from [10] will become useful.

Definition 4 (Analytic, Quasi- and Non-Analytic AFs). Given some AF \( F = (X, A) \), semantics \( \sigma \) and arguments \( x, y \in X \). If there is no \( S \in \sigma(F) \) with \( x, y \in S \), we say that \( x \) and \( y \) are in conflict in \( F \) for \( \sigma \). If \((x, y) \in A \) or \((y, x) \in A \) we say that the conflict \( \{x, y\} \) is explicit, otherwise the conflict is called implicit. An AF \( F = (X, A) \) is called analytic for \( \sigma \) if all conflicts of \( \sigma(F) \) are explicit in \( F \). \( F \) is called quasi-analytic if there is an analytic AF \( G = (X, A_G) \) such that \( \sigma(F) = \sigma(G) \). Finally \( F \) is called non-analytic if it is not quasi-analytic.
Example 4. For na and cf semantics and the AF \( F \) from Example 1 we have that all conflicts are explicit, hence \( F \) is analytic. For all the other introduced semantics \( \sigma \) however as mentioned above there is an implicit conflict between \( a \) and \( d \). However if we add an attack \((a,d)\), then again all conflicts become explicit, hence \( F \) is quasi-analytic for \( \sigma \).

3 Hunt for Non-Analytic AFs

In this section we investigate ways of enforcing implicit conflicts in abstract argumentation. In particular we are interested in for some semantics non-analytic AFs. According to our definitions, as shown in Example 4 every semantics provides quasi-analytic AFs. For cf and na however it is easily to be seen that implicit conflicts occur only in correlation with self-attacking arguments and thus every implicit conflict can be made explicit. Thus for cf and na every AF is either analytic or quasi-analytic.

For preferred and semi-stable semantics it is possible for extensions not to be naive extensions, which means that some extensions might not incorporate wide areas of the AF under consideration. We can use this knowledge for an implicit conflict that operates on defense. If we allow the conflict to become explicit some non-naive extension necessarily defends some argument it should not contain.

Example 5 (Preferred Semantics, Non-Analytic AFs). Take into account the AF \( F \) as shown in Figure 2. Here we have that the set \( E = \{a_1, a_2, a_3\} \) is a preferred extension. The implicit conflicts marked in dashed snake lines between the \( a_i \) and \( b_j \) can not be made explicit. In one direction because of \( E \) needing to be admissible, in the other direction because it would imply a defense of \( y_j \) by \( a_i \).

The principle used in this example can also be facilitated for semi-stable semantics. To this end for \( i \in \{1, 2, 3\} \) we can add (rejected) arguments \( \bar{a}_i \) and attacks \((\bar{a}_i, \bar{a}_i)\) and \((a_i, \bar{a}_i)\). This simple translation also indicates that in this case the AF being non-analytic is indeed a property of the effective extension-sets.
Any stable extension is maximal in range and admissible. We use the range-maximality to establish an implicit conflict, i.e. there is a naive extension that increases in range if we allow the conflict to become explicit. Observe that in the proof for stable semantics we also need to make use of admissibility, and consequently for stage semantics the investigated AF is quasi-analytic.

Example 6 (Stable Semantics, Non-Analytic AFs). For the AF from Figure 3 we have that \( pr, ss, sg, sb \) semantics coincide. Taking into account totality wlog. we can show that addition of the attack \((a,b)\) adds the formerly only naive extension \(\{a,v_1,v_2\}\). We can also show that if some AF is non-analytic for stable semantics where \(sb\) and \(pr\) coincide it is also non-analytic for preferred semantics, immediately yielding the same result for semi-stable semantics. However the AF used here is only quasi-analytic for stage semantics, i.e. there is an analytic AF with the same arguments and same stage extensions.

The range-technique from the previous example can be tightened, we can work around the admissibility condition by introducing additional arguments.

---

**Fig. 3.** Non-Analytic AF for Stable Semantics.

**Fig. 4.** Non-Analytic AF for Stage Semantics.
Fig. 5. Non-Analytic AF for CF2 Semantics.

However observe that the implicit conflicts of Examples 5 and 6 are of a very strong nature. In their proofs we only make use of required semantical principles, i.e. show that for the desired extension sets the given semantics can not allow an attack regardless of explicitness of other conflicts. For stage semantics it is a bit trickier in that we have to assume some analytic AF for a contradiction.

Example 7 (Stage Semantics, Non-Analytic AFs). The AF from Figure 4 is a slight variation of Example 6, with comparable extension set for all desirable semantics. First we show that this AF is still non-analytic for stable semantics and then, assuming that the implicit conflict \{a, b\} and all other conflicts can be made explicit we show that necessarily stage and semi-stable semantics coincide.

From a historical perspective semi-stable and stage semantics were introduced as a means to fight a weakness of stable semantics. For some AFs (e.g. directed cycle of an odd number of arguments) stable semantics collapses and gives no extension at all, even worse disconnectedly joining two AFs where for one of them stable semantics collapses results in a bigger collapsing AF. Semi-stable and stage semantics coincide with stable semantics where the latter does not collapse, yet give non-empty extensions whenever there are non-empty admissible and conflict-free, respectively sets.

CF2 semantics now was introduced on the observation of directionality, or SCC-recursiveness [3]. If an AF consists of two arguments x and y with x attacking y but not the other way around, it seems reasonable to prefer x over y as it necessarily occurs earlier in any chain of arguments. CF2 semantics facilitates this observation with x and y being subframeworks, or more accordingly strongly connected components (SCCs). Directionality is what also drives the following example. The basic idea here is that for the implicit conflict not being allowed to become explicit we force it to cut short strongly connected components in both directions.

Example 8 (CF2 Semantics, Non-Analytic AFs). Take the AF as graphically represented in Figure 5 as given. For c2 semantics there is an implicit conflict between x and y, as well as between x′ and y′. The latter however can easily be
made explicit. The former necessarily cuts short a strongly connected component, thus adding an unwanted naive extension.

4 Discussion

In this work we have presented an overview of which semantical properties can cause AFs to become non-analytic. As it turns out all slightly more sophisticated properties can show this behaviour, and thus once more na semantics is called naive for a reason.

Future work includes deeper analysis of whether further properties can result in substantially different non-analytic AFs, or whether the given techniques can serve to flesh out characterising properties of non-analytic AFs and subsequently analytic or quasi-analytic AFs.

Examples of applied abstract argumentation are still rare in the literature, however it might be interesting to see if some instantiations can guarantee analytic or quasi-analytic AFs. On the one hand analytic AFs provide substantial insight into computational properties. For instance, every argument that is not self-attacking in some analytic AF will be accepted by some extension. An result that usually proves to be computationally hard. On the other hand, as discussed here, allowing non-analytic AFs means increasing expressiveness. As further discussed in [10], this increased expressiveness does not hold for stable and stage semantics if we allow additional rejected arguments as those can be used to get rid of implicit conflicts.

It should be noted that the proofs for showing that the given AFs are non-analytic are non-trivial. While stable, semi-stable and preferred semantics allow (relatively spoken) straightforward proofs, stage semantics already requires a deeper understanding and high-level investigation. For cf2 semantics the state of the art is even worse and the easiest known version of a proof is of computerised nature, i.e. simply evaluating all $3^{13} = 1594323$ possible versions of analytic AFs given the set of extensions. This hardness is in parts due to the presented AFs being minimal with respect to the techniques of construction.

References


How flexible are categorical models of meaning?

Antonin Delpeuch
University of Cambridge and École Normale Supérieure

Abstract. The categorical semantics of type-driven grammars traditionally rely on tensor-based models, inspired by the foundations of quantum physics. The tensor product gives rise to very high-dimensional parameter spaces, and accommodates poorly with non-linearities, restricting its use in practical applications with distributional semantics. In this poster, we show how free autonomous categories can define alternate models of meaning. The free construction allows us to use the cartesian product despite its incompatibility with the yanking equalities. The models it yields are more flexible as they can include nonlinearities, making them comparable to convolutional neural networks. They also require far less parameters to represent word meanings, as their dimensions rise linearly instead of exponentially with the order of the word type.

1 Introduction

The distributional compositional model of meaning of Coecke, Sadrzadeh and Clark [10,3] is an attempt to provide a syntax-aware composition of distributional word vectors. The syntactic structure of a sentence is represented using a type-driven grammar such as a Pregroup Grammar or a Combinatory Categorial Grammar (CCG).

In type-driven grammars, words are given types. These types can be seen as rich algebraic part-of-speech tags that encode the syntactic roles of the words. The structure of a sentence is embodied by a derivation tree witnessing that the product of the word types reduces to the type of a sentence. From the perspective of category theory, we can see this derivation as a function mapping the representation of individual words to the representation of the sentence. This allows for a syntax-driven composition of word vectors in a robust mathematical framework.

The semantics of these type-logical grammars are defined with autonomous categories. This structure ensures useful properties on the semantics, but introduces a constraint: it is not clear how to construct the appropriate category with a particular model of meaning in mind.

In this poster, we show how any monoidal category freely generates an autonomous category. This expands the range of models of meaning that can be used to any monoidal category, and not just any autonomous category. To build the free autonomous category, we use string diagrams as the arrows themselves.
2 Monoidal and autonomous categories

We briefly recall the notion of monoidal and autonomous categories, the mathematical structures we use to define our semantics.

Definition 1 A (strict) monoidal category $\mathcal{C}$ is:

- a collection of objects $\text{Ob} \mathcal{C} = \{A, B, \ldots\}$
- for all pair of objects $A, B$ a collection of morphisms $\mathcal{C}(A, B) = \{f, g, \ldots\}$
- for all object $A$ an identity morphism $1_A \in \mathcal{C}(A, A)$
- a composition operation $\circ : \mathcal{C}(A, B) \times \mathcal{C}(B, C) \to \mathcal{C}(A, C)$, associative and with identities as neutral elements
- a monoid operation $\otimes$ on objects, with the object $I$ as neutral element
- for all objects $A, B, C, D$ an operation $\otimes : \mathcal{C}(A, B) \times \mathcal{C}(C, D) \to \mathcal{C}(A \otimes C, B \otimes D)$, such that $1_A \otimes 1_B = 1_{A \otimes B}$ for all objects $A, B$

such that the following equation is satisfied when the compositions are defined:

$$(f_1 \otimes g_1) \circ (f_2 \otimes g_2) = (f_1 \circ f_2) \otimes (g_1 \circ g_2) \quad (1)$$

This equation is best explained using the graphical language introduced in [11]. An arrow $f : A \to B$ is represented by the leftmost diagram below. When the domain (respectively the codomain) is the monoidal unit $I$, we depict $f$ as a box without input (respectively without output) and give it a triangular shape. The diagrams for composite arrows are defined as follows:

\[
\begin{array}{c}
A \\
\downarrow^f \\
B \\
\end{array}
\quad = \quad
\begin{array}{c}
A \\
\downarrow^g \\
B \\
\end{array}
\quad = \quad
\begin{array}{c}
A \\
\downarrow^f \\
B \\
\end{array}
\quad = \quad
\begin{array}{c}
A \\
\downarrow^g \\
B \\
\end{array}
\]

With these conventions, the equation (1) takes a very simple form:

\[
\begin{pmatrix}
(f_2 \\
(f_1)
\end{pmatrix}
\cdot
\begin{pmatrix}
(g_2 \\
(g_1)
\end{pmatrix}
= \begin{pmatrix}
(f_2 \\
(f_1)
\end{pmatrix}
\cdot
\begin{pmatrix}
(g_1 \\
(g_2)
\end{pmatrix}
\]

Hence, monoidal categories give a way to compose processes (morphisms) and evaluate them in parallel. They are not sufficient to give semantics to type-logical grammars: we need the notion of adjoint.
Definition 2 In a monoidal category, an object $A$ is a left adjoint of $B$ (and $B$ is a right adjoint of $A$) when there are two morphisms $\epsilon : A \otimes B \to I$ (the counit) and $\eta : I \to B \otimes A$ (the unit) such that

$$(1_B \otimes \epsilon) \circ (\eta \otimes 1_B) = 1_B \quad \text{and} \quad (\epsilon \otimes 1_A) \circ (1_A \otimes \eta) = 1_A \quad (2)$$

Again, we can use diagrams to make these equations clearer, provided we use appropriate representations for $\epsilon$ and $\eta$:

\[ \begin{array}{ccc}
A & \rightarrow & B \\
\downarrow & & \downarrow \\
I & \rightarrow & I \\
\end{array} \quad \begin{array}{ccc}
A & \rightarrow & I \\
\downarrow & & \downarrow \\
B & \rightarrow & A \\
\end{array} \quad \begin{array}{ccc}
I & \rightarrow & A \\
\downarrow & & \downarrow \\
A & \rightarrow & A \\
\end{array} \]

Definition 3 A monoidal category is autonomous when every object has a left and a right adjoint, denoted respectively by $A^l$ and $A^r$.

3 The semantics of pregroup grammar

Autonomous categories define a semantics for pregroup grammars. The word types are objects in the autonomous category and proof trees are morphisms built with $\epsilon$, $\eta$ and identities.

For instance, let us analyze the sentence *Mary likes ships*. We set $N$ and $S$, objects in an autonomous category $C$, to be the types of noun phrases and sentences respectively. The words *Mary* and *ships* are noun phrases so we give them the type $N$. The verb *likes* produces a sentence when combined with a subject on its left and a verb on its right, so we can give it the type $N^r \otimes S \otimes N^l$.

We can use the type reduction to compute the semantics of the sentence by composing it with the individual word meanings $\bar{\text{Mary}} : I \to N$, $\bar{\text{likes}} : I \to N^r \otimes S \otimes N^l$ and $\bar{\text{ships}} : I \to N$, represented in the upper part of the diagram. Finally we get a morphism $s : I \to S$ encoding the “meaning” of the sentence.
The traditional autonomous category

The most popular semantic category for distributional models is the autonomous category of finite-dimensional vector spaces and linear maps between them, denoted by $\text{FinVect}_R$, with the tensor product $\otimes$ as monoidal operation [2, 3].

The choice of this category was motivated in [3] by the observation that tensors allow to encode correlations between arguments. For instance, let us assume that a transitive verb is represented by a linear map $N \otimes N \rightarrow S$, or equivalently a tensor $N \otimes S \otimes N$ as in the previous section.\(^1\) The nature of the tensor product allows to give plausible sentence values to $\text{UNESCO runs conferences}$ and $\text{Usain Bolt runs 100m}$ and in plausible ones to $\text{UNESCO runs 100m}$ and $\text{Usain Bolt runs conferences}$. If arguments are combined with the direct sum $\oplus$ instead, which would mean that a transitive verb would be represented by a function $N \oplus N \rightarrow S$, linearity forbids these correlations. We would have $\text{runs(UNESCO, conferences)} = \text{runs}_S(\text{UNESCO}) + \text{runs}_V(\text{conferences})$ by linearity of the verb function. This is not desirable, for instance because it entails such an identity:

$$\text{UNESCO runs conferences} + \text{Usain Bolt runs 100m} = \text{Usain Bolt runs conferences} + \text{UNESCO runs 100m}$$

The major challenge with the tensor-based model is the number of parameters to learn, as the dimensions of the word vectors grow exponentially with the length of their type. Indeed, wide coverage grammars include many longer types, for instance for prepositions or adverbs, and learning tensors for them is challenging.

Putting aside these dimensionality issues, the model may not be adapted to combine vectors in a meaningful way. In some settings using neural networks to learn the representations [8], word vectors enjoy an additive structure: the authors report that $\text{King} \rightarrow \text{Man} + \text{Woman} \simeq \text{Queen}$. When this kind of relation holds, one might want to represent adjectives such as $\text{female}$ by a function adding the vector $\text{Woman} - \text{Man}$ to their argument, such that $\text{female king}$ would be close to $\text{queen}$ and $\text{female man}$ would be equal to $\text{woman}$. This is impossible in $\text{FinVect}_R$, because adjectives are linear maps and $v \mapsto v + \text{Woman} - \text{Man}$ is not linear but only affine.

More generally, the linearity assumption behind the tensor model is debatable. The success of models based on neural networks, for instance using distributional representations for machine translation [7], shows that non-linear compositions should not be ruled out. Furthermore, when maps are not required to be linear, the direct sum $\oplus$ (or cartesian product) allows to represent correlations, and yields much more compact representations than tensors.

Unfortunately, the category of finite-dimensional vector spaces equipped with the direct sum is only monoidal and not autonomous, because the direct sum is a cartesian product:

\(^1\) We use here the fact that for vector spaces, $N^i \simeq N \simeq N^r$. 

48
Proposition 1 In a cartesian monoidal category, no object $A \neq O$ has a (right) adjoint (relative to the cartesian monoidal structure).

Proof. We denote by $\oplus$ the cartesian product and $O$ the terminal object associated to it. Let $A$ and $B$ be objects, different from $O$, $\epsilon : A \oplus B \to O$ and $\eta : O \to B \oplus A$. As the category is cartesian, there are some $u : O \to B$ and $v : O \to A$ such that $\eta = u \oplus v$. We get

$$(\epsilon \oplus 1_A) \circ (1_A \oplus \eta) = (\epsilon \oplus 1_A) \circ (1_A \oplus u \oplus v) = (\epsilon \circ (1_A \oplus u)) \oplus (1_A \circ v) = 0_A \oplus v$$

As $A \neq O$, there is $v' : O \to A$ such that $v' \neq v$. Now $(0_A \oplus v) \circ v' = (0_A \circ v') \oplus v = 1_O \oplus v = v$ hence $0_A \oplus v \neq 1_A$. In pictures:

Therefore, this product cannot be used directly for semantics.

5 Turning monoidal categories into autonomous categories

In this section, we show how monoidal categories can be turned into autonomous categories, and why this construction is suitable to extend categorical models to alternate composition methods. We give an informal explanation of the construction, a mathematical formalisation can be found in [5].

The main idea is to define a category whose morphisms are diagrams themselves, where the nodes are labelled by morphisms from the original category. In these diagrams, we allow ourselves to use units and counits even when there is no corresponding morphism in the original category.

For instance, given a morphism $f : A \times B \to C$ in a monoidal category $(\mathcal{C}, \times, I)$, we can define the following morphism $\tilde{f} : 1 \to A^c \cdot C \cdot B^l$ in the free autonomous category $(\tilde{\mathcal{C}}, \cdot, 1)$ generated by $\mathcal{C}$:

$$\tilde{f} := \begin{array}{c}
1 \\
\tilde{\mathcal{C}} \cdot C \cdot B^l
\end{array}$$
Intuitively, this morphism represents the original function $f$, with the inputs converted to outputs. In this case, the tensor $(A^r, C, B^l)$ is only a symbolic object. In this state, the morphism is not very useful as we cannot use it in numerical computations. However, we can use it formally to define semantics that will eventually become numerical, thanks to the following theorem:

**Theorem 1 (Blute et al. [1])** Let $f$ be a morphism in the free compact closed\(^2\) category generated by $\mathcal{C}$. If its domain and codomains are products of objects of $\mathcal{C}$, then it is equal\(^3\) to a morphism in $\mathcal{C}$.

As the sentence semantics is defined by a morphism $1 \to S$, this theorem ensures that we can eliminate the formal units and counits as the domain and codomain are plain objects of $\mathcal{C}$.

Let us show how this can be used in practice. In the example above, our goal was to be able to sum distributional representations, so let us define the category $\text{Aff}$, whose objects are finite-dimensional vector spaces and morphisms are affine\(^4\) maps between them. We will equip this category with the direct sum $\oplus$ (or cartesian product) as monoidal operation. Concretely, this means that pairs of vectors are constructed by concatenating the vectors instead of taking their outer product. Let $N$ and $S$ be the vector spaces associated with noun phrases and sentences respectively. The sum operation of vectors in $N$ is a morphism in $\text{Aff}$: $+: N \oplus N \to N := (u, v) \mapsto u + v$. We can therefore analyze the sentence *Pat plants female seeds* as shown below:

In the left-hand shape, the diagram cannot be used to compute the representation of the sentence, because the units and counits do not correspond to

---

\(^2\) Compact closed categories are autonomous categories with a symmetric monoidal operation: for our purpose, we can ignore the difference as the semantic models are symmetric.

\(^3\) Technically speaking, this theorem is more accurately stated by saying that the left adjoint (free functor) in the free-forgetful adjunction is full.

\(^4\) An affine map is a map $f : x \mapsto f(x) + b$ where $f$ is a linear map and $b$ is a constant vector.
actual morphisms. Theorem 1 guarantees that we can eliminate them using the yanking equalities (2), and get the representation on the right-hand side, which is a valid diagram for a monoidal category. We get the expected representation for female: it adds a vector to its argument.

Similarly, we can now represent the $2\text{Mat}$ approach of [9], which uses the direct sum and hence could not be represented in the original framework:

We have applied this construction to allow for affine maps, but in fact it can be used with any monoidal category. In particular, we can even add nonlinearities, leading us to type-driven neural models of meaning. Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a nonlinearity. We define the category $\text{Net}_\sigma$ as the monoidal category generated by $\sigma$ and affine maps. Any neural network using $\sigma$ as nonlinearity can be written as a series of compositions and pairing of affine maps and $\sigma$ and are hence arrows in $\text{Net}_\sigma$.

(a) An arrow in $\text{Net}_\sigma$. $f$ and $g$ are affine maps.

(b) Its traditional representation as a neural network

Fig. 3: A categorical neural model of meaning in pictures
There are various reasons why using this model of meaning could be interesting. First, because it gives us some freedom in the functions we can use to combine meanings. The class of neural networks with one hidden layer can approximate any continuous function, with respect to the uniform norm on a compact set. This result has been obtained first by [4] for the case where $\sigma$ is the sigmoid, a popular choice of activation function ($\sigma(x) = \frac{1}{1+e^{-x}}$) and has been generalized to any non-constant, continuous and bounded function by [6]. Of course, we could also directly use the monoidal category of such functions and freely generate the autonomous structure on it, but for concrete applications we need a finite parametrization of the objects we consider, and neural networks provide one.

Second, this result provides a way to define the representation of a sentence as a neural network whose shape depends on the syntactic structure. This is analogous to the syntax-driven convolutional neural network of [12], where the sentence vector is recursively computed by neural networks at each node of the parse tree (using a Context Free Grammar).

6 Conclusion

We have shown that type-driven, compositional and categorical do not necessarily imply tensor-based, which opens up the framework to more economical and efficient representations. Some of them had already been considered as simplifications of the original model of meaning. Some others can be seen as adaptations of the popular neural models to type-driven syntax. As these models enjoy many of the theoretical properties presented in the seminal work of [3], this is an incentive to reconsider why the community focuses on tensor-based models.

Acknowledgments

This work would not have been possible without the helpful supervision of Stephen Clark, Tamara Polajnar and Anne Preller. I thank the reviewers for their useful comments and suggestions. Some parts of this article were taken from my thesis submitted for the degree of Master of Philosophy in Advanced Computer Science, University of Cambridge.

References

Minimizing Scope Ambiguity Hypothesis*

Milica Denić
École Normale Supérieure & Université Paris-Diderot
milica.denic@ens.fr

Abstract. The Bagel problem items and positive polarity items cannot be interpreted in the scope of clausemate sentential negation. This seems to be related to the existence of an alternative element with the same meaning: respectively, an n-word and a negative polarity item. This paper explores the hypothesis that this pattern could be explained through a competition between these elements whose nature is an attempt to avoid scope ambiguities.

1 Proposal

This paper is about the distribution of two types of polarity sensitive items (PSIs) with a similar empirical pattern: the Bagel problem items (BPIs) and the positive polarity items (PPIs). Namely, BPIs and PPIs cannot be interpreted in the scope of clausemate negation, where an alternative element with the same meaning has to be used: an n-word instead of a BPI, and a negative polarity item (NPI) instead of a PPI. We will explore the hypothesis that this pattern is due to the ability of BPIs and PPIs to move (overly or covertly) above negation and thus create scope ambiguities, while their alternative items cannot do so. This ability is to be understood as a syntactic ability to move from a position lower than negation to a position higher than it — BPIs and PPIs are syntactically able to outscope negation and thus create scope ambiguities with respect to it, regardless of their inverse scope over negation being semantically or pragmatically well-formed in a particular context. As they can move to a position higher than negation and they have alternative items with the same meaning that cannot do so, BPIs and PPIs are prohibited from its scope in an attempt to avoid scope ambiguities — we will refer to this hypothesis as the minimizing scope ambiguity hypothesis.

The paper is structured as follows. Section 2 discusses the n-word-BPI distribution and shows that it can be accounted for with the explored hypothesis. Section 3 does the same for the NPI-PPI distribution. Section 4 discusses the implementation of this hypothesis. Section 5 is the conclusion.

* This paper is based on the work done for my MA thesis supervised by Benjamin Spector and Caterina Donati. I am grateful to them for the discussions we had, which shaped the ideas presented here. I also thank Amir Anvari and Hagen Blix for reading the almost-final version of the paper. All errors are mine.

1 In this paper I am concentrating on local PPIs — see [11] for the global PPIs account.
2 The Bagel problem

The Bagel problem is a phenomenon consisting of the existence of certain NPIs that are unacceptable in the scope of clausemate sentential negation, while they are acceptable in other NPI-licensing contexts, such as the scope of superordinate negation, antecedents of conditionals etc. This is puzzling as downward-entailingness (DE) is considered to be the relevant semantic requirement for the NPI licensing [5]: negation, as an antimorphic operator, should be able to license every NPI in its scope.

This phenomenon seems to be connected with the existence of n-words in the language — n-words (negative concord items) have been argued to be existential indefinites that are obligatorily in the scope of clausemate negation with which they enter into an Agree relation [13].

BPIs that have been identified so far are Serbian i-words [12, 7], Polish kolwiek-pronouns [1], and Russian libo-items [6]. In these three languages every BPI has its n-word alternative: the generalization seems to be that in the scope of an operator that licenses an n-word the BPI is anti-licensed. Let us see the basic pattern in the three languages:

Serbian

(1) Nisam upoznala nikoga/*ikoga.
    neg.aux.1PS meet n-word-anybody/*BPI-anybody
    I haven’t met anybody.

Polish

(2) Ewa nie spotkala nikogo/*kogokolwiek.
    Ewa neg.aux meet n-word-anybody/*BPI-anybody
    Ewa didn’t meet anyone.

Russian

(3) Ja ne videl nikogo/*kogo-libo.
    I neg see n-word-anybody/*BPI-anybody
    I haven’t seen anybody.

As we can see in (1)-(3), in the three languages in the scope of clausemate negation only n-words are licensed. In Polish only n-words are licensed in the scope of the preposition without as well; however, if we accept the theory of negative concord as syntactic agreement [13], the fact that the preposition without

2 There are PSIs which have exclusively free choice (FC) usage (i.e. they lack NPI uses), such as Italian qualunque/qualsiasi, and (probably) French quiconque. As such they cannot be used in the scope of clausemate negation either. These items have been analysed in terms of proper strengthening [2]; it is important to distinguish this phenomenon from the Bagel problem, as BPIs are not exclusively FC items, i.e. they have regular NPI uses, but cannot be used the scope of clausemate negation.

3 In other NPI-contexts only BPIs are licensed.

4 In Serbian and Russian BPIs, but not n-words, can be used in the scope of without.
licenses n-words in Polish means that it contains a negation itself, so the scope of *without* in Polish is to be considered a subcase of the scope of clausemate negation.

To show that BPIs are able to scope above negation, we have to embed negation inside another DE environment — as BPIs are a type of NPIs, when they move above negation they need to be in the scope of a DE operator for semantic reasons. In the examples (4), (5) and (6), negation is in the antecedent of a conditional: as we can see from these examples, in the three languages the BPI is able to outscope negation — it is interpreted as scoping existentially between the conditional operator and negation.

```
Serbian
if ≫ ∃ ≫ ¬, *if≫ ¬ ≫ ∃
(4) Ako išta od toga ne razumeš, pozovi me.
if BPI-anything of this neg understand.2sg, call me
If there is anything you don’t understand, call me.
Polish
if ≫ ∃ ≫ ¬, *if≫ ¬ ≫ ∃
(5) Jesli nie rozumiesz czegokolwiek, po prostu zadzwoń do mnie.
If neg understand BPI-anything, just call to me
If there is anything that you don’t understand, just call me.
Russian
if ≫ ∃ ≫ ¬, *if≫ ¬ ≫ ∃
(6) Esli on ne znal čego-libo, on ne stesnjalsja sprašivat’ drugix.
if he neg know BPI-anything, he neg ashamed ask others
If there was something he didn’t know, he wasn’t ashamed to ask others.
```

The n-word cannot scope existentially above negation even when negation is embedded inside another DE environment, as we can see in (7):

```
Serbian
if ≫ ¬ ≫ ∃, *if≫ ∃ ≫ ¬
(7) Ako ne razumeš ništa od toga, pozovi me.
if neg understand.2sg n-word-anything of this, call me
If you understand nothing, call me.
```

According to the hypothesis explored here, as BPIs can move above negation and n-words cannot, the n-words block the use of BPIs in the scope of clausemate negation in order to avoid scope ambiguities. The sentences (1)-(3) thus end up ungrammatical with a BPI: the BPI is syntactically able to move above negation in these sentences, but when it does, the sentence is semantically deviant because the BPI is no longer in a DE environment. On the other hand, the sentences (4)-(6) only get the reading in which the BPI outscopes negation.

5 This example is taken from [6], footnote 9.
Why not a simple blocking account, rather than blocking due to scope options? A simple blocking account was proposed in [6]. According to this account, as n-words are more specific than BPIs (i.e. n-words are licensed exclusively by negation while BPIs need simply a DE context), when the conditions for the insertion of an n-word are met, the n-word blocks the insertion of a BPI. However, this account cannot explain why blocking does not occur in all negative concord languages. For example, Italian is a negative concord language, yet the n-word *nessun* does not block the use of an NPI *alcun*, as can be seen in (8):

(8) Gianni non ha letto *nessun/alcun* libro.

Gianni neg aux read n-word-any/any book

Gianni didn’t read any book.

Similar argument can be made in French. French *personne* is an n-word, while *qui que ce soit* is an NPI, and they can be used interchangeably in the scope of negation:

(9) Personne n’a vu *personne/qui que ce soit*.

n-word-anyone neg’aux see n-word-anyone/anyone

Nobody saw anyone.

If there was a morphological blocking rule that rules out NPIs in the scope of clausal negation because they are less specific in terms of featural content than n-words, it is not clear why this rule would not apply in Italian and French.

Let us now see how the hypothesis explored here stands with respect to the acceptability of *alcun* and *qui que ce soit* in the scope of clausal negation. Both Italian *alcun* and French *qui que ce soit* have the same scope properties with respect to negation as n-words: they cannot outscope it, as can be seen in (10) and (11). The minimizing scope ambiguity hypothesis thus correctly predicts them not to compete with n-words in the scope of clausal negation.

(10) Se Gianni non ha letto *alcun* mio libro, allora *farà*.

if Gianni neg aux read any my book, then fail.3sg

If Gianni hasn’t read any book of mine, he will fail.

(11) Il n’est pas vrai que Marie n’a pas lu *quoi que ce soit*.

It neg’is neg true that Mary neg’aux neg read anything

It isn’t true that Mary hasn’t read anything.

While it is clear why n-words cannot outscope negation once we accept the agreement theory of negative concord [13], I have no explanation why certain
NPIs can outscope negation while others cannot: this is just an empirical generalization. The prediction of the current account is that if a language has both n-words (or NPIs that cannot outscope negation) and NPIs that can outscope negation, then the NPIs that can outscope it will behave as BPIs. More research is needed to see whether this generalization holds cross-linguistically.

Let us now do away with a potential counter-example: subtrigging. Subtrigging is, descriptively, modifying an indefinite pronoun with an adjective, a prepositional phrase or a relative clause. Namely, Serbian *i*-words and Polish *kolwiek-pronouns* can appear in the scope of negation subtrigged — at that position both BPIs and n-words are acceptable with the same meaning.

Serbian

(12) Nisam upoznala ikoga/nikoga od tvojih prijatelja.
    neg.aux meet BPI-anyone/n-word-anyone of your friends
    I haven’t met any of your friends.

Polish

(13) Ewa nie spotkala kogokolwiek/nikoga z moich
    Ewa neg.aux meet BPI-anyone/n-word-anyone of my
    friends
    Ewa didn’t meet any of my friends.

I will follow [2] in assuming that what subtrigging does is to provide a modal component which allows certain polarity-sensitive items that have free choice usage to scope widely with a universal force (for the technical matters and explanation which free choice items can be subtrigged and which not, see [2]). As n-words have no free-choice usage, this does not apply to them. However, what this means for BPIs surfacing subtrigged under clausemate sentential negation is simply that at LF they are no longer in its scope — for example, the sentence (13) will have the logical form in (14):

(14) \[[kogokolwiek z moich znajomych] [Ewa nie spotkala t, ]\]

So, when subtrigged, BPIs scope with universal force above negation: the reason why they appear to scope existentially below negation in those cases is

6 While the free-choice usage of Polish *kolwiek-pronouns* has already been described [1], the free choice usage is not straightforward for Serbian *i*-words, as they are unacceptable in the scope of possibility modals etc. However, they can be used in the scope of imperatives, which suggests that they are not restricted to the NPI-contexts only, as can be seen in (1). Nevertheless, what rules them out in other free choice contexts remains to be investigated.

Serbian

(1) Molim te, uradi išta od toga!
    please do.imp i-word-anything of that
    Please do any of those things!
because of the logical equivalence between $\forall \neg$ and $\neg \exists$. The generalization thus stays intact — BPIs cannot be used with an existential force in the scope of clausemate sentential negation. According to the hypothesis we are proposing here, this is so because n-words can be used exclusively with an existential force in the scope of clausemate negation, while BPIs can outscope it.

3 NPI-PPI distribution

Let us now see how with the similar reasoning we could explain the distribution of PPIs that have an NPI counterpart. For the simplicity of exposition, we will discuss English data only. Namely, PPIs cannot be interpreted in the direct scope of antimorphic operators — the sentence (15) only has the meaning in which something scopes above negation:

$$\exists \gg \neg, *\neg \gg \exists$$

(15) John didn’t do something.

To get a reading in which an existential indefinite scopes below negation, one would have to use an NPI anything.

$$\neg \gg \exists$$

(16) John didn’t do anything.

If this pattern was really due to the possibility of something to scope above negation, we would expect anything to be the kind of NPIs that is not able to outscope negation with an existential force. Let us embed the negative sentence with anything in the scope of negation into another DE environment, to show that this is indeed the case.

$$if \gg \neg \gg \exists, *if \gg \exists \gg \neg$$

(17) If John doesn’t do anything, he will fail the exam.

The reading ‘If there is something John doesn’t do, he will fail the exam’ is reported not to be available for (17). Anything is thus not able to move above negation — its scope options with respect to negation are more similar to those of n-words than to those of BPIs. As English PPIs are able to move above negation, and English NPIs are not, in line with the minimizing scope ambiguity hypothesis one is forced to choose an NPI in the scope of an antimorphic operator: this would derive the PPI behaviour of something, someone etc.

Let us now discuss the environments where both NPIs and PPIs can be used with the same meaning. These are:

– the scope of non-clausemate negation ($\neg \gg \exists$)

(18) It’s not true that Peter saw something/anything.

– rescuing contexts ($\neg \gg \neg \gg \exists$)

(19) It’s not true that Peter didn’t see something/anything.

– antecedents of conditionals ($if \gg \exists$)
If you see something/anything, let me know.

- nuclear scope of DE and Strawson-DE operators like at most, only NP, few...
  (few/at most... ⋈ ≃ ∃)

(21) a. Few people saw something/anything.
    b. Only John saw something/anything.

- restrictor of the universal quantifier

(22) Everyone who read something/anything will pass.

- questions (Q ⋈ ∃)

(23) Have you seen someone/anyone?

We will try to argue that the fact that the complementary distribution of NPIs and PPIs breaks down in these environments is compatible with the minimizing scope ambiguity hypothesis.

3.1 Superordinate negation, questions, antecedents of conditionals
and restrictors of universal quantifiers

Let us first discuss superordinate negation, questions, antecedents of conditionals and restrictors of universal quantifiers. Namely, these environments seem to be islands for quantifier movement — quantifiers cannot scope out of antecedents of conditionals, questions, restrictors of universal quantifiers and tensed CPs. To see this, let us examine the possible interpretations of (24) and (25).

The sentence in (24) cannot be interpreted as ‘for every book, if you read it, you will pass the exam’ (i.e. ‘if you read any book, you will pass’), which would be the reading obtained if the quantifier every could scope out of the antecedent of conditionals. The same is true for questions — the question in (25) cannot be understood as ‘there are more than three books for which I am asking you if you have read them’, which would be the meaning obtained if the quantifier more than three could scope above the question operator.

\[
\text{if } ⋈ \text{ every, } ^* \text{every } ⋈ \text{ if}
\]

(24) If you read every book, you will pass the exam.

\[
\text{Q } ⋈ \text{ more than three, } ^* \text{more than three } ⋈ \text{ Q}
\]

(25) Have you read more than three books?

If a quantifier cannot scope out of the antecedent of conditionals, questions, restrictors of universal quantifiers and tensed CPs, it cannot create scope ambiguities with respect to the conditional operator, superordinate negation, universal quantifier or the question operator. PPIs are quantificational elements, so these movement locality restrictions apply to them as well. As they cannot

7 Similar examples can be constructed for the superordinate negation and the restrictor of universal quantifiers.
move out of the antecedent of conditionals, questions, restrictors of universal quantifiers and above the superordinate negation, the fact that PPIs and NPIs are interchangeable there is compatible with the minimizing scope ambiguity hypothesis.

It should be noted, however, that certain quantifiers take exceptionally wide scope out of these environments through a mechanism different from movement — the choice functions [8]. The minimizing scope ambiguity hypothesis thus has to be restricted to the scope ambiguities that result from the quantifier movement only.8

3.2 Rescuing

Let us now discuss rescuing — when negation is embedded inside certain environments, PPIs can be interpreted in its scope (PPIs are said to be rescued in those contexts). Rescuing is sometimes possible even in the scope of the preposition without. Let us see a (non-exhaustive) list of these contexts:

\[ \exists \gg \neg, \neg \gg \exists \]

(26) a. If John doesn’t understand something, he will fail the exam.
   b. Few people don’t understand something.
   c. Only John doesn’t understand something.
   d. It’s not true that John doesn’t understand something.
   e. I am surprised that John doesn’t understand something.
   f. I never go to the library without some chocolate.

In these contexts the PPI *something* can be interpreted either above or below negation, and as we have seen, its complementary distribution with NPIs breaks down in those contexts, which seems to be at odds with the minimizing scope ambiguity hypothesis. However, these contexts cease to be problematic for the hypothesis defended here once we accept a particular approach to rescuing phenomenon — the light negation hypothesis, initially proposed in [5]. Namely, according to [5], rescuing involves another negation, which is homophonous with the regular negation. When the PPI is interpreted in the scope of negation in (26), the subordinate negation is this special negation; when the PPI outscopes it, the subordinate negation is a regular negation. As empirical support for this claim, it is pointed out in [5] that negation with a rescued PPI in its scope obligatorily scopes over the subject PPI9, as in (27) — this special rescuing negation thus takes a wider scope than the regular negation.

\[ \neg \gg \neg \gg \exists \gg \neg \gg \neg \gg \exists \gg \neg \gg \neg \gg already \]

(27) You can’t convince me that somebody hasn’t already finished the exam.

8 Modified numerals like *more than three* have been argued not to be able to take exceptional wide scope through choice functions [8], which is why numerous examples I am presenting contain them.

9 But see [4] for a different account for these data.
This proposal has been revived in [9], which is about the rescuing data from German: in rescuing contexts in German negation seems to occupy a syntactically higher position than usual — this is why they call it ‘light negation’.

I will follow this line of reasoning and accept that indeed in rescuing environments negation is at a syntactically higher position than usual. An additional support for this is that the rescuing negation seems to take scope over the subject position even when the subject is a regular indefinite, and not a PPI, as can be seen from (28).\(^\text{10}\)

(28) It’s not true that more than three people haven’t already read Harry Potter.

This sentence contains the rescued PPI already. If more than three people scopes above the subordinate negation, the meaning should be that there are three or fewer people who haven’t already read Harry Potter, while if the subordinate negation scopes above the subject position, the meaning is that more than three people have already read Harry Potter. This sentence is reported as not having the first reading — it seems that the rescuing negation obligatorily scopes above the subject position unless the subject is a DE or a Strawson-DE operator, as in (26b) and (26c).\(^\text{11}\)

Note the interesting consequence of the impossibility of more than three people to scope above the light negation: it seems that quantificational expressions in the scope of light negation cannot outscope it through movement. If quantificational expressions cannot outscope the light negation through movement, that would be the reason why according to the minimizing scope ambiguity hypothesis PPIs would not be ruled out from the scope of negation in rescuing contexts.

One of the arguments against the light negation has been the fact that rescuing sometimes seem to be possible in the scope of the preposition without. However, it seems that when this happens, scope properties of the preposition without with respect to the quantificational elements in its scope are very similar to that of light negation: nothing can outscope it through movement. To see this, compare (29) and (30).

(29) I can’t finish my project without someone’s help three days per week.

(30) I can’t finish my project without anyone’s help three days per week.

Regarding the quantifiers that are involved, consider the following meanings:

1. I can’t finish my project if it’s not the case that there is someone who is helping me three days per week — there is someone who needs to help me three days per week. \((\neg \gg \text{without} \gg \exists \gg \text{three days per week})\)

\(^\text{10}\) Of course, if subject indefinites turn out to be PPIs as proposed in [10], this piece of data makes the same point as the example (27).

\(^\text{11}\) Admittedly, these judgements are difficult, as they contain multiple quantificational expressions, so they deserve to be reverified.
2. I can’t finish my project if it’s not the case that for three days a week there
is at least someone who helps me. \((\neg \exists \text{without} \text{three days per week})\)

3. I can’t finish my project if three days a week I have no helper — I need to
have at least someone helping me at least 4 days a week. \((\neg \exists \text{without} \text{three days a week})\)

The sentence with the rescued quantifier (29) was judged to have the meanings
1 and 2. The meaning 3 seems not to be there for (29). On the other hand, the
meaning 3 was the clearest one available for (30). If this is indeed the case, rescu-
ning contexts cease to be problematic for the hypothesis defended here, as it seems
that quantificational elements from the scope of negation and the preposition
without cannot outscope them in those contexts.\(^{12,13}\)

### 3.3 Nuclear scope of DE and Strawson-DE operators such as few

An environment for which it is harder to explain why the complementary distri-
bution of NPIs and PPIs breaks down is the scope of DE and Strawson-DE
operators, as in (21). For the time being I have only a speculation about these
environments. All these environments can have a positive inference that could
disrupt the DE-ness of the environment. When this positive inference is forced,
only the PPI is licensed, as can be seen from the examples below.

\[(31)\] a. At least 3 and at most 5 people did something/*anything.
   b. Few people, but at least 3, did something/*anything.

As the positive inference disrupts the DE-ness of the environment, it could
be that whenever we have a some-item in these contexts, we have factored in
this positive inference. So the interchangeability of some/any in those contexts
could be due to the optionality of the positive inference — when the positive
inference is not there, any is used; and presumably, when the positive inference
is there, some can be used simply because any cannot be used anymore.

However, the examples in (31) only show that when we have a positive in-
fERENCE any cannot be used — what needs to be shown is that the implication
holds in the other direction as well, namely that whenever we have a PPI, we
have an obligatory positive inference. I am not sure how to test for the presence
of this inference yet, so I leave this question open for the time being.

\(^{12}\) There is an alternative way to deal with rescuing. At least some NPIs cannot be
licensed in the scope of light negation — in [5] it is shown that yet cannot. I am
not sure which NPIs can and which cannot be licensed in the scope of light negation —
if it turns out that any-NPIs cannot, some-PPIs could be licensed in the scope
of light negation simply because any-NPIs cannot be licensed there, so minimizing
scope ambiguity hypothesis cannot prefer any-NPIs over some-PPIs.

\(^{13}\) BPIs in the three languages seem to be the NPIs that cannot be licensed by the light
negation (they cannot be rescued) — see also the footnote 12.
4 Technical implementation

Before concluding, I would like to discuss the technical implementation of the explored hypothesis. It could come in two versions: I will refer to them as item-based and environment-based version of the minimizing scope ambiguity hypothesis. The item-based version would have the competing items organized in the mental dictionary according to their scope liberties: whenever we could, we would use the least ambiguous item in terms of scope liberties, regardless of the syntactic environment in which these items would appear. In other words, an n-word would always block the usage of a BPI, and an NPI would always block the usage of a PPI. This would ultimately be a scope ambiguity version of the blocking account as proposed in [6].

The environment-based version would have the attempt of minimizing scope ambiguities be sensitive to the context in which the PSI appears. This version would thus state that, syntactic structure and logical form (LF) being equal, whenever there is a lexical choice between an item that can create scope ambiguities and an item that cannot create scope ambiguities in that particular syntactic environment, we are forced to opt for the lexical item that cannot create scope ambiguities. If no ambiguities can be created, we are allowed to use either item.

The difference is thus mainly whether the minimizing scope ambiguity competition occurs at the level of the lexical insertion, or whether for the same LF we consider in parallel two alternative derivations which differ only with respect to the selected PSI (e.g. one with a PPI and the other with an NPI in the scope of the clausemate sentential negation), and then reject the one which is potentially scopally ambiguous.

While BPIs and n-words seem to be in the complementary distribution for the same LF, we have seen that NPIs and PPIs can be used interchangeably at a number of positions. Because of this, at least for the NPI-PPI distribution the item-based version of the hypothesis seems not to be right. If the attempt to provide a unified explanation for the BPIs and for the NPI-PPI distribution in terms of not creating scope ambiguities is on the right track, we have to opt for the environment-based version of the minimizing scope ambiguity hypothesis.

However, the environment-based version makes one wrong prediction about Polish BPIs. We have seen that Polish *kolwiek*-pronouns are not acceptable in the scope of the preposition *without*. The prediction of the current account is that *kolwiek*-pronouns are able to outscope this preposition. To check this, we have to embed the preposition *without* into another DE environment to ensure the semantic well-formedness of the *kolwiek*-pronoun when it outscopes the preposition *without* — the Polish equivalent of (32) is predicted to be able to mean ‘I regret that there is someone who she came without’. However, my Polish informants find the sentence (32) simply bad. For the time being, I will leave this problem open.

(32) I regret that she came without kolwiek-anyone.
5 Conclusion

We have explored a hypothesis according to which the impossibility of BPIs and PPIs to be interpreted in the scope of clausemate negation is due to their ability to outscope it, while there is alternative item with the same meaning that cannot do so. While this proposal captures most of the data, I have nevertheless left some empirical questions open — the scope of DE and Strawson-DE quantifiers for PPIs and the scope of the preposition *without* for Polish BPIs.

There are also a few conceptual issues that need to be clarified. The first is the question of light negation: if negation really differs somehow in rescuing environments, we still need to explain its semantics (what is it negating?), and its syntactic position. Secondly, it would be interesting to understand what allows certain NPIs to scope over negation (the Bagel problem items), while others cannot (English *any*, Italian *alcun* and French *-que ce soit* items). Thirdly, it’s still open how this hypothesis could be extended to elements other than these items. These issues are left for future work.

References

10. Spector, B.: Some indefinites are positive polarity items in subject position only. Snippets 9 (2004)

14 In my MA thesis I try to extend it to weak scalar particles based on the generalizations presented in [3].
Negative Coordination in French⋆

Aurore González
Harvard University

Abstract. This paper investigates the distribution and interpretation of the coordinating particle *ni* (‘nor’) in French. As claimed in the literature, *ni* behaves like an NPI coordination. Since on the most commonly held view of NPIs, NPIs are existentials, the polarity sensitivity of *ni* suggests that it is interpreted as a disjunction. However, following [7], we provide in this paper several data challenging both the hypothesis that *ni* is an NPI coordination and the hypothesis that *ni* is interpreted as a disjunction where it is generated. We then suggest an analysis of *ni* that can account for these challenges while maintaining the hypothesis that *ni* is a strong NPI interpreted as a disjunction: namely, the coordinating particle *ni* needs to move at LF to be interpreted in the direct scope of its licensor.

Negative Sensitive Items (henceforth NSIs) raise two core issues: (i) Are NSIs inherently negative, not inherently negative or ambiguous?, (ii) Are NSIs universals outscoping negation (i.e. $\forall > \neg$) or existentials interpreted in the scope of negation (i.e. $\neg > \exists$)?

The goal of this paper is to contribute to this debate by exploring negative coordination in French. When *ni* occurs once introducing the second conjunct of a coordinated structure (so-called single *ni* (‘nor’)), it behaves like an NPI coordination ([4], [11], [13]): its distribution is restricted to the scope of negation ((1a)-(1b)), and it cannot serve as an answer on its own in elliptical contexts ((1c)).

(1) a. Jean n’aime *(pas) le thé ni le café.
Jean NE.likes not the tea nor the coffee
‘Jean doesn’t like tea nor does he like coffee.’

b. *Léo ni Jean n’aime (pas) le café.
Léo nor Jean NE.likes not the coffee

⋆ Many thanks to Gennaro Chierchia, Anamaria Fălăuş and Hamida Demirdache for discussion and insightful comments.

1 We use NSI here as a cover word for *n*-words, Negative Polarity Items (henceforth NPIs) and Negative Concord Items (henceforth NCIs).

2 For previous work on negative coordination, we refer the reader to [14] for discussion of English *nor;* [4], [5], [7], [11], [13] for French *ni* and [1], [9] for Spanish *ni*.

3 Sentential negation in French is expressed by the negative adverb *pas,* optionally co-occurring with the preverbal expletive clitic *ne.*
In contrast, when *ni* is reiterated, introducing both conjuncts of a coordinated structure (so-called double *ni* (‘neither...nor’)), it appears to be inherently negative ([11], [13]): it can introduce negation on its own ((2a)–(2b)) and can serve as an answer in elliptical contexts ((2c)).

(2) a. Jean n’aime **ni** le thé **ni** le café.
   ‘Jean likes neither tea nor coffee.’

b. **Ni** Léo **ni** Jean n’aime le café.
   ‘Neither Léo nor Jean likes coffee.’

c. Que bois-tu? – **Ni** café **ni** thé.
   ‘What do you drink?’ – ‘Neither tea nor coffee.’

The two issues extensively explored for NSIs can thus be reformulated for negative coordination as follows: (i) Is the coordinating particle *ni* inherently negative, not inherently negative, or ambiguous?, (ii) Given that \( \neg (p \lor q) \) and \( \neg p \land \neg q \) are equivalent (De Morgan’s law), the interpretation of a sentence containing the coordinating particle *ni* could in principle arise from two different structures: a disjunction with narrow scope relative to negation, or a conjunction with wide scope relative to negation. Is the coordinating particle *ni* interpreted as a conjunction or as a disjunction?

In this paper, we will focus on single *ni*. It has been shown in the literature that single *ni* is a strong NPI only licensed in the scope of Anti-Additive (henceforth AA) items\(^4\) ([11], [13]). On the most commonly held view of NPIs, NPIs are existentials (e.g. [10]). The polarity sensitivity of *ni* thus suggests that it is interpreted as a disjunction. However, based on [7], we will provide in Sects. 1 and 2 several data that challenge both the hypothesis that single *ni* is an NPI coordination and the hypothesis that single *ni* is interpreted as a disjunction where it is generated. In Sect. 3, we will then suggest an analysis of *ni* that can account for these challenges while maintaining the hypothesis that single *ni* is a strong NPI interpreted as a disjunction: namely, single *ni* does not always have surface scope; at LF, it moves via covert movement to be interpreted in the direct scope of its licensor.

\(^4\) A function *f* is anti-additive iff for any *A* and any *B*, \( f(A \lor B) \leftrightarrow f(A) \land f(B) \).
1 Challenges for the Hypothesis that Single ni is an NPI
Coordination ([7])

As already introduced, if single ni is a strong NPI, we would expect it to only be licensed in the scope of AA items. However, the following example shows that this generalization does not hold.

(3) a. Not every student smokes and not every student drinks. \(\rightarrow\) Not every student smokes or drinks.
   b. Pas tous les étudiants ne fument ni ne boivent.
      not.every the students NE smoke nor NE drink
      ‘Not every students smoke and not every students drink.’

(3a) shows that not every doesn’t denote an AA function. Now, (3b) shows that single ni is licensed in the scope of pas tous (‘not every’), challenging thus the hypothesis that single ni is a strong NPI only licensed in the scope of AA items. Furthermore, in (3b), the coordinating particle ni takes wide scope with respect to the universal quantifier tous (‘tous’). That is, the sentence is interpreted as \(\wedge > \neg > \forall\) \((\leftrightarrow \neg > \lor > \forall)\). Crucially, this reading is not expected under the hypothesis that single ni is interpreted as a disjunction where it is generated. Under the latter hypothesis, we would expect (3b) to be interpreted as \(\neg > \forall > \lor\); this reading entails that some students neither smoke nor drink. However, (3b) only entails that some student don’t smoke and some student don’t drink. Therefore, (3b) challenges both the hypothesis that single ni is a strong NPI and the hypothesis that single ni is interpreted as a disjunction where it is generated.

Secondly, according to [6], non Neg-Raising (henceforth NR) predicates (e.g. to claim in (4a)), unlike NR ones (e.g. to believe in (4b)), cannot license strict NPIs such as in years.

(4) a. *Mary didn’t claim that Bill had left the country in years.
   b. Mary doesn’t believe Bill has left the country in years.

However, (5) shows that the coordinating particle ni is licensed in the scope of non-NR predicates such as affirmer (‘to claim’).

(5) Elle n’affirme pas que Jean ira à la piscine ni au cinéma.
    she NE.claims not that Jean will.go to the pool nor to the cinema
    ‘She doesn’t claim that Jean will go to the pool and she doesn’t claim that Jean will go to the cinema.’

Once again, the hypothesis that single ni is a strong NPI is challenged on the basis of this argument: if single ni were a strong NPI, then we would expect it not to be licensed in the scope of non-NR predicates.

\(\text{Strict NPIs correspond to [15]’s strong NPIs excluding minimizers.}\)
Finally, according to [3] among others, the intervention of a certain class of items between an NPI and its licensor (e.g. quantificational DPs ((6a)), numerals other than one ((6b)), the conjunction and ((6c))) trigger intervention effects thus causing ungrammaticality.

(6)  
   a. ??I doubt that every student will ever have any problems. 
   b. ??I never met twenty students who had read any of my papers. 
   c. ??I doubt that Theo drank the left over wine and any coffee. 

Now, as illustrated below, when the universal quantifier tous (‘every’) ((7a)), the numeral vingt (‘twenty’) ((7b)) and the conjunction et (‘and’) ((7c)) intervene between negation and single ni, the sentences are perfectly well-formed.

(7) a. Jean ne veut pas que tous les étudiants aillent à la piscine. 
   Jean NE wants not that all the students go to the pool 
   ni au cinéma. 
   nor.to.the cinema 
   ‘Jean doesn’t want that every students go to the pool and Jean 
   doesn’t want that every students go to the cinema.’ 
   b. Jean n’a jamais rencontré vingt étudiants à la piscine ni 
   Jean NE has never meet twenty students to the pool nor 
   au cinéma. 
   to.the cinema 
   ‘Jean never met twenty students to the pool and Jean never met 
   twenty students to the cinema.’ 
   c. Jean ne veut pas que Paul et Marie aillent à la piscine ni 
   Jean NE want not that Paul and Marie go to the pool nor 
   au cinéma. 
   to.the cinema 
   ‘Jean doesn’t want that Paul and Marie go to the pool and Jean 
   doesn’t want that Paul and Marie go to the cinema.’ 

(7) shows that single ni, unlike other NPIs, doesn’t show intervention effects. Once again, this generalization is surprising given the claim that ni is a strong NPI.

2 Challenges for the Hypothesis that Single ni Is Interpreted as a Disjunction ([7])

Recall that given that ¬[p∨q] and [¬p∧ ¬q] are equivalent (De Morgan’s law), the interpretation of a sentence containing the coordinating particle ni could in principle arise from two different structures: a disjunction with narrow scope relative to negation, or a conjunction with wide scope relative to negation. We now provide several data challenging the hypothesis that single ni is interpreted as a disjunction where it is generated.
To tease apart the predictions made by the existential vs. universal analysis of wh-NPIs in Japanese, [12] constructs examples where a quantificational adverb (henceforth Q-adverb) together with negation creates a non-AA context.

(8)  a. $Q > \neg > \exists$
    b. $Q > \forall > \neg$
    c. $\forall > Q > \neg$

While the existence of the reading in (8a) is not informative (because it is equivalent to the reading in (8b)), the availability of the reading in (8c) provides an argument in favor of the hypothesis that the NPI is interpreted as a universal (since (8c) has no equivalent interpretation with an existential).

We now adapt [12]'s test to tease apart $\land > \neg$ from $\neg > \lor$, by constructing examples where the Q-adverb *deux fois par semaine* (‘twice per week’) together with negation creates a non-AA context $Q \neg$.

(9)  a. **Deux fois par semaine**, Jean ne joue pas de violon ni deux times per week Jean NE plays not PART. violin nor Marie de piano.
    Marie PART. piano
    ‘Twice per week, Jean doesn’t play violin nor does Marie play piano.’
    b. $\land > Q > \neg$: **Twice per week**, Jean doesn’t play violin and twice per week, Marie doesn’t play piano.

If (9a) allows the scope splitting construal $\land > Q > \neg$ in (9b), it should be judged true in the context in Table 1, where Jean doesn’t play violin on Monday and Tuesday, and Marie doesn’t play piano on Wednesday and Thursday.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>T</th>
<th>W</th>
<th>T</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jean</td>
<td>no</td>
<td>no</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marie</td>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

That (9a) is judged true in the above table shows that this sentence is interpreted as $\land > Q > \neg$. An analysis of single *ni* as a disjunction is here suprising: if *ni* would be interpreted as a disjunction, we would expect (9a) to mean $Q > \neg > \lor$, and therefore, we would expect the sentence to be judged false in the above context.

Now, recall that in Sect. 1, we already provided an example challenging the hypothesis that single *ni* is interpreted as a disjunction where it is generated. A similar data also challenging the latter hypothesis comes from the following example.
Tous les animaux ne sont pas approchables ni apprivoisables.

all the animals are not approachable nor tamable

a. ‘No animals are approachable and no animals are tamable.’ ∀ > ¬

b. ‘It is not the case that all animals are approachable and it is not the case that all animals are tamable.’ ¬ > ∀

The sentence in (10) is ambiguous: it can either receive a reading where the universal quantifier *tous* takes wide scope with respect to negation ((10a)) or a reading where the universal quantifier scopes under negation ((10b)). Now, on the second reading, single *ni* takes wide scope with respect to the quantifier (∧ > ¬ > ∀ ↔ ¬ > ∨ > ∀). Crucially, this reading is not expected under the hypothesis that single *ni* is interpreted as a disjunction where it is generated. Under the latter hypothesis, we would expect (10) to have as a second reading ¬ > ∀ > ∨. That (10) doesn’t get this reading challenges thus the hypothesis that single *ni* is interpreted as a disjunction where it is generated.

3 The Proposal

To begin with, recall the issue. Given that ¬[p ∨ q] and [¬p ∧ ¬q] are equivalent (De Morgan’s law), the interpretation of a sentence containing the coordinating particle *ni* could in principle arise from two different structures: a disjunction with narrow scope relative to negation, or a conjunction with wide scope relative to negation. As already introduced, it has been shown in the literature that single *ni* is a strong NPI, suggesting that it is interpreted as a narrow scope disjunction. However, in Sects. 1 and 2, we provided several data that challenge the latter hypothesis. How do we formally capture these facts?

Two hypotheses are conceivable. According to the first one, the challenging data discussed in Sects. 1 and 2 show that single *ni* is interpreted as a conjunction outscoping negation (as argued by [7]). According to the second one, although single *ni* is interpreted as a disjunction under the scope of negation, the challenging data show that single *ni* is not interpreted where it is generated.

We explore in this paper the second hypothesis. More precisely, we suggest that single *ni* is a strong NPI interpreted as a disjunction, that does not always have surface scope. Namely, the coordinating particle *ni* moves at LF to be interpreted in the direct scope of its licensor. To this effect, we suggest the following denotation for the coordinating particle *ni*.

\[
\text{[ni]} = \lambda R_a. \lambda S_a. \lambda P_{<a,t>} \cdot \exists Q \in \{R, S\}: P(Q)
\]

To illustrate the proposal, consider the following sentence.

(12) a. Jean n’est pas malin ni courageux.

Jean NE-is not smart nor brave

‘Jean isn’t smart nor is he brave.’

b. \[\text{[ni]} = \lambda R_{<e,t>} . \lambda S_{<e,t>} . \lambda P_{<e,t>} . \exists Q \in \{R, S\}: P(Q)\]
We first assume that since in (12a), the coordinating particle *ni* takes two properties as arguments, it has the denotation in (12b). Now, for a reason of type mismatch, the coordination phrase (i.e. *malin ni courageux*) cannot be interpreted where it is base-generated. It has to move via covert movement outside TP to be interpreted, as illustrated in (12c). The sentence in (12a) then receives the interpretation in (12d).

We now show that our hypothesis that single *ni* is a strong NPI interpreted as a disjunction that does not always have surface scope can account for both the surprising distribution of single *ni* and the surprising interpretations of sentences containing it.

To begin with, recall that under the hypothesis that single *ni* is a strong NPI, we expect the latter to only be licensed in the scope of AA items. However, as illustrated again in (13a), although *pas tous* (‘not every’) doesn’t denote an AA function, single *ni* can occur in its scope.

(13) a. **Pas tous** les étudiants ne fument **ni** ne boivent.
   not.every the students NE smoke nor NE drink
   ‘Not every students smoke and not every students drink.’

b. [pas [[fument **ni** boivent] 1 [tous les étudiants t1]]]

c. \( \neg \exists Q \in \{\lambda x. \text{smart}(x), \lambda y. \text{brave}(y)\}: Q(\text{Jean}) \)

We assume here that in (13a), the coordination phrase (namely, *fument ni boivent*) moves under the direct scope of negation to be interpreted, as illustrated in (13b)\(^6\). Since at LF, single *ni* has wide scope with respect to the universal

\(^6\) Notice that in (13b), we assume that French *pas tous* isn’t a complex determiner; that is, negation *pas* has scope over the whole proposition. This assumption is supported by the following data in Québécois French (from [2]).

(1) a. *Personne* est venu **pantoute**.
   n-one is come at.all
   ‘No one came at all.’

b. **Pas** personne est venu **pantoute**.
   not n-one is come at.all
   ‘No one came at all.’

In Québécois French, *pantoute* is a strong NPI with a particular distribution: it can only occur in the scope of sentential negation *pas,* *plus* (‘no more’), *rien* (‘nothing’) and *sans* (‘without’). As illustrated in (1a), *pantoute* cannot occur in the scope of the NCI *personne* (‘no one’). However, when negation *pas* precedes *personne,* *pantoute* is now licensed. This suggests that *pas personne* doesn’t form a complex determiner; rather, negation *pas* has scope over the whole proposition, and can thus license the NPI *pantoute*.

72
quantifier\(^7\), but narrow scope with respect to negation, it now occurs in an AA context, and consequently, we expect it to be licensed.

Furthermore, as already discussed, that the sentence in (13a) is interpreted as \( \land > \neg > \forall \) (\( \leftrightarrow \neg > \lor > \forall \)) was not expected under the hypothesis that single ni is interpreted as a disjunction where it is generated. Under our proposal, given the LF in (13b), we derive for the sentence in (13a) the interpretation in (13c) (i.e. \( \neg > \lor > \forall \)). Therefore, our hypothesis that single ni is a strong NPI interpreted as a disjunction that moves at LF to be interpreted in the direct scope of its licensor, can account for both the surprising reading of (13a) and the surprising distribution of ni.

Secondly, recall that single ni, unlike other strong NPIs, is licensed in the scope of non-NR predicates such as affirmer (‘to claim’).

\begin{enumerate}
\item a. Marie n’affirme pas que Jean ira à la piscine ni au cinéma.\\
\textit{‘Marie doesn’t claim that Jean will go to the pool nor to the cinema.’}\\
\item b. [pas [à la piscine ni au cinéma] 1 [elle affirme que [Jean ira t]]]]
\end{enumerate}

As illustrated in (14b), under our proposal, single ni moves via covert movement out of the scope of the predicate affirmer, to be interpreted in the direct scope of negation. After this movement, single ni does no longer occur under the scope of the non-NR predicate ((14b)). Our proposal thus predicts single ni to be licensed in (14a).

This covert movement of the coordination phrase also accounts for the fact that single ni doesn’t show intervention effects. As repeated below, when a universal quantifier intervenes between single ni and sentential negation pas, its licensor, the sentence is perfectly well-formed.

\begin{enumerate}
\item a. Jean ne veut pas que tous les étudiants aillent à la piscine ni au cinéma.\\
\textit{‘Jean doesn’t want that every students go to the pool nor to the cinema.’}\\
\item b. [pas [à la piscine ni au cinéma] 1 [Jean veut que [tous les étudiants aillent t]]]]
\end{enumerate}

\(^7\) To ensure that single ni moves under the direct scope of negation, and consequently outscopes the universal quantifier, we assume, following [8] that NPI ni carries an uninterpretable [+neg] feature that needs to be checked at LF against a negative head.
As illustrated in (15b), since the coordination phrase `à la piscine ni au cinéma' moves via covert movement under the direct scope of negation, at LF, the universal quantifier *tous* does no longer intervene between NPI *ni* and its licensor. That single *ni* doesn’t show intervention effect is thus expected under our proposal.

Notice that this proposal that single *ni* moves via covert movement to be interpreted in the direct scope of its licensor makes a straightforward prediction: long-distance licensing of single *ni* should show island sensitivity. The following example shows that this prediction is indeed borne out.

(16) a. *Marie ne se demande pas pourquoi Jean regardera
Marie NE RECP wonder not why Jean will.watch
Blade Runner *ni* Summertime.
Blade Runner nor Summertime

b. *[pas [[Blade Runner *ni* Summertime] 1 [Marie se demande pourquoi
[Jean regardera t1]]]]

(16a) shows that single *ni* cannot be licensed across an island. Under our proposal, these island effects are expected. Since in (16a), single *ni* is embedded within an island, the movement of the coordination phrase (namely, *Blade Runner ni Summertime*) from the embedded clause (where it is base-generated) to the matrix clause (where it is interpreted) is not allowed: the derivation crashes.

Finally, consider again the following sentence containing both single *ni* and the quantifier *tous* (*‘every’*) in subject position. Recall that this sentence is ambiguous: it can either receive a reading where the universal quantifier *tous* takes wide scope with respect to negation ((17a)) or a reading where the universal quantifier scopes under negation ((17b)).

(17) *Tous les animaux ne sont pas approchables ni apprivoisables.*

a. ‘No animals are approachable and no animals are tamable.’ ∀ > ¬
b. ‘It is not the case that all animals are approachable and it is not the case that all animals are tamable.’ ¬ > ∀

(18) ∀: animal(x) → ¬ ∃Q ∈ {λx. approachable(x), λy. tamable(y)}: Q(x)
(19) ¬ ∃Q ∈ {λx. approachable(x), λy. tamable(y)}: ∀: animal(x) → Q(x)

Our proposal that single *ni* is a strong NPI interpreted as a disjunction that moves at LF to be interpreted in the direct scope of negation, nicely derives the two readings of the sentence in (17). To begin with, when the universal takes wide scope with respect to negation ((17a)), it follows that single *ni* is interpreted in the direct scope of negation. As illustrated in (18), the reading ∀ > ¬ > ∀ is derived. Now, when negation takes wide scope with respect to the universal ((17b)), the coordination phrase (namely, *approchables ni apprivoisables*) moves via covert movement out of the scope of the universal to be interpreted in the direct scope of negation. In that case, as illustrated in (19), the reading ¬ > ∀ > ∀ is derived.
4 Conclusion and Outstanding Issues

To conclude, in this paper, we investigated negative coordination in French, and in particular the distribution and interpretation of single ni (‘nor’). Following [4], [11] and [13], we first assumed that single ni is a strong NPI coordination. Under this hypothesis, we would expect it to be interpreted as a disjunction having narrow scope relative to negation. However, we discussed several data that challenge both the hypothesis that ni is an NPI coordination and the hypothesis that ni is interpreted as a disjunction where it is generated. We then suggested an analysis of single ni that can account for these challenges while maintaining the hypothesis that single ni is a strong NPI interpreted as a disjunction: namely, single ni does not always have surface scope; it moves via covert movement to be interpreted in the direct scope of its licensor.

Notice that one issue still remains: why is the sentence in (9a) (in which the quantifier deux fois par semaine (‘twice per week’) co-occurs with single ni and sentential negation pas) interpreted as $\land > Q > \neg > \lor$? Under our analysis, we would expect this sentence to be interpreted as $Q > \neg > \lor$, contrary to facts. This issue needs to be investigated in further research.

References


Experimental investigations
of probability expressions:
A first step in the (probably) right direction

Michele Herbstritt

University of Tübingen

Abstract. This paper is concerned with the semantics of probability expressions such as probably and likely. According to a recent theory proposed by [6], the meaning of a sentence such as probably \( \varphi \) is sensitive to its context of utterance, and in particular to the set of \( \varphi \)'s contextually salient alternative outcomes. We report the results of three experiments specifically designed to investigate this context-sensitivity. On the basis of our results, we discuss possible directions for both experimental and theoretical future work.

1 Introduction

The vast majority of real world reasoning and decision-making tasks takes place under uncertainty. It is not surprising, then, that linguistic expressions communicating uncertainty (and probability expressions such as probably and likely in particular) are ubiquitous in communication, from everyday talk about the weather to more serious situations of risk communication and management.

From a linguistic point of view we are interested in the meaning and use of probability expressions. In this respect, we can distinguish two classic approaches to the semantics of probably and likely. In formal semantics, the common view dates back to [5] and her qualitative treatment of probability expressions as modals. In psychology, the tendency has been to implicitly assume that a sentence such as probably \( \varphi \) simply means that the probability of \( \varphi \) is bigger than 0.5. In contrast with these two trends, in recent years a number of authors have argued in favor of the adoption of a probabilistic semantics for probability expressions (see [13], [17], [18], [6] and [9], among others). We will come back to these three trends in some more detail in the following section.

This paper does not advocate one particular position on the semantics of probably and likely, as it is intended to set the stage for a thorough empirical investigation of these expressions. However, we will focus on the theory proposed

* The work reported in this paper has greatly benefited from discussions with Michael Franke, Anthea Schöller, Margherita Isella and Judith Degen. Many thanks go to them, and to Fabian Dablander for his substantial help with the experiments. Financial support by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63) is gratefully acknowledged.
by [6]. As we will see, his approach can fit within the probabilistic semantics trend, and it posits a particular kind of context sensitivity of the meaning of *probably* and *likely* which yields clear predictions, well suited for experimental testing.

The paper is structured as follows. In Section 2 we lay out the theoretical background of our work, summarizing Kratzer’s approach to probability expressions and a basic implementation of the probabilistic semantics mentioned above. Moreover, we briefly summarize a crucial empirical finding about how people usually reason when probability expressions are in the picture. This empirical fact motivates Lassiter’s proposal, which is sketched in the last part of Section 2. Section 3 reports our empirical investigations in detail. Section 4 concludes and sets the stage for future work.

## 2 Background

According to a classic approach due to [5], probability expressions such as *probably* and *likely* are to be semantically analyzed as modals, i.e. as quantifiers over possible worlds relative to two contextually determined parameters: the modal base, which fixes the domain of quantification (a set of epistemically accessible worlds) and the ordering source, a set $O$ of propositions used to determine a preorder $\geq_O$ over the worlds in the domain.\(^1\) The preorder $\geq_O$ is lifted to propositions, determining a preorder $\succsim_O$ which is taken to formalize the semantic contribution of the comparative expression *is at least as probable/likely as*: $$\varphi \succsim_O \psi \text{ iff } \forall w \in \psi \exists v \in \varphi \text{ s.t. } w \geq_O v.$$\(^2\)

Based on this definition, the semantics of the strict comparative *is more probable/likely than* is defined as $\varphi \succ_O \psi$ iff $(\varphi \succsim_O \psi) \land \neg (\psi \succsim_O \varphi)$ and finally the meaning of the bare proposition embedding particle *probable/likely* is equated with *is more probable/likely than not*.

The main advantage of Kratzer’s approach is its conservativity. By relying exclusively on a qualitative notion of likelihood, it allows us to stick to the traditional possible worlds framework. Moreover, as shown in details by [18], it validates a number of intuitively valid inference patterns involving probability expressions and their interactions with connectives and other modals. However, Yalcin points out several limitations of Kratzer’s approach too, and he ultimately argues for the adoption of a semantic system which incorporates probability measures over propositions.\(^3\)

Let $W$ be a finite set of possible worlds and $w \in W$; the pair $\langle E_w, Pr_w \rangle$ is a probability space, where $E_w \subseteq W$ is the set of worlds epistemically accessible

---

\(^1\) The notation is taken from [18]. The set $O$ contains the propositions assumed to be normally or stereotypically true in the given situation. The preorder $\geq_O$ is defined as follows: $w \geq_O v$ iff $\{ \varphi \in O \mid w \in \varphi \} \supseteq \{ \varphi \in O \mid v \in \varphi \}$.

\(^2\) In the present section we stick to the common assumption that *probable* and *likely* are synonymous.

\(^3\) But see [3] for different qualitative approaches that avoid the problems shown by Yalcin.
from $w$ and $Pr_w : \mathcal{P}(E_w) \to [0;1]$ is an additive measure which assigns 1 to tautologies. The semantics for probable/likely is then defined on the basis of the comparative construction, as follows (the reference to $w$ is dropped):

i. at least as probable/likely as: $\varphi \succcurlyeq \psi$ iff $Pr(\varphi) \geq Pr(\psi)$

ii. more probable/likely than: $\varphi \succ \psi$ iff $Pr(\varphi) \geq Pr(\psi) \land \neg(Pr(\psi) \geq Pr(\varphi))$

iii. probably/likely $\varphi$ is true iff $\varphi \succ \neg \varphi$

It’s worth noticing that once again the meaning of probable/likely is equated to more probable/likely than not. In this setting, however, this is the same as saying that probably/likely $\varphi$ means that the probability of $\varphi$ is bigger than 0.5. On the one hand this seems intuitively plausible and it is indeed what is assumed in several works in psychology concerned with probabilistic reasoning ([14], [16], among others). On the other hand, in the very same works it has been observed that subjects consistently behave in ways seemingly contradicting this assumption. In particular, Teigen finds that subjects are willing to judge as probable events with a probability lower than 0.5 and Windschitl & Wells isolate what they call the alternative-outcomes effect: even if the probability of an outcome $A$ is kept constant, participants tend to judge the probability of $A$ to be higher when $A$ is one of the most likely among a set of alternatives than when $A$ is compared to another more likely alternative $B$.

These observations are generally explained in line with a large body of literature that posits that humans are bad performers in probabilistic reasoning: the meaning of probably/likely $\varphi$ is indeed that $\varphi$ is more probable than 0.5, but humans do not take this into account when reasoning with these expressions. Recently, [6] challenged this line of explanation, proposing that what is responsible for the observed behavior might actually be the semantics of probable/likely and in particular its sensitivity to certain aspects of the context in which these expressions are used.

In a nutshell, Lassiter’s proposal is that probably/likely $\varphi$ means that the probability of $\varphi$ is (significantly) bigger than a threshold $\theta$ which is determined as the average probability of the elements of the event space to which the outcome $\varphi$ belongs. In other words, $\theta$ is computed as $\frac{1}{|Alt|}$, where $|Alt|$ is the cardinality of the set containing the salient alternative outcomes. This set of alternatives is determined by the context in which probably/likely $\varphi$ is used. For example, consider two different contexts (or scenarios) instantiating the setting of the alternative-outcomes effect:

**Background:** A fair lottery with 1000 tickets, one of which is the winning ticket.

---

4 Additivity means that $Pr_w(\varphi \cup \psi) = Pr_w(\varphi) + Pr_w(\psi)$ for any two disjoint $\varphi, \psi \in \mathcal{P}(E_w)$.

5 But see [4] for an interesting exception.

6 The actual explanations are clearly more sophisticated than this. See for example the cited paper by Windschitl & Wells for an explanation of the alternative-outcomes effect based on the assumption that humans can resort to (at least) two kinds of reasoning systems, one being rule-based and the other associative.
**Dual scenario:** John has 420 tickets; Kate has 580 tickets.

**Plural scenario:** John has 420 tickets; 580 other people have one ticket each.⁷

Even if \( Pr(\text{John wins}) \) is equal to 0.42 in both scenarios, it has been observed that a sentence such as *John will probably win* is evaluated as false in what we called the *dual* scenario and true in the *plural* scenario. Lassiter’s proposal can account for this, in that there are two alternative outcomes in the dual scenario and 0.42 is smaller than \( \frac{1}{2} \), while there are 581 alternative outcomes in the plural scenario and 0.42 is significantly bigger than \( \frac{1}{581} \).

Lassiter goes a step further and proposes that in many situations the set of salient alternatives is determined by the denotation of the question under discussion (QUD), similarly to what happens with the explanation of the focus-sensitivity of *always* provided by [1].⁸ The empirical support for this proposal comes from informal surveys conducted by Lassiter where a plural scenario such as the one above is paired with two sentences with different focus patterns, each of which is assumed to be congruent with only one type of QUD (*polar* vs *wh*):⁹

\[
(1) \quad \begin{align*}
&a. \text{ It is likely that} \ [\text{John will win the lottery}] \\
&\implies \text{answers the question} \ \text{Will John win or lose?} \\
&b. \text{ It is likely that} \ [\text{John will win the lottery}] \\
&\implies \text{answers the question} \ \text{Who will win the lottery?}
\end{align*}
\]

According to Lassiter’s findings, the sentence in (1b) is more likely to be evaluated as true than the sentence in (1a) and this can be explained if the set of alternative outcomes, on the basis of which \( \theta \) is computed, is determined by the QUD: a polar QUD structures the event space into two alternatives, setting \( \theta \) to \( \frac{1}{2} \) while a *wh* QUD structures it into many more alternatives, setting \( \theta \) to a much lower value.

Summing up, Lassiter’s theory makes (at least) two crucial predictions about how the context affects the meaning of *probable/likely*. First, what we call the *scenario effect*, analogous to the alternative-outcomes effect introduced above: a sentence such as *probably/likely \( \varphi \)* (where \( \varphi \) has a certain probability known to the listener) is more likely to be judged as true against a *plural* scenario (where \( \varphi \) is compared with several, less likely, alternatives) than a *dual* scenario (where \( \varphi \) is compared with only one, more likely, alternative). Second, what we call the *QUD effect*: if an explicit QUD shapes the way in which the scenario is conceptualized by the speakers, then a sentence such as *probably/likely \( \varphi \)* (where \( \varphi \) has a certain probability known to the listener) is more often judged as true as an answer to a *wh*-question (e.g. *Who will win the lottery?*) than to a *polar* question (e.g. *Will John win the lottery or not*?). In order to test these predictions we ran three experiments, whose methods and results are reported in the following section.

---

⁷ Adapted from [18].

⁸ Applying a number of tests developed by Beaver and Clark, *probable/likely* seem to pattern with pragmatically focus-sensitive *always* (i.e. QUD relative) rather than grammatically focus-sensitive *only*.

⁹ Capitalization indicates focus. The concept of congruence with the QUD is taken from [12].
3 Experiments

We ran three experiments collecting truth-value judgments from native English speakers on Amazon Mechanical Turk. The first two experiments shared largely the same setting, a fair lottery with 1000 tickets (one of which was the winning ticket) where an agent had a significant amount $n$ of tickets. In both experiments we were interested in testing the effects of the context (scenario and QUD) on the judgment rate of sentences involving probability expressions. To this end, the first experiment collected data relevant to three combinations of scenario type and QUD type, while in the second experiment we presented the participants with only one kind of scenario (plural) in order to restrict our attention to the distinction between polar and wh QUDs.

At the same time, we investigated possible differences between linguistic constructions involving probability expressions: in the first experiment we tested three different constructions with probable/-y: the adjectival proposition embedding probable (e.g. It’s probable that Bob will win the lottery), and two variants of the adverbial probably (e.g. Bob will probably win, Bob will probably be the winner). In the second experiment we only tested one kind of adverbial construction, while introducing the distinction between probably and likely.\footnote{See \cite{8} for recent experimental work suggesting that probably and likely are not synonymous, contrary to what is often assumed.}

In a third round of experiments we focused solely on probably and solely on adverbial constructions such as Bob will probably win the lottery. The setting was similar to the one adopted in the first two experiments, except for the total number of tickets, which we changed to 100 hoping to obtain a more plausible setting. The crucial difference between our third experiment and the previous two, however, concerned the way in which we tried to enforce the relevant difference in the conceptualization of the event space, as will become clear below.

3.1 Experiment 1

\textit{Method.} We collected truth-value judgments from 258 participants. Each participant was randomly assigned to one of the three relevant conditions (⟨wh, dual⟩, ⟨wh, plural⟩ and ⟨polar, plural⟩), and saw three control trials and three critical trials. Each trial consisted of the description of the lottery scenario where an agent had a number $n$ of tickets randomly picked from the set {150, 300, 450}, followed by a dialogue between two agents (Mary and John). The scenario was construed in accordance to the scenario type associated with the participant. The participant was asked to read the dialogue and evaluate John’s answer to Mary’s question. In the critical trials, the question was construed in accordance with the QUD condition associated with the participant and the answer instantiated one of the three constructions that we were testing. Each participant saw each $n$ value and each construction exactly once, in random order. An example trial, as seen by the participants, is displayed in Fig.1.
There are 1000 tickets in the lottery and only one of them is the winning ticket. Bob has 300 tickets. 700 other people have one ticket each. John knows this, but Mary doesn’t. Here’s their dialogue:

Mary: Will Bob win the lottery?
John: Bob will probably win.

Is John’s answer true or false?

Fig. 1. Example trial in Experiment 1, plural scenario, polar QUD.

Results. We excluded data points obtained from nine participants.\textsuperscript{11} The mean responses broken down by $n$ values are plotted in Fig. 2.\textsuperscript{12} An overall (and expected) tendency is clear: bigger values of $n$ correspond to higher endorsement rates of John’s answer. The left panel of the picture visualizes the data obtained in the wh QUD condition, broken down by scenario type, and one can clearly see the expected scenario effect: the values for the plural scenario are visibly bigger than the corresponding values in the dual scenario. The right panel visualizes the data obtained in the plural scenario condition, broken down by QUD type, and returns a less clear picture: the values obtained for $n = 300$ and $n = 450$ exhibit a tendency consistent with the QUD effect introduced above, but the error bars are visibly overlapping and the tendency is even reversed for $n = 150$.

\textsuperscript{11} We excluded data obtained from participants who failed at least two out of three control trials.

\textsuperscript{12} The data obtained for different linguistic constructions were aggregated in these plots; error bars represent standard errors.

We analyzed our data with logistic (binomial) regression using the function \texttt{glm} from the package \texttt{stats} under the R programming language ([11]). We used
to compute the fully saturated model, including every factor \((\text{scenario}, \text{question}, n \text{ and item})\) and interaction thereof. The obtained model was given as input to the \texttt{step} function from the same package, which automatically attempts to drop factors and interactions from the full model in order to output the best fitting model according to Akaike’s Information Criterion ([2]). The model we obtained included only two factors, i.e. the \(n\) value and the scenario type, essentially confirming what we informally observed from the plots: the probability of \(\varphi\) and the scenario have a significant effect on the judgment rates of a sentence such as \(\text{probably } \varphi\) \((p < 0.001 \text{ for both factors})\), whereas the expected (and partially observed) difference between QUDs failed to reach significance.\(^{13}\)

### 3.2 Experiment 2

Inspired by the experimental work reported in [15], we ran a second experiment with a modified setup, with the goal of making the QUD more salient for the subjects. The main difference with the previous setup was that the dialogue was hidden at first, and split into two subsequent screens: the participant was asked to click on a button in order to reveal Mary’s question first (e.g. \textit{Who will win the lottery?}), then John’s answer (e.g. \textit{Bob will likely win.}), and then asked to evaluate John’s answer to Mary’s question.

\textit{Method.} We collected truth-value judgments from 168 participants. Each participant was randomly assigned to either diagonal of the condition matrix resulting from crossing two QUDs \((\text{wh and } \text{polar})\) with two item types \((\text{probably and likely})\). Each participant saw two training trials, followed by two control trials and two critical trials.\(^{14}\) Each trial consisted of the description of the lottery scenario followed by the dialogue between Mary and John. In the critical trials, the question and the answer were construed in accordance with the condition associated with the participant. Each participant saw critical items for two different values of \(n\), randomly picked from the set \(\{150, 300, 450\}\).

\textit{Results.} We excluded data points obtained from eight participants.\(^{15}\) The mean responses broken down by item types and \(n\) values are plotted in Fig.3. The panels in the picture share a common feature, i.e. they both display, for \(n\) values equal to 150 and 450, a small tendency consistent with the expected QUD effect: values in the \textit{wh} QUD conditions are slightly bigger than values in the \textit{polar} QUD condition; moreover, in the data collected for \textit{likely}, this holds for \(n = 300\) as well (however, error bars overlap in every condition). The obtained values are puzzling for two reasons: the mean responses obtained in the \textit{wh} condition are

\[^{13}\] The same goes for the difference between different linguistic constructions involving \textit{probable-y}. However, repeating the reported analysis on the data obtained in the \textit{wh} QUD condition only highlight a slightly significant effect of the type of construction, which we take as an indication that further investigations are needed in this respect.

\[^{14}\] After each training trial, participants were given feedback about their answers.

\[^{15}\] We excluded data obtained from participants who failed at least one out of two control trials.
lower than expected (and observed in Experiment 1); the mean responses do not correlate with the number of tickets, especially in the data collected for likely. These results are difficult to interpret and we are unsure as to whether they may point to an interesting difference between likely and probably or to just pin them down to some artifact specific to our experimental setup.

We analyzed our data as in Experiment 1. Consistently to what we observed from the plots, the expected difference between QUD types failed to reach significance.

3.3 Experiment 3

In Experiment 3 we focused solely on probably and solely on adverbial constructions such as Bob will probably win the lottery. Once again our goal was to investigate the sensitivity of the meaning of probably to the way in which the event space is conceptualized by the speakers. The crucial difference with the previous experiments is that in Experiment 3 we tried to enforce the wanted difference in conceptualization (polar vs wh) not only by having the speaker explicitly asking the appropriate question but also by describing the lottery context and the involved speakers in a way that made clear why the speaker asked the question she asked, the reason behind her request of a certain body of information. More in detail, we paired polar questions about a specific person with contexts in which the speaker was really interested in what would happen to that person (e.g. because she is his mother or partner); on the other hand, wh-questions were paired with more generic contexts in which the speaker had no apparent reason to be interested in anyone in particular.

Method. We collected truth-value judgments from 117 participants. We ran two different versions (A and B) of Experiment 3. Each participant was randomly assigned to one of the two versions and to either condition polar or wh. Each participant saw three control trials and three critical trials (one for each value
of $n$ in the set $\{25, 35, 45\}$). Each trial consisted of the description of the lottery scenario followed by a dialogue between Mary and John. The question asked by Mary was determined in accordance to the condition. The differences between version A and B concerned John’s reply to Mary and the question we asked to the participants. In version A the reply was always a sentence such as *Bob will probably win* and the participants were asked to evaluate its truth. As an illustration, consider the example trial in Fig.4.

First prize is a brand new bicycle. The tickets were sold to six people, as summarized in the following table [...]. Mary is Bob’s parent and knows how much Bob wants a new bicycle. Mary is chatting with John, who organized the lottery. Here’s their dialogue:

Mary: Do you think that Bob will win the first prize, or not?
John: Bob will probably win.

Is John’s answer true or false?

Fig. 4. Example trial in Experiment 3, version A, polar QUD.

In version B we kept this structure but we manipulated not only the description of the context and speakers and the question asked, but also the answer given by John. For example, in the *polar* condition, Mary always asked a question such as *Do you think that Bob will win the first prize, or not?* and John replied with *Bob has less than half of the tickets,* whereas in the *wh* condition Mary always asked a question such as *Who do you think will win the first prize?* and John replied with *Bob has more tickets than anybody else.*

In both conditions the participants were asked to evaluate the truth of a sentence such as *Bob will probably win the first prize.* The idea behind this manipulation was that in the *polar* condition the reply highlights the number of Bob’s tickets in opposition to the total number of tickets owned by the other players, thereby facilitating a -so to speak- polar conceptualization of the scenario, whereas in the *wh* condition the reply highlights the number of Bob’s tickets in opposition to each of the other players, thereby facilitating a -so to speak- wh conceptualization of the scenario.

Results. We excluded data points obtained from fifteen participants.\(^{16}\) The mean responses broken down by versions and $n$ values are plotted in Fig.5. First we can observe that, overall, the tendency in the data is consistent with the expected QUD effect, to a somewhat clearer degree than what we observed in the previous experiments. With the exception of the data collected in version A for $n = 25$, the mean responses in the *wh* condition are always higher than the mean responses in the *polar* condition. However, error bars overlap in most conditions, and the data show some puzzling features as well. In particular, it is not clear...

\(^{16}\) We excluded data from participants who failed the majority of control checks.
to us why the mean responses obtained in version B are in general lower than the ones obtained in version A, which in turn are higher than expected (and observed in Experiment 1).

We analyzed our data using the functions `glm` and `step`. We computed the fully saturated model including every factor (`version`, `condition` and `n`) and interaction thereof. Then we called the `step` function on the obtained model. The output model included all factors. First of all, the number of tickets `n` was included but its effect failed to reach significance, in accordance to what we observed above. Moreover, as expected, the difference between version A and B was significant (`p < 0.001`). Finally, the overall effect of the QUD was marginally significant as well (`p < 0.05`). However, it has to be noted that when we analyzed the data from version A and version B separately, as if they belonged to two altogether different experiments, the QUD effect was significant only in version B.

Nonetheless, from the point of view adopted in this paper, i.e. a first study of the context-sensitivity of probability expressions, we believe that these results are more promising than the ones obtained in the previous experiments. In fact, even if an explicit manipulation of the question asked by the speaker (together with an explicit description of why that question matters to her) is not enough to elicit a significant difference in the interpretation of `probably`, we observe that QUD needs not be the only factor that concurs to shape the way in which speakers conceptualize the event space. As a consequence, further manipulations of the linguistic context where an utterance of `probably` takes place can enforce different conceptualizations which ultimately affect the interpretation of the utterance.

4 Conclusions

We ran three experiments in order to investigate the effects of the context on the interpretation of probability expressions such as `probably` and `likely`. At the
same time we investigated possible differences in the interpretation of different linguistic constructions involving these expressions.

Starting from the second point, we cannot draw a conclusion yet but we believe that our results, especially in Experiment 1, invite us to further investigate the issue whether different constructions are interpreted differently and how.

As for the first point, the real focus of this paper, we believe that our results are interesting for a number of reasons. First of all, the significant scenario effect observed in the first experiment is in accordance with much previous literature and confirms that our experimental setup, essentially a truth-value judgment task in a lottery situation, is sound and can be taken as a starting point for future developments. However, we believe that a thorough investigation of probability expressions should take into consideration more diverse uncertainty situations as well, ranging from colored balls in an urn to weather forecasts, to name only the most obvious ones.

Moving to Experiment 2, our failure to elicit a significant QUD effect opens up interesting possibilities for future work. From an experimental perspective, the challenge is to explore different experimental setups more sensitive to QUD-induced differences. We began to do so in Experiment 3, whose results are consistent with the expected QUD effect albeit enforced by a manipulation of other components of the linguistic context as well. Clearly, more work needs to be done in order to disentangle what seem to be several different components of the context-sensitivity of probability expressions.

Finally, from a theoretical perspective we believe that it would be interesting to pursue a more nuanced theory of the context-sensitivity of probable/likely, possibly following a lead suggested by Lassiter (p.c.) according to which the threshold $\theta$ is not semantically fixed but pragmatically inferred by the speakers, similarly to what happens in recent probabilistic theories of adjectival vagueness ([7], [10]).

References

On the Semantics and Pragmatic Underpinning of Essentially Plural Nouns

Yuki Ito
University of Maryland, College Park, USA.
yito@umd.edu

Abstract. Nouns such as friends and sisters are ambiguous between the essentially plural noun (EPN) reading “friends of each other” and the non-EPN reading “friends of someone.” Building on Hackl’s (2002) compositional analysis, this paper argues that English speakers can derive the EPN reading through a grammatical means and that this derivation is subject to a pragmatic condition. The virtue of the analysis is that it provides an account of Hackl’s generalization regarding the possible relational noun sources for EPNs. Empirical evidence against a lexicalist treatment is provided.

1 Analysis

In Hackl’s taxonomy of three kinds of plurality, essentially plural predicates like friends (with reciprocal meaning) are on the one hand similar to genuine collective predicates such as team and committee in being only true of collections of individuals and on the other hand similar to pluralized individual predicates like boys in bearing plural marking (see also Dowty 1987, Winter 2002, Champollion 2010 for distinctions between pluralities).

The singular counterpart of EPNs has a relational usage (e.g. John is a friend of Sue). Correspondingly, sentences like (1) John, Mary, and Phil are friends are ambiguous between the reading where friends is interpreted as an EPN and the one where it is understood as a pluralized relational noun whose internal argument is provided by the discourse.

(1) John, Mary, and Phil are friends.

* Editorial Note: To our greatest shock the author passed away in untimely fashion shortly before presenting his work at the ESSLLI Student Session. In order to commemorate him we decided to include the last version of his paper that he had sent us into this volume.

1 Compatibility with all is known to diagnose between essentially plural predicates and genuine collective predicates across categories:

(i) (a) All the students are friends/similar/gathered. (essentially plural predicates).
    (b) *All the students are a good team/numerous/formed a pyramid. (genuine collective predicates)
a. EPN-reading: John, Mary, and Phil are friends of each other.
b. Non-EPN reading: John, Mary, and Phil are friends of someone salient in the discourse.

There is also another kind of non-EPN reading where the noun is understood as a pluralized relational noun whose internal argument is existentially quantified rather than provided by the discourse. This reading is salient with nouns like *brothers*:

(2) John, Max, and Tom are brothers.
   a. EPN-reading: John, Max, and Tom are brother of each other.
   b. Non-EPN reading: John/Max/Tom is a brother of someone.

Hackl proposes to derive EPNs from their relational counterpart:

(3) a. John, Mary, and Phil are friends.

   b. John, Mary and Phil
      are
      pro
      pro
      **pro
      **pro
      friends
      friends

   c. friends = **friend
   d. friend = λxλy : y ≠ x. y is a friend of x
   e. **R(x)(y) = 1 iff R(x)(y) = 1 or ∃x1x2y1y2 : x1 ⊕ x2 = x & y1 ⊕ y2 = y & **R(x1)(y1) = 1 & **R(x2)(y2) = 1
   f. [[[pro3[friends pro3]]] = λx. x **(is a friend of) x
      (the set of plural individuals whose parts stand in a friend relation to each other)
   g. j ⊕ m ⊕ p ∈ {x : x**(is a friend of) x = 1}
      (the individual denoted by John, Mary, and Phil is a plural individual whose parts stand in a friend relation to each other)

The co-indexing of pro and the pluralization create number-neutral reflexivization g(3) **(is a friend of) g(3). Because of the irreflexivity presupposition associated with the relational noun source (one cannot be one’s own friend), the lambda abstract created from this ((3f)) can only be satisfied by a plural individual (cf. Beck 2001 on the inheritance of presuppositions under application of the ** operator). The irreflexivity presupposition also serves to exclude the atomic reflexive x is a friend of x from the plural reflexivization x **(is a friend of) x. This exclusion of the atomic reflexive from the plural reflexivization gives rise to the reciprocal interpretation, a characteristic property of EPNs.

2 (3b) adapts Hackl’s syntactic analysis. To create the lambda abstract, Hackl basegenerates the subject in the external argument position and raises it to the surface position. Postulation of pro in the external argument position is my modification to deal with non-predicative uses such as *Three sisters came in.*/Sisters should always try to have separate bedrooms.*
What Hackl has shown is that EPNs can be compositionally derived without a designated rule. Yet, there is still a sense in which EPNs are idiomatic, for the coindexing in (3b) does not follow from independent principles. I propose that this paradoxical situation arises because while some EPNs are possibly stored in the lexicon as such, English speakers can also derive the EPN reading in the style of Hackl. I argue that the availability of the structure along (3b) is subject to a pragmatic condition; specifically that it is the existentially quantified non-EPN reading we have seen above that conditions the EPN reading.

The motivation for the analysis comes from Hackl’s generalization regarding the possible relational noun sources for EPNs:

(4) **Hackl’s generalization**

Inherently symmetric relations that have a presupposition of non-identity on their arguments have essentially plural predicate counterparts. (e.g. Relational nouns that encode relations that are not inherently symmetric do not have EPN usage: *John, Mary, and Phil are fans/reviewers. ≠ John, Mary, and Phil are fans/reviewers of each other.*)

Given the irreflexivity presupposition’s role in deriving reciprocal meaning we have seen, this part of the generalization makes sense. My proposal is that the remaining part of the generalization-limitation to lexically symmetric predicates—is the key in understanding the EPN reading. The idea is that (3b) is a way to resolve the referential indeterminacy that occurs with the existentially quantified non-EPN reading. Notice that the EPN reading entails the existentially quantified non-EPN reading:

(5) a. (3g) reduces to “John, Mary, and Phil are friends of each other”

b. If “John, Mary, and Phil are friends of each other”, then “John/Mary/Phil is a friend of someone”

In this sense, the EPN reading is a strengthening of the existentially quantified non-EPN reading. What is crucial to observe here is that the amount of information added by this strengthening is smaller with symmetric than with nonsymmetric predicates. This is because via lexically encoded symmetry, “John/Mary/Phil is a friend of someone” entails “Someone is a friend of John/Mary/Phil” whereas non-lexically symmetric predicates like *reviewers* do not.

---

3 I follow Schwarz (2006) in taking the generalization to be based on Strawson-symmetry:

(i) A relational noun N is Strawson-symmetric iff “X is a N of Y” Strawson-entails “Y is a N of X”

φ Strawson-entails Ψ if and only if the conjunction of φ and the presupposition of Ψ entails Ψ.

Schwarz’s point is that the gender information on apparently nonsymmetric predicates like *sister* is presuppositional:

(ii) a. Kim isn’t his sister.

b. Perhaps Kim is his sister.

c. Is Kim his sister?
not have this entailment (“John/Mary/Phil is a reviewer of someone” does not entail “Someone is a reviewer of John/Mary/Phil”). Thus, the amount of information added in the EPN reading is as follows:

(6) a. lexically symmetric predicates
   \{J/M/P is a friend of someone among J,M,P\} \& \{Someone among J,M,P is a friend of J/M/P\}

b. non-lexically-symmetric predicates
   \{J/M/P is a reviewer of someone among J,M,P\} \& \{Someone among J,M,P is a reviewer of J/M/P\}

So, on the one hand, there is an impetus to resolve the referential indeterminacy with lexically symmetric predicates and on the other hand, with non-lexically-symmetric predicates there is an information-theoretic consideration not to add too much information without phonetic exponents. Thus, if the EPN reading is pragmatically conditioned by the existentially quantified non-EPN reading, the limitation of EPN usage to lexically symmetric predicates is explained.

While it is possible that this pragmatic consideration has only played a diachronic role such that lexically symmetric predicates such as *friends* have undergone lexicalization as EPNs, I will show that the EPN reading has a certain level of productivity, a fact that argues for the synchronic reality of structure (3b) (note that this does not deny that they are stored as words at the same time).

2 Evidence

If EPNs are stored as such in the lexicon, the interpretation of EPNs cannot reference syntax. However, there exists a case where the EPN reading has to refer to syntactically composed structure. This comes from EPNs involving conjunction:

(7) *These 100 people are brothers and sisters.*
    TRUE if all the males among the 100 are brothers of some female among the 100 and all the females among the 100 are sisters of some male among the 100.

For instance, note that the referential indeterminacy does not arise under embedding by negation, which predicts, under the current analysis, the absence of the EPN reading, contrary to fact. However, it may be that this problem does not arise because EPNs are treated as words. Nevins and Myler (2014) discuss expressions like *brown-eyed*. These expressions are parasynthetic in the sense that modification is obligatory (e.g. *eyed, *legged). The authors suggest that this is due to a semantic/pragmatic restriction requiring informativeness (cf. ??*John has eyes*), which explains the existence of some modifierless versions (e.g. *bearded, winged*). What is relevant for our current concerns is that if informativeness in this sense is evaluated at the level of the whole sentence, sentences like *John isn’t eyed* and *This snake is legged* are predicted to be allowed. Nevins and Myler propose that because these expressions are words, informativeness is evaluated at the level of the word. The same account could extend to our case.
Here, reciprocity is defined over the entire conjunction, not within each conjunct. That is, there is no set of people who are brothers of each other; nor is there any set of people who are sisters of each other (see Matushansky and Ionin 2011 for a similar observation). Given the fact that predicate conjunction can have weak interpretations:

(8) a. The birds are above the cloud and below the cloud. (Winter 2001:338)
   b. Mary and John sent and received these letters. TRUE if these letters
      refers to letters A and B; A was sent by Mary and received by John; B
      was sent by John and received by Mary (Winter 2001:349)

this reading can be derived via the structure in (3b) (*brothers and sisters* in place of *friends*), but a lexicalist treatment fails to do so because by definition it cannot operate on syntactically composed structures (conjunction here). 5

Since the compositional derivation of the weak interpretation of predicate conjunction is beyond the scope of this paper (see e.g. Chaves 2012 for a proposal), the following descriptive generalization suffices:

(9) Weak interpretation of predicate conjunction (two-place predicates) 6

\[
\llbracket [\text{DP}_1 \text{Pred}_1 \text{DP}_2] \rrbracket = \lambda x. x (**(R_1 \text{ or } R_2)(x)(y))
\]

where \( \llbracket \text{DP}_1 \rrbracket = y, \llbracket \text{DP}_2 \rrbracket = x, \llbracket \text{Pred}_1 \rrbracket = R_1, \llbracket \text{Pred}_2 \rrbracket = R_2 \)

\( \llbracket (8\text{b}) \rrbracket = **(\text{sent or received})(\text{these letters})(\text{Mary and John}) \)

With this weak interpretation, the desired reading is derived as follows:

(10)
Note that it is of course also possible to apply the Hackl-style derivation within each conjunct. For instance, by virtue of the DN and N construction (e.g., *Ten men and women got married today*, see Heycock and Zamparelli 2005, Champollion 2014), the following is coherent (because of structural ambiguity):

(11) No brothers and sisters came. But these 100 people came. These 100 people are brothers and sisters.

The point is rather that the reciprocity taking scope over a non-lexical predicate is not expected if EPNs are only stored as words in the lexicon.  

3 Concluding remarks

In this paper, I have argued that English speakers can derive the EPN reading through a grammatical means and that this derivation is subject to a pragmatic condition, providing an account of Hackl’s generalization regarding the possible relational noun sources for EPNs.

Before finishing, two remarks are in order concerning the scope of the current analysis. First, under this analysis, what is crucial for the “essential” plurality of EPNs are filling in of the two argument positions of their relational counterpart with pro. Accordingly, the analysis extends to those essentially plural adjectives (EPAs) that have a relational counterpart such as similar and different, but not to essentially plural verbs (EPVs, e.g. gather, disperse), for which it is difficult to identify a relational source. This is in fact a welcome consequence. Empirical evidence for the “essential” plurality of EPNs comes from the interaction of EPNs and group nouns: EPNs can take group nouns as their subjects, only when the group noun induces a plural agreement:

(12) a. *The committee is friends.* (Winter 2002)
    b. The team are friends on track as well as off track, and are as much family as we are friends. (British English, de Vries 2013)
    c. Half of the committee are friends. (cf. Pearson 2011 on plural agreement with partitives)

Under the common view that group nouns denote pluralities only when they show plural agreements (with their standard denotations being atoms/singularities) (e.g. Barker 1992, Schwarzschild 1996), this is explained if EPNs denote predicates of pluralities because a predicate of pluralities cannot be true of a singularity. Now, whereas EPAs show the same pattern as EPNs regarding the compatibility with group nouns, EPVs can take group nouns with a singular agreement as their subject (Winter 2002):

---

An issue that arises with this analysis, pointed out by a reviewer, is that the weak interpretation of predicate conjunction creates symmetric predicates for cases such as employers and employees and teachers and students, which leads to the prediction that the Hackl-style derivation should be available with these phrases (and yet these phrases sound worse than brothers and sisters). One possibility is that this is a case of garden-path effects. These phrases are also compatible with a parse as reciprocal conjunction (cf. fn.5); this could possibly block the parse as an EPN.
Thus, this fact seems to suggest that EPVs like *gather* require a separate treatment.

Next, covert reciprocal verbs such as *kiss* and *meet* (see Dimitriadis 2008) pose an interesting question. As we have seen, the Hackl-style derivation creates predicates of plural individuals. Given the broad recognition that verbs are predicates of events rather than individuals (see Maienborn 2010 for a recent overview), whether the current analysis extends to covert reciprocal verbs is an issue that merits a closer investigation.

References

A Family of Exclusives in Hebrew

Dina Orenstein

Bar Ilan University, Ramat Gan 52900, Israel
dinaorenstein@gmail.com

Abstract. The paper deals with three Hebrew exclusives: rak (roughly, only), stam (roughly, merely), and be-sax ha-kol (roughly, in sum the whole). These particles give rise to a complex set of semantic and pragmatic effects. I suggest that despite the apparent differences between them, they all share a core semantic exclusive meaning, which is a modified version of the scalar accounts of Beaver & Clark 2008 [2] and Roberts 2011 [14]. I further suggest that the differences between these expressions are attributed to three varying parameters: (a) type of scale - entailment or evaluative, (b) type of alternatives - internal or external and (c) origin of the scale - supplied by context or by the lexical structure.

Keywords: exclusives, alternatives, scale

1 Introduction

In this paper I discuss three Hebrew expressions: rak (roughly, only), stam (roughly, merely), and be-sax ha-kol (literally: in sum the whole). In some cases these three expressions seem to lead to an identical semantic effect, as can be seen in (1):

(1) Rina rak/stam/be-sax ha-kol [pkida]F
    Rina rak/stam/be-sax ha-kol [clerk]F
    “Rina is only / merely / be-sax ha-kol a [clerk]F.”

With all three expressions in (1) we get the reading that Rina is a clerk, and not more than that, e.g. she is not a doctor. However, we also find many interpretational and distributional differences between the three expressions, as I show in detail in the next section.

To account for a full range of data I suggest that these three expressions constitute a family of exclusive expressions. I.e., I propose that the similarities between them are due to their shared core meaning, and the differences are due to three parameters along which these expressions differ. In particular, all three expressions are ‘scalar exclusives’, i.e., they reject stronger alternatives than the prejacent on a salient scale. They differ along the following three varying parameters.

* Special thanks to Yael Greenberg for her unconditional support and insightful comments.
1. Type of scale (entailment or evaluative)
2. Type of alternatives in the scale (Roothian / ‘external’ or ‘internal’)
3. Origin of alternatives (supplied by context / by the lexical semantics of the associate)

The paper is structured as follows. In Sect. 2 I present the data, and in Sect. 3 I briefly present the core definition of exclusives. Each of the varying parameters is discussed in Sections 4-6. Finally, in Sect. 7 I summarize the analysis.

2 Data

As we have seen in (1) above, it seems that the exclusives rak (only), stam (merely) and be-sax ha-kol (‘in sum the whole’) can all be used to express the same kind of meaning. In other sentences, however, these exclusives differ in their interpretation and distribution. As rak is the standard typical exclusive, I compare the other two exclusives to rak. In what follows I first discuss the data of stam and then the data of be-sax ha-kol.

2.1 Stäm.

Looking at the data, we reveal two instantiations of stäm in Hebrew: unaccented stäm and accented stäm. Accented and unaccented stäm differ in their interpretation and distribution. I discuss each instantiation of stäm in the subsections below.

Unaccented Stäm. Consider (2) below:

(2) Danny hu rak/stäm [student]F
Danny he rak/stäm [student]F
“Danny is rak/stäm a [student]F.”

Example (2) with rak has two readings. The first reading is that Danny is a student, and he does not have any other occupations. E.g., he does not have a job in addition to being a student. The second reading is that Danny is a student, but not more than that, i.e. he is not ranked higher on the academic scale. E.g., he is not a post doc., or a professor. In contrast, (2) with stäm has only the second reading, namely, that Danny is a student and that he is not ranked higher on the academic scale. This interpretational difference may lead to felicity differences. For example, (3) is felicitous with rak but not with stäm.

(3) hu rak/#stäm zaxa [be-pras Nobel]F
he rak/#stäm won [in.prize nobel]F
“He only/#stäm won the [Nobel Prize]F.”

Example (3) with stäm is infelicitous because stäm can only have the “nothing stronger than” reading. But as no prize is considered more important than the
Nobel Prize, the use of *stam* is uninformative and leads to infelicity. In contrast, *rak* is felicitous here, as it implies that Danny won no prize besides the Nobel Prize, e.g. he did not win the Wolf Prize and he did not win the Israel Prize etc.

Finally, we observe that contrary to *rak*, unaccented *stam* is infelicitous with numerals, as can be seen in (4):

(4) ha-simla ola 100 shkalim ve-li ye$ rak/#stam [50]
    The.dress costs-3.f.sg 100 shekels and.me there.are rak/#stam [50]
    shkalin]_F
    shekels]_F
    “The dress costs a hundred shekels and I have rak/#stam [50 shekels]_F.”

**Accented Stam.** There are clear interpretational differences between unaccented *stam* and accented *stam*. To illustrate, consider (5) and (6)

(5) ze ma she-rina kanta? ze stam SAON!
    This what that.rina bought? This stam WATCH
    “That’s what Rina bought? It’s *stam* a WATCH.”

(6) ze ma she-rina kanta? ze STAM Saon!
    This what that.rina bought? This STAM watch
    “That’s what Rina bought? It’s STAM a watch.”

The only difference between (5) and (6) is in accent placement. (5) with unaccented *stam* implicates that (a) Rina bought a watch and that (b) she did not buy anything more valuable / expensive than a watch (e.g. not a dishwasher). But unlike (5), (6) implicates that (a) Rina bought a cheap / simple watch, and (b) she did not buy any more valuable watch, i.e. not any more expensive / special watch.

Similar differences between *stam* and *STAM* are found in the following examples, where the main predicate is verbal and not nominal:

(7) dani stam RAKAD im rina
    Danny stam DANCED with Rina
    “Danny stam DANCED with Rina.”

(8) dani STAM rakad im rina
    Danny STAM danced with Rina
    “Danny STAM danced with Rina.”

In (7) with *stam* we get the implication that dancing with Rina is considered less significant / noteworthy than other activities with Rina (e.g. kissing Rina). In contrast, the version in (8) with *STAM*, can have at least two potential implications: (a) that John’s manner of dancing with Rina was simple or not noteworthy, and (b) that John’s dancing with Rina was unintended and casual.
Interestingly, the interpretational difference sometimes results in distribu-
tional differences. In Sect. 2.1 above I have shown that (3) with unaccented stam is infelicitous, because no prize can be stronger than the Nobel Prize. But, if we put accent on stam in the sentence in (3), as in (9) below, the sentence is perfectly felicitous:

(9) hu STAM zaxa be-pras nobel
    he STAM won-3.m.sg in.prize nobel
    “He STAM won the Nobel Prize.”

The reading we get in (9) is that he got the Nobel Prize unjustly, e.g. he did not really deserve the prize. Notice that this phenomenon is not found with rak. It is definitely possible to accent rak in the context of (9) (inspired by an example in Kadmon & Sevi 2011 [9], but the reading is basically the same as with unaccented rak:

(10) a. ma yeS ba-kufsa?
    what there.is in.box?
    what’s in the box?

b. Saon. ha-beaya hi Seze rak/RAK Saon
    Watch. the.problem she that.it rak/RAK watch
    ‘A watch. The problem is that it’s rak/RAK a watch.’

2.2 Be-sax Ha-kol

Remember that in many cases the use of be-sax ha-kol leads to a standard exclusive reading similar to rak, as we have seen in (1) above. However, I observe that it can yield two additional readings, not found with rak and with stam: an ‘approximative’ reading and a ‘precise’ reading. When be-sax ha-kol combines with gradable adjectives it many times yields an approximative reading (similar to more or less). The approximative reading of be-sax ha-kol is illustrated in e.g. (11):

(11) ha-xeder be-sax ha-kol [naki]_F
    The.room be-sax ha-kol [clean]_F
    “The room is be-sax ha-kol [clean]_F.”

Example (11) has the reading that though the room is considered clean, it is not maximally clean. Note that this reading is not obtained with rak or with stam/STAM:

(12) ha-xeder rak/stam/STAM [naki]_F
    The.room rak/stam/STAM [clean]_F
    “The room is rak/stam/STAM [clean]_F.”
An interesting point to note with respect to the approximative reading of be-
sax ha-kol is that there are gradable adjectives with which it cannot yield this reading, as in (13):

(13)  

* #ha-xeder be-sax ha-kol [meluxlax]_F  
  the.room be-sax ha-kol [dirty]_F

  “The room is be-sax ha-kol [dirty]_F.”

The precise reading of be-sax ha-kol is obtained when it associates with numeral expressions, as illustrated in (14):

(14)  

* higiu be-sax ha-kol [30]_F anaSim la-mesiba  
  came-3rd.pl. be-sax ha-kol [30]_F people to.party

  “be-sax ha-kol [30]_F people came to the party.”

The implication of (14) is that exactly 30 people came to the party. This reading is puzzling when compared to the two other readings of be-sax ha-kol. On the one hand, it differs from the approximative reading. As we have just seen, when be-sax ha-kol associates with adjectives it implies that the degree to which the adjective holds of the subject is less than the maximum. E.g., in (11) be-sax ha-kol implies that the room is more or less clean (not completely, ‘precisely’ clean), whereas on the precise reading it implies that the cardinality denoted by the numeral it associates with is precise. On the other hand, it also differs from the exclusive reading. Remember that on the exclusive reading of be-sax ha-kol it marks that nothing stronger than p holds. Consequently, it is felicitous in contexts where stronger alternatives are under consideration, and not vice versa, as illustrated in the contrast between (15) and (16) below:

(15)  

* xasavti Se-Rina rofaa, aval hi be-sax ha-kol/rak [pkida]_F  
  thought-1.sg. that.rina doctor, but she be-sax ha-kol/rak [clerk]_F

  “I thought that Rina was a doctor, but she is be-sax ha-kol/only a [clerk]_F.”

(16)  

* #xasavti Se-Rina pkida, aval hi be-sax ha-kol/rak [rofaa]_F  
  thought-1.sg. that.rina clerk, but she be-sax ha-kol/rak [doctor]_F

  “I thought that Rina was a clerk, but she is be-sax ha-kol/only a [doctor]_F.”

However, surprisingly, when combined with numerals, be-sax ha-kol (but not rak), is felicitous in the presence of a weaker alternative in the context:

(17)  

* tsipiti Se-yagi 10 orxim. ba-sof hayu  
  expected-1.sg. that.come 10 guests. In.the.end were

  #rak/be-sax ha-kol [20]_F

  rak/be-sax ha-kol [20]_F

  “I expected 10 guests to arrive. Eventually there were #rak/be-sax ha-kol [20]_F.”
3 The Core Meaning of Exclusives

The definition I suggest in (18) is in the spirit of scalar accounts of only (e.g., Beaver & Clark 2008 [2], Kadmon & Sevi 2011 [9], Coppock & Beaver 2014 [4] and mainly Roberts 2011 [14]) with some modifications. Note that the underlined material is not at issue, and $q >_s p$ is true if $q$ is higher than $p$ on the scale $s$. We take $C$ to be a relevant subset of the focus semantic value of $p$, but unlike Rooth 1985 [16]; 1992 [17], not necessarily supplied by the context.

(18) The lexical entry for exclusives:

\[
\lambda C.\lambda p.\lambda w \forall q[q \in C \land q \text{ is salient}] \rightarrow q >_s p \land w \in p \land \forall q \in C \quad q >_s p \rightarrow w \notin q.
\]

Where \( C \subseteq \|p\|^F \land \|p\|^0 \in C \land \exists q \ q \neq p \land q \in C \)

Given this definition, an exclusive particle is a sentential operator (following Beaver & Clark 2008 [2], Roberts 2011 [14])\(^1\), which combines with a proposition $p$ (the prejacent), $C$, i.e., a relevant set of focus alternatives to $p$, and a world, $w$. It has two not at issue components saying (a) that every salient proposition in $C$ is stronger than the prejacent, and (b) that the prejacent is true in $w$. It asserts that every alternative $q$ stronger than the prejacent on a salient scale is false in $w$.

The scale we refer to in (18) can be either entailment based, or non-entailment / evaluative based (Beaver & Clark 2008 [2]). Note that this definition overcomes some problems in previous definitions. See Orenstein 2015 [11], Orenstein & Greenberg 2012 [12] for a detailed discussion.

Though all three Hebrew exclusives share the core meaning in (18), we suggest that they differ along three parameters. In what follows I present each of the parameters and show how they account for various differences between these exclusives.

4 Parameter #1: Type of Scale: Entailment vs. Evaluative (non-Entailment)

I suggest that rak (and only) can operate on both entailment and evaluative scales, but stam is restricted to non-entailment / evaluative scales. Be-sax ha-kol is compatible with both scales, but seems to prefer evaluative scales. First, stam’s inability to operate on entailment scales is supported by its infelicity with numerals as in (4) above, clearly because numerals usually trigger an entailment scale. This restriction accounts also for the interesting felicity difference between stam and rak in (3), repeated here as (19),

(19) hu rak/#stam zaxa [be-pras Nobel]_F
he rak/#stam won [in.prize nobel]_F

“He only/#stam won the [Nobel Prize]_F.”

\(^1\) But see Coppock & Beaver 2014 [4] for a non-sentential analysis. This analysis takes different exclusives to differ along the type parameter, in addition to other parameters.
In principle, exclusives cannot combine with maximal elements, because in such cases the main contribution of the exclusive is vacuous (it has nothing to reject). Now, on an evaluative scale, winning the Nobel Prize is maximal, so *stam* cannot associate with it. But if we consider an entailment scale (ordered by the number of prizes one wins), *rak* can be felicitous, as in e.g. *Yossi won both the Nobel prize and the Israel Prize. Danny only won the Nobel Prize*.

5 Parameter #2: Types of Alternatives in the Scale:  
(‘External’ vs. ‘Internal’)

It is standardly assumed that focus sensitive expressions operate on alternatives which are identical to the prejacent, besides the focused (usually the stressed) element. This element is replaced by an element of the same semantic type (Rooth 1985 [16]). But in addition to expressions which are sensitive to such ‘Roothian’ / ‘external’ alternatives, some expressions can also operate on ‘internal’, non-Roothian alternatives (Cf. Greenberg & Khrizman 2012 [7], Greenberg 2014 [6], Chierchia 2012 [3]). I observe that two of the members of the family of Hebrew exclusives operate on internal alternatives, namely, *be-sax ha-kol* and accented *stam*. I will start with *be-sax ha-kol*. As we have seen in 2 above, while *rak* only has an ‘exclusive’ effect, *be-sax ha-kol* can also have an ‘approximative’ effect (similar to e.g. *more or less*). This observation is clearly illustrated in the scenario in (20):

(20) Context: John and Mary booked a room in a hotel for their important guests and asked that the room will be clean, large, with view to the sea. After John checks the room he tells his wife:

a. ha-xeder *rak* [naki]_p
   The.room only [clean]_p
   “The room is only [clean]_p .”

b. ha-xeder *be-sax ha-kol* [naki]_p
   the.room *be-sax ha-kol* [clean]_p
   “The room is *be-sax ha-kol* [clean]_p .”

The implication of (20a) is that the room is clean, but not more than that. The potential alternatives to *p* here are e.g. *The room is clean and large / The room is clean and large and has a view to the sea*. Alternatively, John tells his wife (20b). The implication of (20b) is that the room is not maximally, very clean. The potential alternatives to *p* here are *The room is very clean / The room is maximally clean*. I assume, then, that with the approximative reading *be-sax ha-kol* still functions as an exclusive operator, rejecting higher alternatives, but all alternatives are ‘internal’ – different versions of the prejacent. In the case of (20b) the prejacent is (21a), and more formally (21b) (following e.g., Kennedy & McNally 2005 [10]):

(21) a. The room is pos clean
b. \( \exists d \geq \text{stand(clean)} \land \text{clean(the room)}(d) \)

Briefly, the different alternative interpretations of (21a) result from the potential variability in the characterization of stand in (21b). be-sax ha-kol rejects (21b) under an interpretation of stand (clean) as \( d_{\text{max}} \), which is the salient standard for upper closed adjectives (following Kennedy & McNally 2005 [10]). It accepts (21b) under a lower characterization of stand(clean). We end up with The room is not maximally clean, but its degree of cleanliness is still high enough to be considered clean. To account for the felicity effect, i.e. the fact that on the approximative reading be-sax ha-kol is infelicitous with lower closed adjectives, we will apply the definition of exclusives to (13).

(13) \#ha-xeder be-sax ha-kol [meluxla]_F
the.room be-sax ha-kol [dirty]_F
“The room is be-sax ha-kol [dirty]_F.”

Following Kennedy & McNally 2005 [10], the salient standard for such adjectives is the minimal point in the scale (just above the zero point). In the case of (13), then, the presupposed salient alternative is The degree to which the room is dirty is equal to or higher than the minimal point, and crucially, this alternative is required to be stronger than the prejacent, i.e. stronger than The degree to which the room is dirty is equal to or higher than the standard of dirtiness. But, unlike what we saw with upper closed adjectives like clean, in this case, using a different standard will not work: If the salient standard is the minimal degree of dirtiness, then no standard can be lower than it. Thus, there is no way to satisfy the presupposition, and the sentence is infelicitous due to presupposition failure.

The second particle that operates on internal alternatives is accented stam. Remember that accenting stam results in a different interpretation. Example (5) above with unaccented stam implies that it’s a watch and not more than that. The potential alternatives in this case are e.g. It’s a watch < It’s a dishwasher < it’s a car < it’s a house. In contrast, when stam is accented, as in (22),

(22) kibalti Saon, ha-beaya hi Se-ze STAM shaon!
Got.I watch the.problem she that.it STAM watch
“I got a watch. The problem is that it’s STAM a watch!”

The implication is that it’s a simple / cheap watch, and not more than that. The potential alternatives in this case are e.g. {It’s a cheap watch < it’s an expensive watch} or alternatively {It’s a simple watch < it’s a special watch}. I suggest, then, that just like unaccented stam, accented STAM is also an exclusive operator (rejecting higher alternatives on an evaluative scale), only here the alternatives are ‘internal’, in the sense that they are different versions of This is a watch, with different sub-categories of watch. I further suggest that unlike the case with be-sax ha-kol, here the internal alternatives are constructed from the interaction of the semantics of exclusives with information structural and prosodic factors. In particular, I propose, following Egg & Zimmerman 2011 [5] and Greenberg 2014 [6], that accent shifts to stam because its associate is given in the context, and hence, de-accented. Crucially, however, as an exclusive it
still associates with it’s a watch in (22). Assuming that the associate of accented *stam* is given means that a correlate of it is salient in the discourse. I.e., the salient alternative to the prejacent is identical to it, as in e.g. (23):

(23) \{ it is a watch (prejacent); it is a watch (salient) \}.

This seems to pose a problem for the first *not at issue* component of the definition of exclusives. Remember that this component requires all salient alternatives to be stronger than the prejacent, but as we see in (23), the salient alternative is identical to the prejacent, and consequently as strong as it. To satisfy this requirement we suggest to re-interpret the alternatives in such a way that the salient alternative will be ranked higher than the prejacent on an evaluative scale. This can be done by modifying *it’s a watch* with two different modifiers MOD$_1$ and MOD$_2$ as in (24):

(24) *It’s MOD$_1$ watch > it’s MOD$_2$ watch*

The values assigned to MOD vary with context. In the case of (22) we may get e.g. (25):

(25) *This is a designer’s watch > This is a standard watch.*

Notice that the existence of similar covert evaluative modifiers has been independently proposed in the literature for the interpretation of e.g. *exclamatives* (Rett 2008 [13]), and *The same* (Barker 2012 [1]). Along the same lines we can account also for the interpretation of *STAM VP* constructions, as in (8) above.

6 Parameter #3: Origin of the Scale: (Supplied by Context or by the Lexical Structure)

I observe that *rak* and *stam* operate only on scales that are supplied by context, but *be-sax ha-kol* can operate also on scales that originate from the lexical semantics of its associate. This accounts for two observations. First, it accounts for the unpredicted felicity of (17) with *be-sax ha-kol*. In particular, I propose that in (17) *be-sax ha-kol* can ignore the weaker alternative in the context because it has the option to operate on another salient scale. Crucially, the salient scale it seems to operate on is supplied by the lexical structure of its associate, and NOT by context. In such a case, *be-sax ha-kol* operates on a scale of alternatives that are automatically activated by a numeral expression (Spector 2013 [18]), as numbers naturally trigger a ‘Horn scale’. Assuming that the basic meaning of *n* is *at least n* (e.g. Horn 1976 [8]; van Rooij & Schulz 2006 [15]), in e.g. (17) *be-sax ha-kol* presupposes that *at least 20 people arrived*, and asserts that the stronger alternative *at least 21 people arrived* is false. The result is that *exactly 20 people arrived*. With *rak* this sentence is in felicitous because *rak* cannot operate on lexically driven scales. It must relate to the salient alternative in the context. The second observation that this parameter can explain is that we don’t get an approximative reading when accented *STAM* associates with gradable adjectives ((12) above), despite the fact that as discussed in Sect. 5,
it is compatible with internal alternatives. However, as STAM cannot operate on lexically driven scales, it cannot operate on the internal alternatives that originate from the lexical semantics of gradable adjectives, as does be-sax ha-kol.

7 Summary

In this paper I have suggested that the Hebrew exclusives rak, stam and be-sax ha-kol all have the semantics of scalar exclusive operators, as defined in (18) above. I have further claimed that the various differences between them are derived from the different specification of three parameters, as summarized in Table 1. Further research should investigate these parameters with more particles cross linguistically. Cf. Coppock & Beaver’s 2014 [4] work on English exclusives, as well as Tomaszewicz 2012 [19] analysis of Polish exclusive expressions.

Table 1. Specification of three varying parameters

<table>
<thead>
<tr>
<th>Parameter #1</th>
<th>Parameter #2</th>
<th>Parameter #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of scale</td>
<td>Type of alternatives</td>
<td>Origin of alternatives</td>
</tr>
<tr>
<td>rak</td>
<td>Entailment and evaluative</td>
<td>external</td>
</tr>
<tr>
<td>stam</td>
<td>Only evaluative</td>
<td>External and internal</td>
</tr>
<tr>
<td>be-sax ha-kol</td>
<td>Entailment and evaluative</td>
<td>External and internal</td>
</tr>
</tbody>
</table>

References

Changes in the Distributional Patterns of the Spanish Copulas: the Influence of Contextual Modulation

Sara Sánchez-Alonso

Yale University

sara.sanchez.alonso@yale.edu

Abstract. I investigate the influence of contextual information in the acceptability and online processing of Spanish copula sentences with adjectival predicates, such as *La carretera es/está ancha* “The road *ser/estar* wide”. The focus is on those adjectival predicates, typically considered individual-level, that are acceptable and preferred with *ser* without contextual information. I examine two main questions: (1) how and what kind of contextual modulation influences native speaker’s judgments of copula sentences and (2) whether lack of such contextual information will translate into online processing costs. Study 1 addresses the first question using a questionnaire with two different tasks: a rating and fill-in-the-blank tasks. The results show that native speakers have a preference for *estar* to appear with a supporting context, i.e. a context that provides a set of contrasting alternatives to the current discourse situation. Study 2 is a self-paced-reading study that provides further support for the type of contextual modulation required by *estar*. In particular, *estar* sentences preceded by a neutral context engender a higher cost overall than both baseline sentences with *ser* and *estar* sentences preceded by a supporting context. No difference in terms of processing costs were found for *ser* sentences regardless of context type. Overall, the data indicates that when *estar* sentences with individual-level predicates are preceded by the required contextual information, they are not harder to process than *ser* sentences. I consider how these results contribute to further our understanding of the diachronic changes in the distributional patterns of the Spanish copulas, as well as the copula uses observed in modern Spanish.

1 Introduction

Spanish has two copulas, *ser* and *estar*, that exhibit distinct distributional patterns: locative predicates typically appear with *estar*, while nominal predicates typically appear with *ser*. Adjectival predicates sometimes, but not always, may combine with either copula, and this gives rise to differing interpretations. Crucially, individual-level adjectival predicates such as (1-a) and (1-b) are typically unacceptable with *estar* without a supporting context and they will be the focus of the present study.
The results from the two studies will shed light on the distributional patterns of the two copulas in modern Spanish, and it will also provide some insights into their diachronic development.

2 The Diachronic Development of the Spanish Copulas

The domain of use of *estar* has been part of a continuous process of lexical diffusion that started in the 12th century ([14], [13]). It originated from the weakening of the Latin verb *stare* “to stand”, and since then it has been encroaching on the domain of use of *ser*. In Latin, the verb *stare* was most commonly used to indicate physical position and has often been translated as “stand, stand still, remain standing” as in (2-a) and (2-b). The copula *ser* originates from Latin *esser*, which has been translated as “be, exist, be there” and it was used in Latin to indicate both location (2-c) and existence (2-d) ([13]).

(2) a. *Sto ad ianuam.* “I am in front of the door”
   b. *Pugna stetit.* ‘The battle continued’.
   c. *Roma in Italia est.* “Rome is in Italy”
   d. *Erant omnino itinera duo.* ‘There were only two tracks’.

The verb *esser* appeared with INDIVIDUAL-LEVEL (3-a) and STAGE-LEVEL (3-b) copulative sentences. These two types of predicative sentences were originally defined by Carlson ([4]) as denoting two different types of entities. Briefly, stage-level predicates apply to a spatially and temporally bounded manifestation of something (3-a), as opposed to individual-level predicates, which apply to individuals as a whole, regardless of their temporal/spatial manifestations (3-b) ([13]).

(3) a. *Bene est.* ‘(He) is well’.
   b. *Ueri amici fideles sunt.* “Real friends are faithful”

In the first written records of Spanish from the 12th century, *ser* still serves a wide range of functions (for an overview: [5], [2], [3] i.a), and *estar* maintains the original meaning from Latin and thus appears regularly in locative sentences. At this stage of the language, *ser* is also found in locative constructions, and sometimes both copulas are used interchangeably in the same text as in (4-a) and (4-b) ([11]).

(4) a. *...al Criador que en el cielo está.* “...to the Lord that in the sky *estar*”
   b. *El señor que es en el cielo.* ‘The Lord that *ser* in Heaven’.
During this period, *estar* and *ser* begin to coexist in predicate copulative sentences that are stage-level as in (5-a) and (5-b) respectively. *Estar* will become the preferred form with this type of predicates in later centuries, approximately around the 16-17th centuries, as evidenced by a majority of combinations of *estar* with stage-level predications in comparison to *ser* ([11]).

(5)  
   a. ¿_et por qué _eres triste_ et demudado? “and why _ser_ you sad and upset?”
   b. ¿por qué _estas triste_? ‘Why _estar_ you sad?’

However, not all stage-level adjectives will show the same early acceptability with *estar*. Some adjectives only begin to appear with *estar* at a later period, around the 15-16th centuries. This is the case of adjectival predicates that have been described in the literature as “evidential” or “emotional” ([14], [8], [12]; i.a.). These are uses in which the property is attributed on the basis of an immediate sensorial experience as in (6) and (7) below ([1], [6]).

(6) ¡_Qué linda estas_! “How pretty you _estar_!”
(7) Los pollos _están_ muy ricos con tomate en este tiempo. “Chicken _estar_ delicious with tomato at this time of the year.”

The changes in the distributional patterns described so far are still present in modern Iberian Spanish. In the next section, I will focus on copula patterns in modern Spanish, and particularly on those adjectives that appear with both copulas.

2.1 Copula Distribution in Modern Iberian Spanish

In modern Spanish, some adjectival predicates may combine with both copulas, giving rise to two different interpretations. In (8), for example, *ser* combines with the adjective *guapo* “handsome” to indicate that handsomeness is an intrinsic property of Pedro. In contrast, the use of *estar* in (9) expresses that handsomeness is a property that applies to Pedro at a particular time, for instance at the time when he is wearing a nice suit.

**Context:** I have always thought that Pedro should work as a model.

(8) _Es guapo y tiene buen estilo._  
   ‘(He) _ser_ handsome and has very good clothing style’.

**Context:** Pedro is wearing a really nice suit today.

(9) _Está guapo y tiene buen estilo._  
   ‘(He) _estar_ handsome and has very good clothing style’.

109
The use of *estar* with adjectival predicates, however, does not always describe a property as temporary. In some cases, *estar* seems to express a change or to describe a contrasting property. For example, in (10-a), the property of being bald cannot be considered temporary. Instead, the intuition is that *estar* points to a change or establishes a contrast with a prior state of Guillermo, in which he was not bald. In (10-b), the apples are described as sour, which contrasts with an expected state of the apples in which they are ripened.

(10)  

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Guillermo está completamente calvo. “Guillermo <em>estar</em> completely bald.”</td>
<td>b. Estas manzanas están agrias. ‘These apple are sour.’</td>
</tr>
</tbody>
</table>

Some of the adjectives that appear with both copulas have been traditionally classified as individual-level. These are adjectives that, without context, show a strong preference for *ser*. However, when the appropriate contextual information is provided, individual-level adjectives are acceptable with *estar*. For example, the adjective “británico” typically appears with *ser*, but there are instances in which the use of *estar* is acceptable, as in (11-a).

**Context:** Juan approaches wearing a bowler hat and with a walking cane in his hand, and someone says to him ([7]):

(11)  

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ¡Vaya! ¡<em>Estás</em> muy británico! ‘Wow! You <em>estar</em> very British!’</td>
<td></td>
</tr>
</tbody>
</table>

This type of adjectives present a puzzle within the distributional patterns of the *estar* and they have been usually analyzed as a type of “coercion” ([7], [12]). In the next section, I will introduce an analysis that aims to understand copula alternation in modern Spanish, particularly in combination with adjectival predicates, by appealing to the pragmatics properties of *estar*.

### 3 A Discourse-Based Analysis on Copula Alternation

Several analyses have been proposed to understand the differences between *ser* and *estar*, but here I will focus on a particular account that locates the distinction in the pragmatic presuppositions associated with *estar* ([9], [10]). This analysis is one of the most formally explicit proposals of the phenomenon and it has been able to account for the majority of copula uses described in the previous section. On this analysis, the two copulas exhibit the same lexical semantics, the only difference is that *estar* carries the presupposition that the embedded proposition is restricted to a specific discourse situation, whereas *ser* remains neutral on this issue. Maienborn (2005) defines a specific discourse situation as “the relevant situation to which the speaker restricts his/her claim”.

A proposition will be restricted to a specific discourse situation if the context allows for alternative situations in which the proposition need not apply. Therefore, *estar* licensing contexts will be those in which the discourse situation described by *estar* can be contrasted against alternative discourse situations. For example, *estar* felicitously combines with the individual-level predicate *wide* in
(12-a) because it contrasts the discourse situation with alternative situations in which the property of being wide may not apply to distinct parts of the Panamericana highway.

Context: A journalist is reporting on the Panamericana and she is now near Lima.

(12) a. *La carretera está ancha.* ‘The road *estar* wide.’

On Maienborn’s analysis, the felicity of *estar* sentences with individual-level adjectival predicates crucially depends on the contextual conditions and the accessibility of contrasting alternatives.

4 Processing Predictions

On the basis of Maienborn’s analysis, the acceptability of *estar* sentences with individual-level adjectival predicates is expected to increase if the context supports a link to a particular discourse situation that contrasts to alternative situations. A context that is neutral with respect to the existence of contrasting alternative situations will decrease the acceptability of *estar* with individual-level adjectival predicates. The acceptability of *ser* should remain unchanged regardless of contextual modulation.

With respect to the online processing of *ser* and *estar*, the two main predictions are as follows. (1) *Estar* sentences preceded by a neutral context that does not provide a set of contrasting alternatives will lead to a higher processing cost than *ser* sentences. This cost is expected to arise right after retrieval of the adjective, which provides the necessary information to solve the presuppositional content of *estar*. In addition, (2) neutral contexts followed by *estar* sentences will be costlier to process than *estar* sentences preceded by a supporting context. The presuppositional information of *estar* should be naturally resolved and without cost when the necessary contextual information is provided.

These predictions were tested in two studies: **Study 1** explores whether native speaker’s judgments are influenced by contextual modulations, and **Study 2** investigates the online processing of the copulas using a self-paced reading paradigm.

5 Study 1: Questionnaire

Forty native speakers of Iberian Spanish (17 men and 23 women) were presented with pairs of sentences. Each pair consisted of a context sentence, followed by a test sentence with either *ser* or *estar*. Since the goal of the study was to explore the influence of the context on the acceptability of *estar* with individual-level sentences, only individual-level adjectives were included. Two contexts were distinguished: SUPPORTING CONTEXTS (EstarSupp), which facilitated reference to a particular discourse situation, and NEUTRAL CONTEXTS (Neutral), which remained neutral on the specificity of the situation.
Half of the sentences were presented in a **scoring task**, in which participants had to assign the sentence a number from 1-5 depending on whether the sentence could or could not be said by a native speaker of Spanish. The other half of the sentences were presented in a **fill-in-the-blank** task, in which participants had to complete the sentence with one of the two copulas. Examples of the sentence pairs are presented below.

1. **EstarSupp + ESTAR:**
   a. **Context:**
      
      *Pedro hizo dieta durante meses*  
      ‘Pedro went on a diet for six months.’
   b. **Sentence:**
      
      *Le vi ayer, está delgado.*  
      ‘I saw him yesterday, (he) **estar** skinny.’

2. **Neutral + ESTAR:**
   a. **Context:**
      
      *Por fin he conocido al marido de Lucía.*  
      ‘Finally, I have met Lucia’s husband.’
   b. **Sentence:**
      
      *Le vi ayer, está delgado.*  
      ‘I saw him yesterday, (he) **estar** skinny.’

3. **EstarSupp + SER:**
   a. **Context:**
      
      *Pedro hizo dieta durante meses*  
      ‘Pedro went on a diet for six months.’
   b. **Sentence:**
      
      *Le vi ayer, es delgado.*  
      ‘I saw him yesterday, (he) **ser** skinny.’

4. **Neutral + SER:**
   a. **Context:**
      
      *Por fin he conocido al marido de Lucía.*  
      ‘Finally, I have met Lucia’s husband.’
   b. **Sentence:**
      
      *Le vi ayer, es delgado.*  
      ‘I saw him yesterday, (he) **ser** skinny.’

### 5.1 Results

In the **scoring task**, acceptability scores were significantly higher for [EstarSupp+ESTAR] in comparison to [Neutral+ESTAR] (*p < .001*). In addition, [Neutral+SER] sentences were rated higher than [EstarSupp+SER] sentences (*p < .001*), indicating a preference for **estar** with supporting contexts.

In the **fill-in-the-blank task**, **estar** was the preferred choice with EstarSupp contexts in 65% of the sentences. **Ser** was the preferred choice with Neutral contexts in 80% of sentence pairs. These results, and particularly the higher percentage of **estar** with supporting contexts, indicate that the choice of copula is modulated by contextual information. It is particularly important to remember that all adjectives are individual-level and therefore should be acceptable and preferred with **ser** without contextual information.
6 Study 2: Self-Paced Reading Study

Forty native speakers of Iberian Spanish (19 men and 21 women) were presented with pairs of sentences following the design in Study 1. Sentences were presented word by word using the moving-window technique. Reading times were recorded for each displayed segment. There were four crucial segments, all of them in the test sentence that appeared with either ser or estar: the copula (COP), the adjective (ADJ), one word after the adjective (ADJ+1) and two words after the adjective (ADJ+2). After each sentence, a yes-no question or statement about the sentence was presented to ensure that participants had read and understood the sentence.

6.1 Results

A repeated-measures ANOVA revealed a significant effect of context type on reading times (RTs) right after the retrieval of the adjective, that is in the following two words (ADJ+1 and ADJ+2) (Figure 1). In comparison to the baseline, [Neutral+Estar] sentences engendered significantly higher RTs than [Neutral+Ser] sentences (ADJ+1: \(p = .039\)). This cost persists two words after the adjective (ADJ+2: \(p = .046\)). In addition, [Neutral+Estar] sentences led to a higher cost in terms of reading times than [EstarSupp+Estar] (ADJ+2: \(p = .004\)). The results are summarized in Table 1.

![Fig. 1: Mean Reading Times for each Main Segment by Sentence Type (n=40)](image-url)
7 Discussion

The results of the questionnaire indicate that speakers have a preference for *estar* to appear with a supporting context, i.e. a context that provides an accessible set of alternative situations that contrast with the particular discourse situation that the speaker has in mind. In contrast, *ser* sentences received higher ratings and were chosen more frequently with neutral contexts. It is important to bear in mind that all the adjectives included in the study show a preference for *ser* when presented without contextual information. Although *ser* is expected to be a neutral copula, and therefore the prediction is that it will be acceptable with both types of contexts (supporting and neutral), the results from the fill-in-the-blank task indicate that *ser* is mostly preferred with neutral contexts. I explain this to be a result of the pragmatic division of labor (Horn 1984), in which *ser* functions as a more general copula. In the fill-in-the-blank task, participants were asked to choose between the two copulas. If the speaker chooses *ser*, the inference is that, due to pragmatic economy principles, s/he does not want to restrict the proposition to a specific discourse situation, otherwise, s/he would have chosen *estar*. As a result of this inference, *ser* is usually reserved for those copula uses in which the proposition is not restricted to a particular situation.

These findings are also supported by the real-time data: [Neutral + Estar] sentences engender a higher cost than both (1) baseline sentences with *ser* and (2) *estar* sentences preceded by a supporting context. These costs are first observed right after the retrieval of the adjective, which is the point at which the participant has processed all the information required to resolve *estar*’s presuppositional content.

No differences in reading times were found between *estar* sentences with individual-level adjectives that were preceded by a supporting context and any of the sentences with *ser*. From this data, we can conclude that when *estar* predications with individual-level adjectives are provided with the necessary contextual information, they are not harder to process in comparison to *ser* sentences. The appropriate context for a *estar* sentence is one that provides a set of contrasting alternatives that allows the restriction of the proposition to a particular discourse situation.

As expected, no difference in terms of reading times were found for *ser* sentences regardless of the context. This result supports the neutral role of *ser* with respect to the restriction of the proposition to a discourse situation.

Altogether, the data provides evidence for a pragmatic analysis of *ser* and *estar* in which both copulas have the same lexical semantics, and differ only in

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Segment: $p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Neutral + Estar] &gt; [Neutral + Ser]</td>
<td>ADJ+1: $p = .039$</td>
</tr>
<tr>
<td>[Neutral + Estar] &gt; [Neutral + Ser]</td>
<td>ADJ+2: $p = .046$</td>
</tr>
<tr>
<td>[Neutral + Estar] &gt; [EstarSup+Estar]</td>
<td>ADJ+2: $p = .004$</td>
</tr>
</tbody>
</table>

Table 1: Self-Paced Reading Results by Contrast (n = 40).
that estar presupposes a relation to a specific discourse situation. Estar predications will be restricted to a particular discourse situation if there are alternatives in which the proposition need not apply. This account of copula alternation explains the felicitous use of estar with individual-level adjectives. It is one of the most parsimonious accounts in that it relies on minimal differences between the two copulas and predicts similarities with copulas in other languages, such as English be and German sein.

7.1 Implications for Copula Variation

The results from the two studies presented above shed light on the diachronic development of the distributional patterns of the copulas and the copula uses in modern Spanish. The results support the process of lexical diffusion observed in the domain of use of estar and its encroaching onto the domain of ser.

In combination with adjectival predicates, Maienborn’s ([9], [10]) analysis predicts that a felicitous use of estar requires an accessible set of alternatives that contrast with the relevant discourse situation that the speaker has in mind. This set of contrasting situations is naturally provided by stage-level predicates, which explains why they were the first type of adjectival predicates to combine with estar. For example, in (13), the speaker knows that “sadness” is typically a property that applies at different points in the life of an individual, and thus it naturally contrasts with other points in the life of “the boy”.

\[
\text{(13)} \quad \text{El niño está triste.} \quad \text{‘The boy estar sad.’}
\]

This use of estar with stage-level predicates was already attested in the 12th century. The second type of stage-level predications, the so-called “evidential” or “emotional” uses of estar such as (14), started to appear in written texts much later (around the 15th century).

\[
\text{(14)} \quad \text{La paella está riquísima.} \quad \text{‘The paella estar delicious.’}
\]

These copula uses are usually instances in which the property is attributed on the basis of the speaker’s expectations about the referent. The contrast is therefore built on a set of alternatives based on the personal experience of the speaker. A main difference between these two types of stage-level adjectives is how the set of contrasting alternatives is built. Further studies are needed, however, to investigate how these two ways of creating sets of contrasting alternatives may have influenced the acceptability of estar with stage-level predications.

In modern Spanish, we observe a change in the type of adjectives that combine estar. In particular, I have shown that typical individual-level adjectives are acceptable with estar when preceded by a context that provides contrasting alternatives to the specific discourse situation. This use of estar can be as natural to process as any ser sentence with individual-level adjectives.
References

A Formal Semantics of the Final Rise*

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

Abstract. This paper presents a formally precise model of how the final rise affects the discourse structure of a dialogue. Our account makes precise the informal claims from previous discussions which have characterised the final rise as signalling ‘incompleteness,’ ‘uncertainty’ or ‘insufficiency’ in various senses (e.g., Pierrehumbert and Hirschberg [11], Hobbs [8], Bolinger [3], Westera [19]). Inspired by their analyses, we give a formal semantics in the SDRT framework (see [1]) that models ‘incompleteness’ as an underspecified notion that is resolved to specific interpretations in context.

1 Introduction

In this paper, we give a formal semantics of the final rise in spoken English dialogue, as part of our ongoing work on pitch contours. Our general claim is that the pitch contour of an utterance perturbs the standard inferences from surface form to illocutionary force and perlocutionary effect. Hence, by computing these perturbations, we can derive the implicatures that a final rise is usually taken to convey. The following examples (adapted from [15]) are cases in point:

(1) A: You’re a millionaire.
a. B: I’m a \textsc{millionaire}, \textsc{H* LL%} ‘Yes, I am.’
b. B: I’m a \textsc{millionaire}, \textsc{H* LH%} ‘Really?’

(2) A: Are you rich?
a. B: I’m a \textsc{millionaire}, \textsc{H* LL%} ‘In particular, I’m rich.’
b. B: I’m a \textsc{millionaire}, \textsc{H* LH%} ‘Does that count?’

The utterance in (1a) and (2a) is intonated in the standard ‘high focus, final fall’ contour (H* LL%), leading to the usual effects of indicatives: It commits B to the proposition ‘B is a millionaire,’ and therefore in (1a) establishes a shared public commitment (i.e., agreement) on that proposition, and completes

* The research presented in this paper has been funded by the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 567652 ESSENCE: Evolution of Shared Semantics in Computational Environments (http://www.essence-network.com/).

1 To describe our examples, we use the ToBI annotation scheme for pitch contour (see [14]). An asterisk (*) marks the focus accent, and a final rise is denoted by LH%.
a question-answer pair in (2a). In contrast, the same utterance with a final rise (LH%) in (1b) does not make any such commitment (cf. [6]). Therefore, there is no established agreement between A and B in (1b) and B’s utterance functions as a clarification request. Then again, in (2b), B does make the commitment to ‘B is a millionaire,’ but does not commit to this necessarily answering A’s question, i.e., the illocutionary force of an answer does not immediately obtain, and it is left open whether it does (pending A’s response).

Based on these and similar examples, a body of work has described the final rise in English as marking an utterance as ‘incomplete,’ broadly construed. In particular, Pierrehumbert and Hirschberg in [11] gloss this as ‘to interpret an utterance with particular attention to subsequent utterances,’ which Hobbs in [8] plainly calls ‘open.’ Westera in [19] characterises a final rise as indicating the failure of a Gricean maxim, which corresponds to our intuition that standard perlocutionary inferences do not obtain. Our goal in this paper is to give a fully formal model of these effects.

We proceed as follows. In the next section, we expand our informal discussion above to additional examples. Along the way, we further discuss earlier treatments of their effects and start motivating the choices we made for our formal account. We motivate our choice of SDRT as our formal framework and give a brief introduction into that formalism in Section 3. In Section 4, we present our formal theory in and show that its results correspond to our discussion from Section 2. We conclude with some pointers towards further work in Section 5.

2 Informal Discussion

In this paper, we are primarily interested in the effects of a final rise (LH%) in colloquial English and ignore the effects of nuclear pitch accents (e.g., the high focus accent H*). We consider ‘high focus, final fall’ (H* LL%) to be the default contour (for indicatives) and contend that it has conversational significance only in non-standard situations. The data and analysis presented here are not new in themselves. The following discussion merely rehearses previous accounts to motivate our formal model.

In (2b), the final rise indicates that the utterance is not necessarily an answer to the preceding question. Such utterances allow the dialogue to proceed in two different ways: either the addressee confirms (or rejects) that they take up the answer, or the speaker continues to supply more information. In either case, it can be said that the final rise utterance itself is not a complete answer. Hence, we follow the literature in characterising the final rise as signalling ‘incompleteness,’ but we consider incompleteness itself to be an underspecified property. Put simply, an incomplete utterance needs to be addressed—the incompleteness needs to be resolved—in the dialogue, but it is left open how. Which resolutions are possible depends on the context, and the speakers negotiate online which possible resolutions are adequate. The following two possible continuations of B’s utterance (adapted from [7]) exemplify the effect:
In both examples, B’s first utterance can be considered incomplete in the sense outlined above. In (3), B is expressing by way of a final rise that ‘Skokie’ is not the full answer, but that it is ‘Skokie, Illinois.’ In (4), B isn’t sure whether the answer ‘Skokie’ is sufficient, and hence leaves the answer open for confirmation or further questions. As analysed by Hirschberg and Ward in [7], the incompleteness of the final rise utterance in (3) and (4) lies in B being unsure whether A can resolve the proper noun ‘Skokie’ in a meaningful way. Note that B is inviting a response from A in (4): an appropriate gloss of the incompleteness in B’s utterance is ‘does this answer your question?’ However, while this means that the utterance has question force, in some sense, a reading as a polar question (‘am I from Skokie?’) is inappropriate.

In contrast, there is an apparent tendency to infer polar clarification requests (requests for confirmation or elaboration) from a final rise intonation, as in (1b) where the gloss ‘am I a millionaire?’ is appropriate. So in some contexts, it is licensed to interpret a final rise utterance in indicative mood with propositional content \( p \) as having the conversational effect of the question ‘Is it the case that \( p \)?’ As mentioned already, part of our solution is to leave the incompleteness of the final rise underspecified. We merely stipulate that it projects a follow-up, but leave open what speech act precisely is being projected. First of all, this neatly captures the use of a final rise in ‘list intonations’ as in the following example: (cited from [19]):

(5) A: What did you do today?
   B: I sat in on a history class.
   B: I learned about housing prices.
   B: And I watched a cool documentary.

In Example (5) and similar ones, a speaker simply provides the follow-up projected by the final rise intonation themself, and we can gloss the contribution of the final rise as ‘I am not done.’ If this is not the case, the onus to provide the projected follow-up is on the addressee of the final rise utterance. We assume that, by default, this projection corresponds to a polar clarification question wherever the context allows this, i.e., that the speech act projected is an answer to a polar question. But we also explicitly leave room for other possible continuations, e.g., elaborations (as in 3) or acknowledgements (as in 4). This defaulting to polar questions can be abused for comedic effect, as in the following example:\(^2\)

The example is taken from the TV show The Big Bang Theory, but anyone who has participated in an oral exam should be familiar with utterances like B’s.
B's utterance in (6b) taken on its own admits different interpretations: it might be taken to express the polar question 'Is it Radon?' or that B is unsure whether the answer is adequate w.r.t. A's intentions. The speaker A is apparently (or purposefully pretending to be) unsure of the right interpretation. This further shows the inherent underspecification in interpreting a final rise.

Aside from this influence on the discourse structure, there are also accounts that characterise 'incompleteness' as displaying an 'uncertain attitude' of the speaker (cf. [18], [7]). Westera in [19] has given a compelling account of such uncertainty in terms of the Gricean maxims: by intonating an utterance with a final rise, a speaker announces that they cannot vouch for the truthfulness (Quality), sufficiency (Quantity) or appropriateness (Relation) of their utterance. Since we do not formalise the Gricean maxims precisely, we take a broader standpoint: in our formal account, the speaker is uncertain regarding the successful uptake of their utterance. For our formalisation, we choose a model that keeps track of discourse structure and cognitive attitudes separately. Hence, our model will describe 'incompleteness' according to both accounts.

The final rise also prominently features in utterances that are correcting or contradicting the previous speaker: the rise-fall-rise contour has been dubbed contradiction contour by Liberman and Sag in [10]. This fits into our discussion so far as follows: After a correction, the discourse is in a 'state of crisis' (cf. [5]) that needs to be addressed. In that sense, a correction prompts a follow-up, aligning with our characterisation of the final rise. Furthermore, a correcting speaker is almost by necessity uncertain whether their contribution will be taken up cooperatively.

3 Framework

Our theory of pitch contours is implemented in Dialogue Segmented Discourse Representation Theory (DSDRT) (established in [1] and [9]). Our rationale goes as follows. SDRT models back-and-forth information flow between a logic of information content (a dynamic semantics) and a logic for cognitive modelling (modelling beliefs and intentions) by way of the so-called glue logic that connects the different sources of information. This allows us to model perturbations of the information flow: by stipulating new glue axioms for the final rise, we can block defeasible standard inferences, and impose novel restrictions on the semantic content. In addition, the glue logic allows for defeasible reasoning, and models underspecified information. We now give a brief overview over the most important concepts.
The SDRT language of information content is used to express the (truth-conditional) logical form of a discourse. It consists of the following components:

- A standard first-order alphabet, plus a modal operator (\(\square\)), and operators for imperatives (!) and interrogatives (?)
- A set of label variables. We conventionally denote these by Greek letters, \(\alpha, \beta, \lambda, \pi_1, \pi_2, \ldots\)
- A finite set of predefined discourse relations, e.g., Explanation, Elaboration, Correction.

The well-formed formulae are obtained by the standard syntax on the alphabet without discourse relations (treating labels as variable symbols), then adding formulae of the form \(R(\pi_1, \ldots, \pi_n)\) where \(R\) is an \(n\)-ary discourse relation and \(\pi_i\) are labels, and then closing under booleans and quantification (of first order variables, label variables, and discourse relations).

These formulae are given a dynamic semantics that also associates pre-defined truth-conditions with every discourse relation. For instance, the truth of \(\text{Elaboration}(\pi_1, \pi_2)\) must also render \(\pi_1\) and \(\pi_2\) true.

A logical form in SDRT is a segmented discourse representation structure (henceforth, SDRS). An SDRS consists of a set of labels \(\Pi\) and a function \(F\) mapping labels from \(\Pi\) to formulae in the logic of information content (sometimes we write \(K_{\pi}\) for \(F(\pi)\)). Intuitively, the labels of an SDRS mark discourse segments that can be connected by discourse relations. These relations are also represented in the language of information content. Hence, we can use \(F\) to define a order on \(\Pi\): \(\pi_1 \triangleright \pi_2\) iff either \(R(\lambda, \pi_2, \lambda)\) or \(R(\pi_2, \lambda)\) appear in \(F(\pi_1)\) for some \(R\) and \(\lambda\). We write \(\pi_1 \trianglerighteq \pi_2\) for the reflexive and transitive closure of \(\triangleright\) and read this as \(\pi_1\) outscopes \(\pi_2\). Intuitively, \(\pi_1\) outscoping \(\pi_2\) means that \(\pi_2\) is a subsegment in the larger discourse segment labelled by \(\pi_1\). On well-formed SDRSs, this order is additionally required to be anti-symmetric, i.e., \(\trianglerighteq\) is a partial order.

Outscoping, together with the discourse relations, is used to reason about the structure of a discourse and not just its contents. We allow the outscoping relation \(\trianglerighteq\) (with the above truth-conditions) to be used in the logic of information content. DSDRT extends this model to dialogues where the information content of the two interlocutors might differ. In a dialogue, every interlocutor vouches for the truth of certain logical forms, i.e., a speaker makes commitments to SDRSs. The logical form of a dialogue turn is a set of SDRSs (one for each interlocutor), and the logical form of a dialogue is the sequence of the logical forms of its

---

3 See [1, Appendix D] for a list.

4 Such quantification is a slight deviation from the presentation in [9] where only first-order variables are quantified. The change is made here to model a public commitment to a yet unknown discourse relation in a transparent way: by existential quantification. This is a conservative extension. Since the logic of information content includes an event calculus that allows us to quantify over speech events, this changes neither the proof-theoretic properties of the logic nor its expressiveness. The equivalent construction in standard SDRT would use the '?' variables of the glue logic to express unknown relations and labels.

5 Again, a slight deviation from the norm, but also with no further consequences.
turns. Thereby, each interlocutor’s commitments over the course of a dialogue are recorded individually. The following example (from [2]) is a simple DSDRS:

\[
A: \text{Max fell.} \quad B: \text{John pushed him.}
\]

We gloss the assignment function \(F\) in each individual SDRS by a colon (so \(F(\pi) = K\) is \(\pi : K\)). In Example (7) above, \(\pi_1\) labels the information content of ‘Max fell’ and \(\pi_2\) that of ‘John pushed Max’ and the speakers are committed to their individual utterances. However, by general principles of dialogue coherence, we also infer that B meant to explain the event described by A, which is recorded by the contents of \(\lambda\). The dynamic semantics of \(\text{Explanation}(\pi_1, \pi_2)\) entail the contents of \(\pi_1\) and \(\pi_2\) and that \(\pi_2\) answers ‘why \(\pi_1\)?’ Hence we can infer from the SDRS in (7, turn 2) that B is also committed to the proposition that ‘Max fell.’ This means that we can read off the logical form in (7) that A and B agree on that proposition, despite this not being linguistically explicit in B’s utterance.

The inferences used to construct logical form—particularly which discourse relations connect which labels—are not drawn in the logic of information content, but in the glue logic. Such inferences are by their very nature defeasible, as novel information could change the interpretation of the dialogue at any time. Hence, the glue logic has a defeasible conditional \(>\) which we read as ‘normally.’ To be precise, from \(\varphi\) and \(\varphi > \psi\) we infer \(\psi\) only if that conclusion is not blocked by other information in the current context.\(^6\) In addition, glue logic formulae can use a special variable ‘?’ to denote elements of the logic of information content that are not yet fully specified, but whose structural properties can already be circumscribed. The derivation of discourse relations is facilitated by stipulating axioms in the glue logic of the following form:

\[
(\lambda : ?(\alpha, \beta) \land \text{Info}(\alpha, \beta)) > \lambda : R(\alpha, \beta)
\]

This schema reads as: “if \(\alpha\) and \(\beta\) are rhetorically connected somehow to form a part of the (extended) discourse segment \(\lambda\), and their contents satisfy \(\text{Info}\), then normally, they are connected by \(R\).” These axioms are used to construct the form (structure) of a logical form. The following concrete axiom is used to derive the SDRS in (7, turn 2) and stipulates that if it is known that \(\beta\) attaches to \(\alpha\) (by a yet unknown discourse relation), and there is evidence that \(\beta\) can cause \(\alpha\), we (defeasibly) infer that \(\beta\) attaches as an explanation to \(\alpha\).

**Explanatory Axiom.**

\[
(\lambda : ?(\alpha, \beta) \land \text{cause}_D(\beta, \alpha)) > \lambda : \text{Explanation}(\alpha, \beta).
\]

Finally, the separate cognitive modelling logic is used to model the speakers’ cognitive states. It includes a number of modal operators (see [2] for details):

\(^6\) This leads the glue logic to admit a nonmonotonic proof theory which is detailed in [1, Chapter 5].
KD45 modal operators for beliefs: \( B_S \) for a speaker \( S \).
- K45 modal operators for public commitments: \( P_S \) for a speaker \( S \).
- Modal operators for intentions: \( I_S \) for a speaker \( S \).

Glossing over some details, we write \( I_A P_B \varphi \) if \( A \) wants \( B \) to commit to \( \varphi \) and \( I_A B_B \varphi \) if \( A \) wants \( B \) to believe that \( \varphi \) holds. The cognitive modelling logic interfaces with the glue logic: facts from the cognitive modelling logic can block inferences using the defeasible conditional \( > \) in the glue logic. For instance, if \( P_S \neg \varphi \) in the cognitive model, then the glue logic cannot infer a discourse relation in \( S \)'s SDRS that would entail \( \varphi \). Conversely, information from information content can be used to infer cognitive states, e.g., if \( A \) has asserted that \( p \), then the glue logic infers that \( B_A p \) — but this inference too can be defeased if it is known that \( A \) is being insincere. For ease of notation, we will also use functions \( S(\pi) \) and \( H(\pi) \) mapping a label to its speaker and hearer, respectively.

4 A Formal Model of the Final Rise

Based on our informal discussion in Section 2, we claim that the final rise has an influence both on the structure of the dialogue (incompleteness) and on the displayed attitudes of its speaker (uncertainty). To formalise incompleteness, we assign the following semantics to the final rise in the logic of information content:

**Semantics of the Final Rise.**

\[
LH\%(\pi) \rightarrow \exists R, \pi', \pi'' (R(\pi', \pi'') \land \pi' \geq \pi).
\]

That is, the final rise does not affect the information content of the utterance itself, but it enforces that there is a yet unknown follow-up response that stands in some yet unknown discourse relation to the current dialogue. We also leave it underspecified what label this relation attaches to on the left side, only that the current utterance (labelled \( \pi \)) is in its scope. This underspecification is necessary because the continuation can attach both to the current utterance itself or to a wider discourse relation, when it is that relation itself that is uncertain. For example, in (3) the follow-up attaches directly as an elaboration, but in (4) the follow-up attaches to the whole question-answer pair consisting of the first two utterances; see Figure 1 below. We consider this model to be a rather faithful formalisation of the informal discussion in [11].

These semantics account for the forward projecting function of the final rise, e.g., in list intonations as (5). However, we also observed a backward looking property: asking for clarification, thereby indicating uncertainty regarding an earlier utterance. We model this effect by stipulating an appropriate glue logic axiom. The following rule, if it applies, turns a proposition \( p \) in indicative mood into the polar question \( ?p \) as in Example (1b). For the sake of exposition, we shorthand the above semantics of the final rise as ‘\( \pi : LH\% \)’ (read as: the label \( \pi \) includes the final rise semantics). We also use a function \( \text{prop} \) that maps a polar question \( ?p \) to its propositional content \( p \).
(Ax1) Clarification from Final Rise.
\[(\pi : LH\% \land \lambda : \square(\alpha, \pi) \land \alpha \rightarrow prop(\lambda)) \land \lambda : CR(\alpha, \pi)).\]

This axiom stipulates that if an utterance has a final rise, and it directly attaches to a previous utterance in some way, then it has the force of a clarifying polar question if this is consistent and the polar question is truth-conditionally appropriate. The appropriateness constraint \(K_\alpha \rightarrow prop(\lambda)\) is required to explain the incoherence of (8b):

(8) A: You are rich.
    a. B: I’m rich? ‘Am I?’
    b. B: # I’m a millionaire? ‘Am I?’

(9) A: You are a millionaire.
    a. B: I’m rich? ‘Am I?’
    b. B: I’m a millionaire? ‘Am I?’

Both (9a) and (9b) are licensed because, conventionally, ‘millionaire’ implies ‘rich,’ hence the question in (9a) is reasonable. Conversely, ‘rich’ does not necessarily imply ‘millionaire,’ so B’s utterance in (8b) cannot be taken to (necessarily) ask for clarification of A’s assertion. The only other permissible interpretation of (8b)’s surface form would be one indicating assent, but such interpretations are blocked by (Ax2) below. It is noteworthy that (Ax1) only applies to questions that ask for clarification of an earlier event. Specifically, we require the presence of an appropriate antecedent. We do not stipulate a rule that would infer question force from a final rise in general, as such utterances are incoherent when spoken ‘out of the blue,’ i.e., without an antecedent (see [6, p. 85]):

(10) a. A: Did you go to the cinema last night? ‘Am I?’
    b. A: # You went to the cinema last night? ‘Am I?’

The utterance (10b) is incoherent without an antecedent despite the corresponding interrogative mood utterance (10a) being appropriate in the same context. An ancillary conclusion to draw from (10) is that the antecedents \(\alpha\) we require for (Ax1) to apply need not be linguistic in nature (cf. [17]) as, e.g., an openly visible cinema ticket would render (10b) coherent. This is not to say that utterances with a final rise cannot be uttered out of the blue in general—they can, but then they cannot be interpreted as polar questions.

(11) A: (to a receptionist) My name is Mark Liberman.
    # ‘Is my name Mark Liberman?’

\[7\] We assign the relation CR the dynamic semantics of elaborating questions (Q-Elab), see [1, p. 468]. We gloss over more detailed properties of clarification questions (cf. [12]), as we are here only interested in the rather simple subset of polar questions.

\[8\] It is necessary to map \(K_\beta\) to its propositional content, as once question force is inferred, the logic of information content will represent \(K_\beta\) in question semantics.
While (11) *can* be uttered out of the blue, it should not be assigned question semantics. Nevertheless, the utterance expresses a request for a response/follow-up and this expectation of an adjacent action is adequately captured in our semantic postulate for the final rise.\(^9\) Lastly, we also describe the attitude displayed by a final rise, uncertainty, in the cognitive modelling logic:

\[(Ax2) \text{ Cognitive Contribution of the Final Rise.} \]
\[
(\pi : LH\% \land \lambda : R(\alpha, \pi) \land \neg \pi : ?K_{\pi}) \supset P_{S(\pi)} B_{S(\pi)} I_{H(\pi)} P_{H(\pi)} R(\alpha, \pi).
\]

This stipulates that if the utterance with the final rise is not a question,\(^10\) and is presumed to directly attach to an antecedent by some relation, then the speaker is usually conveying that they are uncertain whether the hearer is (or should be) willing to commit to that relation. This allows us to account for the uncertain answers (2b) and (4) where the speaker is uncertain whether the hearer is willing to commit to the Question-Answer-Pair (QAP) relation. In addition, this applies to utterances correcting or rejecting a previous utterance: typically, the corrected speaker cannot be assumed to accept the correction immediately. In other words, (Ax2) states that the speaker of the final rise utterance displays that they do not assume that normal cooperativity assumptions obtain—because those would usually lead to the addressee taking up the utterance. This is in alignment with Westera’s model in [19], but notably different from accounts that put ‘epistemic uncertainty’ in the spotlight (e.g., [18]). The latter discussions use epistemic uncertainty to account for the lack of commitment to a declarative proposition intonated with a final rise. We achieve this effect instead through (Ax1), as the dynamic semantics of the CR relation do not result in any commitments.

Note that (Ax1) and (Ax2) cannot apply simultaneously: (Ax1) infers question force, and (Ax2) explicitly does not apply if this is the case. This property of our formalism allows us to separate the implicatures we associated with our initial minimal pair (1b) vs. (2b). The final rise utterance in (1b) satisfies the antecedents of (Ax1) and is thereby rendered in question semantics. Consequently we do not infer a commitment of its speaker to ‘B is a millionaire.’ In contrast, the same utterance in (2b) does not satisfy the (context-sensitive) appropriateness constraint of (Ax1), and hence (Ax2) applies. Therefore, the speaker does make the usual commitments associated with a declarative utterance, but explicitly displays that they are not sure if the discourse relation of question-answer applies, \emph{i.e.}, if the addressee is willing to take up the utterance as an answer.

As a more verbose application, we present the DSDRT logical forms of (3) and (4) in Figure 1; the final SDRSs contain superfluous conditions that we left in for the sake of clarity. In both cases, A’s initial question is resolved by question semantics, and hence projects an answer. Then, in B’s first utterance, the final rise semantics stipulate that B’s turn is in some sense incomplete, but the available information is not sufficient to make that incompleteness precise. Due

\(^9\) Again, the response might also be a non-linguistic action such as looking up a reservation.

\(^10\) On questions, even when posed in indicative mood, a final rise is part of the default contour and we do not take it to signal a particular attitude in that situation.
to the appropriateness constraint, (Ax1) does not apply. The cognitive con-tribution (Ax2) prevents the answer ‘Skokie’ to attach to the question answer pair at this point. In the third turn, the logical forms diverge: the incompleteness is resolved to a missing elaboration in (3), and to an acceptance of the answer in (4). The cognitive contribution does not prevent these resolutions: in (3), the full answer is under another label, and in (4), the speaker A’s commitments are not subject to the cognitive restriction. In the next turn of (4), B will defease the attitude displayed through (Ax2), as A has now publicly displayed that the uncertainty is resolved.

5 Conclusion

We have presented a formal model that accounts for illocutionary and perlocutionary effects of the final rise in English, modelling in particular the rendering of indicatives to questions, and uncertainty when answering questions. To account for the variety of observable effects, our model postulates strongly underspecified semantics. In connection with SDRT’s glue logic, these underspecifications can be resolved in context to make concrete claims that are strong enough to predict incoherence. The particular novel contribution of our model lies in its formality, as we give a fully formal model of previously informal characterisations.

A notable shortcoming of our model, as it is presented here, is that we do not take the focus accent into account. The final rise is a part of some complex pitch contours that have received substantial attention in the literature, e.g.,
rise-fall-rise in [7] and [4] or, more generally, the fully compositional system of [15, 16]. However, we believe that the underspecified semantics presented in this paper are sufficiently broad to be consistent with these observations and that our formal model can therefore be expanded to cover more specific interpretations of complex pitch contours. This is part of our ongoing work and we address some of these concerns in [13].

References

Latin Long-Distance Reflexives in Messenger Reports

Per Erik Solberg
Department of Philosophy, Classics, History of Arts and Ideas
University of Oslo

Abstract. This paper presents the results of a study of Latin long-distance reflexives (LDRs) in so-called messenger reports, the reported words of a messenger speaking on behalf of someone else. Such LDRs often refer to the sender of the messengers, and this has been difficult to account for on a compositional logophoricity analyses of Latin LDRs. Based on this study, I argue that the antecedence options of LDRs are more restricted that what has been previously assumed, and I sketch a logophoricity analysis where sender-oriented LDRs are explained in terms of plural semantics and utterance predicate accommodation.

1 Introduction

The personal reflexive pronoun se is frequently used as a long-distance reflexive (LDR) in speech, thought, perception or emotion reports in Classical Latin: Rather than having as antecedent its local subject, it is bound by a nominal constituent in a superordinate clause or a previous clause [6,10,20]. There have been a few attempts at accounting for Latin LDRs as logophoric pronouns [10,20]. Logophors are pronouns restricted to reports, which refer to the person whose words, thoughts, perception or emotion is reported. A few languages have dedicated pronouns with this function, while other languages use the reflexive pronoun (see e.g. [3,4,19]).

The attempts at analyzing Latin LDRs as logophors haven’t succeeded in explaining the antecedents of Latin LDRs in what we might call messenger reports, reports of the speech delivered by a messenger on behalf of someone else: In such cases, it seems that LDRs have the ability to refer to the sender of the messenger rather than the messenger himself, even when the messenger is the subject of the report predicate. An antecedence pattern like that is challenging, not only for a logophoricity analysis, but probably for any compositional theory of the phenomenon.

This paper presents results from a corpus study of LDRs in Latin messenger reports, which suggests that LDRs in messenger reports are less problematic

* I wish to thank Dag Haug, Marius Johndal, Hazel Pearson and my colleagues in the Synsem project for fruitful discussions on this topic. Also, a big thanks to the very constructive ESSLLI reviewers. For correspondence, please use p.e.solberg@ifikk.uio.no.
than previously assumed. I furthermore sketch a semantic logophoricity analysis of Latin LDRs. A group of the examples with seemingly deviant antecedents in messenger reports can be accounted for as plural logophors. The antecedents in the remaining examples can be explained by mechanisms needed in the grammar for independent reasons.

The paper is organized as follows: Section 2 presents the basics of LDRs in Latin and Sect. 3 reviews the previous research on antecedents to Latin LDRs in messenger reports; in Sect. 4, I present relevant findings from my corpus study, which serve as basis for the outline of an analysis presented in Sect. 5; Sect. 6 concludes the paper.

2 Latin LDRs, the Basics

Complement clauses to report predicates in Latin are either subjunctive clauses or so-called *accusativus-cum-infinitivo* clauses (AcIs), that is, clauses with an infinitive verb and an accusative subject¹. Adverbial clauses and relative clauses in reports will be subjunctive-marked, provided that they are interpreted as being part of what is reported. LDRs are very frequent in report complements, both AcI and subjunctive, cf. (1) a and b.

(1) a. *[De numero eorum omnia se habere explorata] Remi dicebant.*

   ‘The Remi said that they had knowledge of everything concerning their number [the number of members of another tribe].’ (Caes. B.G. 2.4.4)

b. *[Ubii ... magnopere orabant /ut sibi auxilium ferret].*

   ‘The Ubii entreated with insistence that he should bring them help.’ (Caes. B.G. 4.16.5)

In addition to complement reports, Latin also has what could be called unembedded reports [1], reports stretching over multiple sentences, which are common e.g. when a speech is reported. These also alternate between subjunctive clauses and AcIs, and readily contain LDRs, cf. (2). Unembedded reports can be introduced by a report predicate, but this is not obligatory, as we will see.

¹ Syntactically, AcIs are more similar to finite clauses such as subjunctive clauses than e.g. control infinitives [10, Chap. 3], and it is therefore reasonable to consider reflexives with non-local antecedents within them as LDRs.
Ariovistus, ad postulata Caesaris paucā respondit,

de suis virtūtibus multa praedicavit: ... [16 sentences] ... about his virtue many. proclaimed.

'Àriovistus responded to Caesar’s demands in few words, and proclaimed his own virtue in many: ... What did he [Caesar] want from him?' (Caes. B.G. 1.44.1:8)

LDRs also occur in relative clauses and adverbial clauses embedded within reports, provided that they are subjunctive-marked [20].

LDRs refer to the individual whom we could characterize pretheoretically as the author of the report, the person whose words, thoughts, perception or emotion the report expresses. This will often, but not always be the nominative subject of an active report predicate [20, p. 22-29].

LDRs in messenger reports are not straightforwardly captured by this generalization, however. In (3a), an unembedded report is introduced by a speech predicate which has the messenger, Divico, as its subject. But LDRs in the report refer to the Helvetii, the tribe which sends Divico, not to Divico. In (3b), Caesar sends messengers. The messengers are the subject of the speech predicate, but Caesar is the antecedent of the LDRs in the complement clause and the subjunctive relative clause embedded within it.

'The Helvetians sent messengers to him. The leader of this embassy was Divico. He talked with Caesar in these terms: ... They had learned from their fathers and their elders that they should rather contend with valour than rely on trickery and plots.' (Caes. Gal. 1.13.2-4)
b. *Ad quos cum Caesar i nuntios (j) misisset,*
to whom when Caesar.NOM messengers.ACC send.PST.PRF.SBJV
*qui postularent eos /qui sibi *... bellum
who ask.PST.SBJV.PL them.ACC who REF.L.DAT war.ACC
*intulissent/ sibi dederent*,
inflict.PST.PRF.SBJV REF.L.DAT surrender.PST.SBJV answered

'When Caesar sent messengers to them, who were to ask that they that they surrender to him, those who had attacked him, they answered:' (Caes. B.G. 4.16.3)

There might be an intuitive sense in which the sender of the messenger is the author of the report. After all, the message presumably expresses the thoughts and intentions of the person or group which sends it. But if we are to account for long-distance binding in a compositional approach, we must account for how long-distance binding can skip the regular antecedent, namely the subject of the report predicate, and take a different antecedent, which in cases like (3a) is introduced several sentences before.

3 Previous Accounts of Antecedents to Latin LDRs

In grammars and the philological literature, it is sometimes mentioned that LDRs in messenger reports can refer to the sender of the messenger rather than the subject of the speech predicate [12, p. 608]; [14, p. 127]. Riemann [17, p. 139] points out that the messenger often speaks on behalf of a group of which he himself is a member, and that the LDR in that case is somewhat similar to a first person plural. In cases where the antecedent is not a group of this sort, Riemann suggests that the messenger repeats verbatim the words of the sender, turning him into a pure medium by which the sender conveys his message.

In my master’s thesis [20], I argued that Latin LDRs are logophors, and Chap. 5 of the thesis presented the outlines of a compositional analysis based on Schlenker’s theory of context shift [18]. I didn’t manage to make sense of the messenger reports, however, as it was difficult to come up with a variable with the denotation of the sender.

Jøhndal [10, chap. 4] agreed with me that a logophoricity analysis of Latin LDRs was justified, and proposed a syntactic logophoricity approach to Latin LDRs within Lexical Functional Grammar. The lexical entry of report predicates contains constraints that specify the clausal complement as a licit domain for long-distance reflexives and designate the subject, or another argument, as antecedent. Jøhndal pointed out that the antecedents in messenger reports can be problematic, because they are not necessarily arguments of the report predicate. He suggested that the subject of the report predicate, the messenger, is designated as the antecedent in the syntax. The reference is resolved to the

---

2 I also argued that there is a different strategy for long-distance reflexivity in Latin, which is not directly relevant to the current topic [20, p. 30-39].
actual antecedent through a pragmatic mechanism outside of syntax. This is possible, because there is some kind of metaphorical relation between the actual antecedent and the messenger, as the latter is a representative of the former.

It appears that it is hard to account for LDRs in messenger reports in a compositional logophoricity analysis, without recurring to pragmatic ad hoc mechanisms, and this calls into question if such analyses are appropriate after all.

4 A New Look at the Data

I will here present new data from a recent data collection, which suggest that messenger reports constitute less of a problem for a logophoricity account. This data collection uses three sources: 1. A review of all instances of the personal reflexive *se* in the Classical Latin subcorpora of the PROIEL treebank [8], consisting of approx. 84,000 tokens. 2. Examples from the philological literature. 3. Queries in PHI Latin Texts (http://latin.packhum.org), a collection of all Latin texts written prior to A.D. 200, approx. 9 million words. I have determined the antecedent of the reflexive myself in examples retrieved from all three sources.

As I mentioned in the previous section, Riemann [17] suggests that LDRs in messenger reports with group antecedents, such as (3a) or (4), are somewhat similar to the first person plural: They refer to the messenger and the group to which he belongs.

\[(4) \text{Ibi ei praesto fuere Atheniensium legati orantes} \]
\[
\text{there him.DAT ready were Athenians.GEN messengers.NOM praying}.
\]
\[
\text{[ut se in obsidione eximeret].}
\]
\[
\text{that REFL.ACC siege.ABL free.PST.SBJV}
\]

‘There he met messengers from the Athenians who begged him to free them from the siege.’ (Liv. 31.14.3; [17, p. 139])

If it could be shown that sender-oriented LDRs are restricted to such group contexts, a plural semantics is all we need to add to the theory.

In (3a) and (4), the messengers serve as subjects for the report predicates, and these predicates are not embedded in a clause with an intensional semantics of any kind. Interestingly, I haven’t found evidence of sender-oriented LDRs in clauses with these properties when the sender is an individual. Sender-oriented LDRs with individual senders occur in two environments in my data. The first environment is utterance predicates embedded within clauses expressing the intention of the sender for sending the messengers. This is exemplified in (3b), where the speech predicate is the subjunctive verb of a relative clause. The subjunctive marks the relative clause as intentional; it expresses the purpose Caesar had for sending the messengers. When it is talk of sending messengers, spies or
letters, LDRs are in fact readily licensed in adjunct intentional clauses, cf. (5), a and b:

(5) a. ... proi, ad Scipionem Pompeium=que nuntios mittit [ut to Scipio Pompeius=and messengers send COMP

  sibi, subsidio veniat].

  REFL.DAT help.DAT come.SBJV

  'He, sends messengers to Scipio and Pompeius, so that they can come
to his aid.' (Caes. B.C. 3.80.3; [10, p.132])

b. ... proi Nicaeam Ephesus=que mittebat [qui rumores

  Nicaea.ACC Ephesus.ACC=and sent who.NOM rumors.ACC

  ... celeriter ad se referrent].

  quickly to REFL.ACC report.SBJV

  'He sent [spies] to Nicaea and Ephesus, who were to report rumors
quickly back to him.' (Cic. Deiot. 9.25; [12, p. 607])

Cases like (3b), where a report predicate is embedded within such an intentional adjunct clause, can quite plausibly be analyzed as parallel to examples such as (5): The sender-oriented LDR are licensed in the intentional relative clause rather than by the report predicate postularent.

The second environment where LDRs refer to an individual sender, is unembedded reports following mitto, 'send', with no overt report predicate:

(6) proi, misit enim puerum: se ad me venire.

  sent PCL boy.ACC REFL.ACC to me come.INF

  'He sends the boy [to say that] he will come to me.' (Cic. Att. 10.16.5;
[10, p. 132])

In this example, mitto is followed by an AcI report. AcIs are used in complements or unembedded reports, they cannot be adverbial clauses. Since mitto does not have the semantics of a complement-taking predicate, it is reasonable to take the AcIs in such examples as unbedded reports.

Unembedded reports which are not introduced by an utterance predicate, are common in Latin, as in (7a), in particular with expressions of sending and coming of messengers and letters, as in (7b):

(7) a. ... sese omnes; flentes Caesari ad pedes proiecerunt:

  REFL.ACC all.NOM crying Caesar.DAT to feet.ACC threw

  non minus se non id contendere ...

  not less REF.ACC it gói strive.INF

  'Crying, they threw themselves at Caesar’s feet: They strived no less
for ...' (Caes. B.G. 1.31.2)

3 Unpronounced, or pro-dropped, subjects are marked with pro in examples when they are relevant.
b. *... litterae ei redduntur a Pompeio: ... properaret ad letter him.DAT give.PASS from Pompeius hurry.SBJV to sei cum exercitu venire ...
REFL.ACC with army come.INF*

‘Letters from Pompeius came to him: ... He should quickly come to him with his army.’ (Caes. Civ. 3.33.1; [6, p. 210])

Reports similar to (3a) and (4), but with individual senders, do occur in my data. In such examples, however, regular personal pronouns, proper names and definite descriptions are used to refer to the sender instead of LDRs.

5 A Sketch of an Analysis

5.1 Centred Worlds Semantics

As we saw in Sect. 2, Latin LDRs occur in reports and refer to what I called the *author of the report*, which often, but not always, coincides with the subject of a report predicate. In this section, I will draw the outline of an analysis based on the semantics of reports, with the aim of showing how the phenomena in Sect. 4 may be accounted for without the addition of pragmatic ad hoc operations.

Semantic theories of logophoricity, such as Pearson’s [16] and Schlenker’s [18], link logophoric reference to the attitudinal semantics of reports: Speech, thought, perception or emotion reports depend on a *propositional attitude predicate*, a predicate which establishes a particular relation, a propositional attitude, between an individual, the attitude holder (AH), and the complement proposition. The interpretation of the complement is relativized to the AH. The proposition *it is raining* in (8) must hold in Peter’s belief worlds for the sentence to be true, but not necessarily in the actual world:

(8) Peter believes that it is raining.

Similarly, complements to predicates such as *hope*, *say*, and *glad* depend on the attitude holders desire, utterance and emotional state respectively.

According to the standard analysis originally proposed by Hintikka [9], attitude predicates are universal quantifiers over worlds which are compatible with the AH’s thought, utterance, emotion etc. Attitude complements are possible world propositions bound by the attitudinal quantifier.

To account for pronouns with logophoric reference, something slightly more sophisticated than world quantification is needed, as we must have a semantics which provides a variable denoting the AH. Lewis [13] argues that propositional knowledge isn’t quite enough to account for first-personal beliefs and desires etc. It is also necessary for the AH to identify herself in logical space. However, to any given proposition, a corresponding property can be constructed, namely that of inhabiting a world in which the proposition holds. If the object of attitudes is properties of this kind rather than the corresponding proposition, the AH
can locate herself, or rather, the counterparts of herself, in the worlds which are in accordance with her attitude. Formally, this can be modelled as a world-individual pair \(<w,x>\) such that it is compatible with what the AH believes (or desires, fears, asserts etc.) for the AH to be \(x\) in \(w\). Such a pair is often called a centred world, and the individual variable is referred to as the center (cf. e.g. [15]).

Pearson [16] gives a formal account of logophoricity using a Lewis-style semantics: Attitude predicates quantify over centred worlds. She assumes a denotation for \(\text{say}\) as in (9):

\[
[[\text{say}]]^g = \lambda P_{e<s.t>}\lambda x_w\lambda w_x, \forall <w',y> \in \text{say}_{x,w}, P(y)(w')
\]

where \(\text{say}_{x,w} = \{<w',y>\mid \text{what x says in w is true in w' and x identifies y as herself in w'}\}\) (cf. [16, Ex.s (13) and (70)])

Logophors are pronouns bound by the center. (10a) is a simplified English translation of (1a), where the logophoric reflexive is translated as \(\text{SE}\). (10) b and c give the computation of its interpretation:

(10) a. Marcus\(_i\) says that \(\text{SE}_i\) has knowledge of everything concerning their number.

\[\lambda w_1 [w_1 \text{Marcus says } \lambda <w_2,x_3> [w_2 \text{ that } \text{SE}_3 \text{ has knowledge of everything concerning their number}]] = \]

b. \(\lambda w_2 \forall <w',y> \in \text{say}_{\text{Marcus},w}, y \text{ has knowledge of everything concerning their number in } w'\)

c. \(\lambda w \forall <w',y> \in \text{say}_{\text{Marcus},w}, y \text{ has knowledge of everything concerning their number in } w'\)

As the logophoric reflexive is bound by the center, the \(y\) variable in (10c), it will obligatorily refer to the individual that Marcus identifies as the counterpart of himself in each world compatible with his utterance.

I will in Sect. 5.2 - 5.4 suggest ways to account for the phenomena described in Sect. 4 within the framework presented here.

### 5.2 Plural Logophoric Reflexives

I argued above that there is a parallel between LDRs in messenger reports and 1st and 2nd person plural. Such plurals do not usually denote pluralities of speakers and addressees, but groups which include speakers and addressees [11, p. 224]. Kratzer [11] captures this with the pronominal plural feature \([\text{group}]\):

\[
[[[\text{group}]]]^{g,c} = \lambda x.\text{group}(x)(c) = \text{the group which includes } x \text{ and } x\text{'s associates with respect to } c \quad ([11, p. 224]; [21, p. 432])
\]

\(^4\) I disregard certain elements of Pearson’s analysis, most importantly, how binding of the logophor by the attitudinal center is made obligatory and how de re readings of logophors are obtained.
Despite the lack of overt plural morphology, there is independent evidence that Latin reflexives can carry plural features. When an LDR with a plural antecedent is the subject of an AcI, it triggers plural agreement\(^5\).

Let us look at what a semantic computation might look like of a messenger report with a plural LDR, given a plural semantics like (11). Example (12) is a constructed example based on (3a). (12b) is the sentence we are interested in. (12a) introduces the messenger context:

(12) a. The Helvetians\(^i\) sent messengers to Caesar. The leader of this embassy was Divico\(^j\).

b. He\(^j\) said that SE\(^i\) had learned not to rely on trickery and plots.

The denotation of (12b) is given in (13), omitting tense and assuming a denotation for \textit{say} as in (9):

\[
\text{[[He said that SE}_{[\text{group}]} \text{ had learned not to rely on trickery and plots}}]^{9, c} = \lambda w. \forall <w', y> \in \text{say}_{\text{Divico}, w}, \text{group}(y)(c) \text{ had learned not to rely on trickery and plots in } w'.
\]

Given the context in (12a), there is a salient group which includes the attitude holder and his associates, namely the Helvetians\(^6\). This approach is very similar to, and inspired by, Stephenson’s treatment of partial control [21, p. 431-433].

It is important to note that this analysis does not avoid pragmatic information entirely: A salient group must be retrieved from the context. However, such a mechanism is needed for independent reasons, to account for the semantics of plurals. Also, the mechanism is less dramatic than the one proposed by Jøhndal [10] and discussed in 3, as it does not shift the reference to someone different from the AH.

Interestingly, this analysis contradicts Stirling’s claim that dedicated logophoric pronouns can refer to a group including the speaker, while logophoric reflexives always require strict identity with the antecedent [22, p. 259].

5.3 Accommodation of Utterance Predicates

We saw in Sect. 4 that LDRs can refer to individual senders in unembedded reports without overt utterance predicates following \textit{mitto}, ‘send’. Unembedded reports which are not introduced by utterance predicates are attested in German [5] and Ancient Greek [1]. (14) is an example from German:

---

\(^5\) Note, however, that Kratzer [11] assumes that reflexives lack \(\phi\)-features.

\(^6\) In the following subsections, I will argue that a dynamic analysis of Latin LDRs is needed. There might be a way to model the context dependency in a dynamic framework in a more precise way than by a context argument to the function \textit{group}. I will present a more elaborate analysis of plural LDRs in future work.
(14) Ich bediente mich des Lautsprechers eines Polizeiwagens: Die Mauer sei härter als Köpfe, die gegen sie anrennen wollten. 'I used the loudspeaker of a police car, [shouting that] the wall would not yield to oncharging heads.' (Example, glosses and translation according to [5, Ex. (31)])

The subjunctive in the second sentence in (14) marks the clause as an utterance report. However, the preceding sentence does not contain an utterance predicate. Nevertheless, it makes an utterance event highly salient in the context, namely one where the first person subject speaks from the police car using a loudspeaker.

(Utterance) reports in German, Latin and Ancient Greek are singled out by special morphological forms: the subjunctive in German, cf. (14), AcI and subjunctive in Latin and optative and AcI in Ancient Greek. Bary and Maier [1] and Fabricius-Hansen and Sæbø [5] provide analyses of unembedded reports in dynamic semantic frameworks. A crucial element of both analyses is that the specialized report forms (subjunctive, optative, AcI) trigger a presupposition that the clause reports someone’s utterance. If the clause is a complement to an utterance predicate or is part of an unembedded report following an utterance predicate, the presupposition is verified against that predicate, either within the clause or in the context. When the presupposition cannot be verified, a speech predicate can be accommodated. This can only happen, however, if a speech event is plausible in the context, as with the mention of a loudspeaker in (14) [5, p. 246-247].

Although the details need to be worked out formally, I believe that a similar dynamic analysis can account for Latin unembedded reports as in (6), where there is talk of the sending of messengers, or the sending or the arrival of letters, as in (7b): An utterance event is highly salient in the context, namely the utterance communicated through messengers or letters.

The sender-reference of LDRs in reports of this kind can be explained in the logophoricity framework set up in 5.1, provided that we take the sender as the agent argument of the accommodated utterance predicate. This is unproblematic in reports like (7b), where letters, not messengers, are involved, as there is no animate individual who competes for the function of utterance agent. The letter serves merely as the physical medium by which the utterance is communicated. This explanation is admittedly less evident with messengers, as the messenger himself is a highly salient candidate for the role of utterance agent, being after all the person who communicates the report to the hearers. In order to evaluate to what extent this is problematic, the analysis has to be properly formalized and paired with a theory of presupposition and accommodation (e.g. [2] or [7]), a task I plan to undertake in the near future.
5.4 Logophoric Reflexives in Intentional Adjunct Clauses

The last environment to account for, is LDRs in reports depending on intentional adjunct clauses explaining why messengers are sent, such as (3b). As I explained in Sect. 4, LDRs with sender reference can be licensed directly in intentional adjunct clauses in this environment, cf. (5). It is therefore reasonable to assume that LDRs in examples like (3b) are licensed in the intentional adjunct clause, not by the embedded report predicate.

It remains to be explained why intentional adjunct clauses which give the purpose for sending messengers, can license LDRs. I see two possible explanation. The first is that intentional adjunct clauses are logophoric licensers in virtue of their intentional semantics, as an intention is a form of attitude. Formally, they could be centred worlds properties in the same way as attitude reports, bound by a covert intentional operator. There would be cross-linguistic evidence for such an analysis, as intentional adjunct clauses are known to license logophors in Ewe [3, p. 154-156]. Latin LDRs are licensed in the complement of all classes of attitude predicates, including embedded commands and desiderative predicates. It would therefore not be too surprising if Latin purpose clauses could license LDRs. There is an empirical problem with this explanation, however: I have found very few LDRs in purpose clauses in my data collection when there is not question of sending messengers or letters.

The second possible explanation is that also these examples involve accommodation of an utterance, namely what the sender says to messengers or writes in letters. I would then need a mechanism of intra-sentential accommodation in my theory (cf. [5]). A desirable advantage of this approach is that LDRs in both unembedded reports and adjunct clauses in messenger contexts are explained by the same mechanism. The choice of one or the other explanation should be made on empirical grounds, however: If further data collection reveals that LDRs can be licensed in purpose clauses more generally, there is no need for a special explanation in messenger contexts.

6 Concluding Remarks

This paper has investigated LDRs in Latin messenger reports, which appear to allow antecedents which are unexpected on a logophoricity analysis of LDRs. Both Jøhndal [10] and I [20] have previously pointed out the difficulty of accounting for what we thought were the correct pattern, without recourse to ad hoc pragmatic operations.

The data collection presented here suggests that the availability of sender-refering LDRs in messenger reports is more restricted than previously assumed, and I have argued that the cases of such LDRs which are attested, can be explained by independently motivated mechanisms: plural semantics, accommodation and logophoric licensing in intentional adjunct clauses. Pragmatic considerations are still needed, but probably no more than what is needed to account for plural pronouns and accommodation in general.
References

1. Bary, C., Maier, E.: Unembedded indirect discourse. In: Urtzi Etxeberria, Ana-
pp. 77–94 (2014)
West-African Languages 10, 141–177 (1975)
5. Fabricius-Hansen, C., Sæbe, K.J.: In a mediative mood: The semantics of the
221 (1987)
8. Haug, D., Eckhoff, H., Hertzenberg, M., Müh, A., Welo, E., Majer, M., Jøhndal,
hf.uio.no/ifikk/english/research/projects/proiel/
bridge (2012)
11. Kratzer, A.: Making a pronoun: Fake indexicals as windows into the properties of
12. Kühner, R., Stegman, C.: Ausführliche Grammatik der lateinischen Sprache, Sat-
(1955)
(1979)
von Thorsten Burkard und Markus Schauer. Wissenschaftliche Buchgesellschaft,
Darmstadt (2000)
15. Ninan, D.: De se attitudes: Ascription and communication. Philosophy Compass
16. Pearson, H.: The interpretation of the logophoric pronoun in Ewe. Natural Lan-
17. Riemann, O.: Études sur la langue et la grammaire de Tite-Live. Ernest Thorin,
Paris (1884)
(2011)
(2010)
Press (1993)
Ex Post “Cheap Talk”: Value of Information and Value of Signals

Liping Tang

Carnegie Mellon University, Pittsburgh PA 15213, USA

Abstract. Crawford and Sobel’s “Cheap Talk” model [1] describes an information communication situation, in which the agents have conflicting interests. A better informed sender will send a possibly noisy signal which may not exactly carries his private information about the true state to the receiver. Then the receiver should process the information from the signal and choose the action which determines the welfare of both. The ex ante (before sender observes the true state) equilibrium properties have been investigated in their work. However, little work has been done from the ex post (after sender observes the true state) view of this game. The main difficult for the ex post analysis is that the original game has no ex post informative Nash equilibrium. Two main reasons are the causes that the game has no ex post informative Nash equilibrium. One reason is the absence of a consistent common model of players’ strategies in the ex post version of this game. The other reason is the confusion resulting from too many signals having close meanings. Therefore, in the ex post version of this game, information and signals can both have negative values with respect to the existence of the informative Nash equilibriums. Corresponding to these two reasons, two methods have been tried to save the ex post informative Nash equilibriums. One way is to build up a nested epistemic model for players. As a result, only pooling equilibrium exits as the iterated levels of players’ beliefs increase. The second way is by restricting players’ information about the world. Under this modification, some informative Nash equilibriums can be found sometimes.

1 Introduction

There are many situations in real life that decisions have to be made by individuals who have conflicting interests. For example, in job market, potential employees send signals about their ability levels to the employer by acquiring their education credentials. If the potential employee is highly qualified for some job, then he should be honest on showing his education background. However, if the potential employee is actually a bad worker, then his credibility might be questioned by the employer. Therefore, under the situation where a bad worker wants to be hired, what information he should signal the employer and how much the employer should trust this potential employee are the kind of questions concerning in Crawford and Sobel’s work (CS model). The CS model describes an
information communication situation, in which the agents have conflicting interests. A better informed sender sends a possibly noisy signal which may not exactly carry his private information about the true state to the receiver. Then the receiver should process the information from the signal and choose the action which determines the welfare of both. The original CS model investigated ex ante solution concepts for this game. Namely, players make decisions before the sender observes the true state. However, the ex ante informative equilibrium is not the ex post equilibrium any more. Very little work has been discussed from ex post point of view (decisions are made after sender observes the true state) for this game. The goal of my work is to explore the “cheap talk” model from the ex post point of view.

There are two main reasons for the absence of the ex post informative equilibriums for this game. One is that players have no ex post consistent model about each other’s strategy structure under common knowledge assumption of their epistemic statuses. By relaxing this common knowledge assumption, a nested epistemic model for players is introduced. Under this new model, one player always believes one level more than the other and the other doesn’t realize his ignorance. As each player believes more and more about each other’s beliefs, it is found that the game behaves like the pooling case after both players reach certain level of belief assumption.

As the second reason, the absence of informative equilibriums also results from the expansion of the set of credible signals in the ex post version of this game. When the set of signals consists of several signals having close meanings, confusions arise. Along this line of exploration, we restrict players’ information about the world while keeping the expanded set of signals. In this restricted information model, sender is receiving a vague information (an interval) about the state rather than a fixed value of the state. Surprisingly, some informative equilibriums can be saved sometimes. In this sense, information about the world has a negative value with respect to the existence of the ex post informative Nash equilibriums. It is because when sender knows less, ex post informative Nash equilibriums exist sometimes.

The remainder of the paper is arranged in the following way. Section 2 introduces the basic settings of CS game and the limitations of this model for the ex post analysis. Section 3 provides two possible ways to overcome the difficulties. Paper ends with a simple conclusion.

2 “Cheap Talk” Model and Its Limitations

2.1 CS Model and Its Main Results

The CS model consists of two players, a sender(S) and a receiver(R). Only S can observe the value of the random variable \( m \) which is uniformly distributed on

---

1 Stalnaker[6] has discussions about credibilities of signals.

2 Details will be provided in section 2 on how the set of credible signals changes for the ex post CS model.
[0, 1]. Then S should send a signal to the receiver, which may be random, but possibly noisy due to their conflicting interests. Based on the signal, R chooses the action \( y\), a real number which determines both players’ payoffs. Their payoff functions are as follows.

\[
U^S = -(y - (b + m))^2 \\
U^R = -(y - m)^2
\]

where \( b \) is a positive real number which is used to measure how nearly players’ interests coincide. From the payoff functions, we notice that for R, he prefers \( y \) to be as close as the true value of \( m \). However, he can not observe \( m \) directly but receive a signal about the value of \( m \) from S. For S, he wishes that \( y \) can be close to \( b + m \), which differs from \( m \) by \( b \). Therefore, S may not send the true value of \( m \) but a signal which leads R to think that \( m \) is close to \( b + m \) such that R can let \( y \) to be closer to \( b + m \). Thus as \( b \) getting bigger, the conflict between the players is getting larger. It is easy to check that truth telling can not be a Nash equilibrium. Except the truth telling case, another extreme situation is that R ignores S’s signal and chooses \( y \) to be the expectation of \( m \), in this case, \( 1/2 \) all the time. Obviously, players’ these strategies form a Nash equilibrium. But this equilibrium is a pooling equilibrium. No information is transferred between players.

Are there any other Nash equilibriums? Crawford and Sobel borrowed the idea of partition equilibriums from Green and Stokey[4]. The idea is the following. We start from the simplest case. Suppose \( a \) is close to the true value of \( m \). However, he can not observe \( m \) directly but receive a signal about the value of \( m \) from S. For S, he wishes that \( y \) can be close to \( b + m \), which differs from \( m \) by \( b \). Therefore, S may not send the true value of \( m \) but a signal which leads R to think that \( m \) is close to \( b + m \) such that R can let \( y \) to be closer to \( b + m \). Thus as \( b \) getting bigger, the conflict between the players is getting larger. It is easy to check that truth telling can not be a Nash equilibrium. Except the truth telling case, another extreme situation is that R ignores S’s signal and chooses \( y \) to be the expectation of \( m \), in this case, \( 1/2 \) all the time. Obviously, players’ these strategies form a Nash equilibrium. But this equilibrium is a pooling equilibrium. No information is transferred between players.

### Theorem 1

Suppose the information transmission model and \( b > 0 \). Then there exists a positive integer \( N(b) \) such that, for every integer \( 1 \leq N \leq N(b) \), there exists one partition equilibrium \((y(n), q(n|m))\), where \( q(n|m) \) is uniform, supported on \([a_{i-1}, a_i]\) if \( m \in (a_{i-1}, a_i) \), such that

1. \( a_i = a_1 i + 2i(i - 1)b \) where \( i = 1, \ldots, N \)
2. \( y(n) = (a_{i-1} + a_i)/2 \) for all \( n \in (a_{i-1}, a_i) \)
3. \( a_0 = 0 \) and
4. \( a_N = 1 \).

An example is helpful to understand this theorem. Let \( b=1/20 \), then by condition 1, we have \( 2N(N - 1)b < 1 \). Thus, the largest integer satisfying this
inequality is $N(b) = 3$. Therefore, according to Theorem 1, for each integer smaller and equal to 3, there is a partition equilibrium. By condition 1 and 2, each case can be calculated as follows.

**Example 1**

1. $N = 1$
   \[ I_{11}[0, 1] \]
   \[ U^S = -\left(\frac{27}{60} - m\right)^2 \]

2. $N = 2$
   \[ I_{21}[0, \frac{24}{60}] \quad I_{22}[\frac{24}{60}, 1] \]
   \[ y = \frac{12}{60} \quad y = \frac{24}{60} \]
   \[ U^S = -\left(\frac{9}{60} - m\right)^2 \quad U^S = -\left(\frac{39}{60} - m\right)^2 \]

3. $N = 3$
   \[ I_{31}[0, \frac{8}{60}] \quad I_{32}[\frac{8}{60}, \frac{28}{60}] \quad I_{33}[\frac{28}{60}, 1] \]
   \[ y = \frac{4}{60} \quad y = \frac{18}{60} \quad y = \frac{34}{60} \]
   \[ U^S = -\left(\frac{1}{60} - m\right)^2 \quad U^S = -\left(\frac{15}{60} - m\right)^2 \quad U^S = -\left(\frac{41}{60} - m\right)^2 \]

Meanwhile, a comparative result for partition equilibriums is also given in Crawford and Sobel’s paper [1].

**Proposition 1.** For a given value of $b$, the partition equilibrium of size $N(b)$ is the ex ante Pareto-superior equilibrium to all other equilibriums.

Two things should be emphasized about this proposition. First, the partition equilibrium of size $N(b)$ is Pareto-superior for both players to all other equilibria. Therefore, this equilibrium should be the optimal one if the players have the opportunity to pick the partition. Second, this Pareto-superior result is ex ante which means this result hold only before the true value of $m$ has been observed. The ex post case requires a different analysis which will be the main discussion in the next section.

### 2.2 Limitations on CS Model

As is mentioned earlier, the analysis of CS model is based on the ex ante perspective. Namely, a commitmennt on which partition equilibrium to play is decided before the true value of $m$ is observed. However, no reason has been given on how and why any agreement on the partition equilibrium should be reached in advance. An explanation depending on Proposition 1 may be given that the partition equilibrium of size $N(b)$ should be chosen since it is Pareto-superior. However, the Pareto-superior result is an ex ante result. After the value of $m$ is revealed, the ex ante Pareto equilibrium may not be optimal any more. For example, let $b = \frac{1}{20}$ and $m = \frac{10}{60}$. If the $N = 3$ partition equilibrium is used, which is the ex ante Pareto-superior equilibrium, sender should send the signal $I_{32}$ according to Example 1. As a result, given receiver’s strategy of taking the middle point of that interval, $U^S = -\left(\frac{5}{60}\right)^2$. But this is not optimal for $S$. If $N = 2$ partition equilibrium is applied, $U^S = -\left(\frac{1}{60}\right)^2$. Thus, apparently, partition equilibrium of size $N(b)$ is not ex post optimal and Proposition 1 can not be applied in the ex post case.
One way to avoid the problem is to drop the commitment such that all the partition equilibriums are available for the players to consider in a particular way. Let’s come back to Example 1. If we pile all those three partition equilibriums together, players are committed to the set of signals to be \(\{I_{11}, I_{21}, I_{22}, I_{31}, I_{32}, I_{33}\}\). For the receiver, he should read signals in an indicative way. Namely, once \(I_{ij}\) is observed, receiver should choose the action \(y\) to be the expectation on the support of the \(j\)th interval within the \(N = i\) partition equilibrium. As a result, comparing to the original case that sender’s credible signals form only from one partition equilibrium, for instance, \(\{I_{31}, I_{32}, I_{33}\}\), sender’s set of credible signals is expanded from \(N\) to \((1+N)/2\). Now we have an ex post CS game without commitment on a particular partition equilibrium in advance and a new ex post decision process is as follows: S observes the true value of \(m\); then S chooses one of \((1+N)/2\) number of available interval signals according to \(b\), \(m\) and \(U^S\); R reads the signal and chooses \(y\).

Let’s check if this process can go through for the current model. It is easier to explain by an example. Let’s take \(b = 1/20\) for illustration again. Sender’s payoff functions are calculated for different signals. What we can do now is to find sender’s optimal strategy according to the value of \(m\) given that receiver’s strategy is to send the middle point of the intervals. This is not difficult to calculate by the structures of the payoff functions. We found the pattern as follows.

\[
\begin{align*}
[0, \frac{5}{60}) & \quad (\frac{5}{60}, \frac{12}{60}) & \quad (\frac{12}{60}, \frac{21}{60}) & \quad (\frac{21}{60}, \frac{33}{60}) & \quad (\frac{33}{60}, \frac{40}{60}) & \quad (\frac{40}{60}, 1] \\
I_{31} & \quad I_{21} & \quad I_{32} & \quad I_{11} & \quad I_{22} & \quad I_{33}
\end{align*}
\]

This pattern shows the correspondences between the value of \(m\) and the optimal interval signals for the sender. For example, if \(m \in [0, \frac{5}{60})\), then sender’s optimal choice of signal should be \(I_{31}\) which is the first interval of partition equilibrium of size 3. So far so good.

Without lose of generality, let’s suppose \(m = \frac{10}{60}\). So we know that sender should send \(I_{21}\). The corresponding interval for \(I_{21}\) is \((\frac{5}{60}, \frac{12}{60})\) for sender. After receiving this signal, what should the receiver do? He knows first that \(I_{21}\) means \(m \in [0, \frac{21}{60})\) by Example 1, thus he should choose \(y = \frac{21}{60}\). Secondly, receiver knows sender’s preference on the payoff functions with respect to \(m\). So he knows \(I_{21}\) means \(m \in (\frac{5}{60}, \frac{12}{60})\) which is a finer and more accurate interval than \(m \in [0, \frac{21}{60})\). Thus, receiver should pick \(y = (\frac{5}{60} + \frac{12}{60})/2 = \frac{8.5}{60}\) which is inconsistent with reading the signal directly. Contradiction arises that sender chooses his optimal signal \(I_{21}\) by using the payoff functions with \(y = \frac{21}{60}\) while receiver picks \(y = \frac{8.5}{60}\) after he gets \(I_{21}\). Obviously, the contradiction exists for any \(m\).

According to the above analysis, the informative Nash equilibriums don’t exist for the ex post CS game. There are two main reasons for the ex post dilemma. One is the absence of players’ consistent model for strategies. The other one is from the expanded set of signals in which some signals having close meanings with respect to certain value of \(m\). Two approaches are tried to solve the problem. One way is to relax the common knowledge assumption about

\[\text{Credibility of signals is discussed in Stalnaker[6], Jäger[3] and Rabin[5]}\]

\[\text{See Example 1}\]
players’ strategy models. A nested epistemic model for players is created where one player always believes one level more than the other. Under this modification, the convergence result of this model is investigated. The other way is to keep the richness of the language but restrict sender’s information about the world (states), namely, the value of $m$. We let sender receive the value of $m$ from the form of a fixed point to the form of an interval. Fortunately, some informative Nash equilibriums can survive sometimes in this model.

3 Two Approaches for the Ex Post Analysis

3.1 Less Information about Opponent’s Beliefs

According to the argument above, the absence of the informative equilibrium in the ex post CS game arises when we assume the common belief of rationality, namely, both players believe each other’s reasoning patterns, rationalities and they both believe each other believes their reasoning patterns, rationalities and they both believe they both believe ···. Therefore, for keeping the ex post analysis going through and avoiding the contradiction, we have to break the common belief chain. A new epistemic reasoning pattern has to be built to avoid the inconsistency. We start with a very naive case when receiver just reads sender’s signal directly, then build up receiver’s epistemic model level by level. An example for illustration is given as follows. We keep using the example of $b = \frac{1}{20}$.

**Example 2**

1. Level 0: Receiver is naive in the sense that he simply reads sender’s signal directly by taking the midpoint of that signal according to the original interval patterns in Example 1.

By assuming level 0, sender will pick one of the six signals according to where the true $m$ is as follows denoted as $L_0 m$ intervals.

$I_{31}$ $I_{21}$ $I_{32}$ $I_{11}$ $I_{22}$ $I_{33}$

2. Level 1: Receiver awares that sender believes his level 0 naiveness and sender is using the $L_0 m$ for picking signals. So Receiver will apply the $L_0 m$ intervals for choosing the proper $y$.

Thus, sender’s utility function will change to the following six ones for different signals.

$I_{31} : U^s = -\left(-\frac{0.5}{60} - m\right)^2$ $I_{21} : U^s = -\left(\frac{3.5}{60} - m\right)^2$

$I_{32} : U^s = -\left(-\frac{19.5}{60} - m\right)^2$ $I_{11} : U^s = -\left(\frac{29}{60} - m\right)^2$

$I_{22} : U^s = -\left(-\frac{33.5}{60} - m\right)^2$ $I_{33} : U^s = -\left(\frac{57}{60} - m\right)^2$

By assuming level 1, sender will pick one of the six signals according to where the true $m$ is and the new set of utility functions are in the following way, denoted as $L_1 m$ intervals.

$I_{31}$ $I_{21}$ $I_{32}$ $I_{11}$ $I_{22}$ $I_{33}$

$\frac{0.5}{60}$ $(\frac{2.5}{60}, \frac{9.5}{60})$ $(\frac{9.5}{60}, \frac{18.75}{60})$ $(\frac{18.75}{60}, \frac{28.75}{60})$ $(\frac{28.75}{60}, \frac{40.25}{60})$ $(\frac{40.25}{60}, 1)$
3. Level 2: Receiver aware that sender believes his level 1 beliefs and sender is using the L1m for choosing signals. So R will apply the L1m intervals for choosing the proper $y$. Thus, sender’s utility function will change to the following six ones for different signals.

$I_{31} : U^* = -(\frac{2.25}{60} - m)^2$
$I_{32} : U^* = -(\frac{11.75}{60} - m)^2$
$I_{22} : U^* = -(\frac{3}{60} - m)^2$

By assuming level 2, sender will pick one of the six signals according to where the true $m$ is and the new set of utility functions are in the following pattern, denoted as L2m intervals.

$$[0, \frac{0.375}{60}) \ (\frac{0.375}{60}, \frac{7.0625}{60}) \ (\frac{7.0625}{60}, \frac{15.9375}{60}) \ (\frac{15.9375}{60}, \frac{26.125}{60}) \ (\frac{26.125}{60}, \frac{39.3125}{60}) \ (\frac{39.3125}{60}, 1]$$

$I_{31}$ $I_{21}$ $I_{32}$ $I_{11}$ $I_{22}$ $I_{33}$

4. Level 3, similar calculation applies to L3m.

$$[0, \frac{4.609375}{60}) \ (\frac{4.609375}{60}, \frac{13.27}{60}) \ (\frac{13.27}{60}, \frac{23.875}{60}) \ (\frac{23.875}{60}, \frac{38.19}{60}) \ (\frac{38.19}{60}, 1]$$

$I_{21}$ $I_{32}$ $I_{11}$ $I_{22}$ $I_{33}$

5. 

At each level, sender models receiver’s beliefs as one level less compared to his beliefs. Meanwhile, sender’s model of receiver’s nested belief model can grow again and again. Each level is a cutting of the common belief chain. As a result, we avoided the inconsistency successfully. But how far can this process continue? Looking at the example more carefully, we found that from L2m to L3m, the number of intervals already changes from 6 to 5 due to the structure of the payoff functions and the range of $m$. If we do continue this process for more steps, we can see that the intervals drop off one by one from the left. As levels go up, we can imagine that all the signals will disappear except $I_{33}$ which is the right most interval of the largest partition equilibrium. Sender will always send the same signal and receiver takes the same action. The game behaves like the ex ante pooling equilibrium, which is the least informative equilibrium. The result holds for any particular $b$.

**Theorem 2.** Given two players’ CS information transmission model and any particular $b > 0$, the ex post nested epistemic model from sender’s point of view leads the game to a pooling equilibrium eventually.

As building up players’ epistemic model, we actually give players more information about other players’ beliefs. However, as the theorem shows, as the levels grow, we eventually end up with the pooling case which is not an efficient Nash equilibrium. Thus, in this sense, more information (beliefs about others’ beliefs) doesn’t bring more benefits. So information actually has a negative value in this sense.

Meanwhile, the result in the theorem above is consistent with the result in the IBR model [2]. As the IBR model predicates, the fix point of players’ strategies under the IBR model is the perfect Bayesian equilibrium. Since the ex post CS model only has the pooling equilibrium, therefore, as Theorem 2 states, players
strategies converge to the only equilibrium of this game even if it is not a good one.

A simulation can be operated to check the speed of convergence. It turns out that convergence occurs very quickly. For example, if $b = 1/20$, at level $L_{14}$, all the divided points become negative, and only pooling equilibrium remains. If $b = 1/50$, at level $L_{42}$, all the divided points become negative. And only pooling equilibrium remains. Therefore, the smaller the $b$ is, the slower the converge speed is.

3.2 Less Information about the World

As we have shown so far, the ex post version of CS game has no informative Nash equilibrium under neither the common knowledge assumption about players’ rationalities nor players’ nested epistemic model. In this section, I want to try the other route to save the informative Nash equilibriums for the ex post version of the game. The strategy here is to restrict the value of $m$ such that sender is given an interval of $m$ rather than a fixed point while keeping the larger set of credible signals.

For example, when $b = 1/20$, we have the following partition equilibriums.

\[ N = 1 \quad I_{11} [0, 1] \]
\[ N = 2 \quad I_{21} [0, \frac{24}{60}] \quad I_{22} [\frac{24}{60}, 1] \]
\[ N = 3 \quad I_{31} [0, \frac{8}{60}] \quad I_{32} [\frac{8}{60}, \frac{28}{60}] \quad I_{33} [\frac{28}{60}, 1] \]

Suppose the available information set about $m$ which is given to sender is \{\([0, \frac{8}{60}]), ([\frac{8}{60}, \frac{28}{60}]), [\frac{28}{60}, 1]\}, which are the intervals from the partition equilibrium of size 3 and receiver knows about it. Then sender sends $I_{31}$ given interval $[0, \frac{8}{60}]$, $I_{32}$ given interval $[\frac{8}{60}, \frac{28}{60}]$ and sends $I_{33}$ given $[\frac{28}{60}, 1]$. Moreover, $I_{31}, I_{32}, I_{33}$ are the best responses for those three intervals to sender given that receiver takes the middle point of each interval. On the other hand, given that $m$ locates in any interval $[a, b]$, receiver’s best response is the expectation of $m$ on $[a, b]$ which is $E_x(m) = \frac{1}{2}(a + b)$, the middle point of the interval. Therefore, sender’s strategy that sending the corresponding signal for the given interval and receiver’s strategy that taking the middle point of the given signal form a Nash equilibrium. And this Nash equilibrium does carry some information about the world.

Therefore, in the example above, we saved one ex ante partition equilibrium which has the finest partitions by restricting the information of $m$. The question is can we save all other partition equilibriums in the same way. Unfortunately, the answer is no. Take $b = 1/20$ for example again. Suppose sender is given the interval $[\frac{24}{60}, 1]$ that is from the partition equilibrium of size 2. If he follows the strategy before, he should send $I_{22}$ which yields $y = E_x(m) = \frac{42}{60}$. However, according to sender’s utility function, $U^s = -(y - (b + m))^2$, the optimal $y$ for him is $E_x(m) + b$. Thus, sender intends to send the signal which yields a result being closer to $E_x(m) + b$ than the true expectation of $m$. So in the case when he is given interval $[\frac{24}{60}, 1]$, he is doing better by sending $I_{33}$ which yields $y = \frac{44}{60}$. Therefore, in this example, sender should not send the original given interval but some other signals. On the other hand, if receiver knows about sender’s strategy, taking the middle point is not the best response for him. Thus, the
partition equilibrium of size 2 can not survive as a Nash equilibrium. Therefore,
the set \( I \) may not include all the intervals from all of the ex ante partition
equilibriums as we have shown. We want to have a systematic method to find
what would the set \( I \) look like for any \( b \). The following proposition provides us
one way to find \( I \).

**Proposition 2.** Suppose under the information restricted model, sender is given
an interval \([a, b]\), which yields \( y_{ab} = \frac{1}{2}(a + b) \). Sender would strictly prefer to send
\( I_{a'b'} \) with respect to \([a', b']\) which yields \( y_{a'b'} = \frac{1}{2}(a' + b') \) if \( 0 < y_{a'b'} - y_{ab} < 2b \).

Step 1: Given any \( b \), find all the partition equilibriums by the original CS
model;
Step 2: Find all the middle point \( d_i \)s for each interval of each partition equi-
libriums. And Compare each \( d_i \) with all other \( d_j \)s.
Step 3: By proposition 2, if \( 0 < d_j - d_i < 2b \), then we say \( d_i \) is replaceable.
Step 4: \( I \) consists all the intervals from the ex ante partition equilibriums
without replaceable \( d_i \)s.

By those steps, we can build a simulation to find \( I \). For example,

<table>
<thead>
<tr>
<th>( b )</th>
<th>( N(b) )</th>
<th>Survived Partition Equilibriums</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.004</td>
<td>11</td>
<td>( N=1, N=3 )</td>
</tr>
<tr>
<td>0.002</td>
<td>16</td>
<td>( N=1, N=2, N=3, N=5 )</td>
</tr>
<tr>
<td>0.0014</td>
<td>19</td>
<td>( N=1 )</td>
</tr>
<tr>
<td>0.00001</td>
<td>224</td>
<td>( N=1, N=2 )</td>
</tr>
</tbody>
</table>

As we have known from the original CS model, the smaller the \( b \) is, the bigger
\( N(b) \) is and more partition equilibriums we have. As \( b \) goes to 0, we will have
infinite many intervals. So the length of the intervals gets smaller and smaller,
and all the middle points become closer and closer. A reasonable conjecture is as
\( b \) goes to 0 but not equal to 0, no partition equilibrium would survive. Thus, no
informative Nash equilibrium exists even for the restricted information model.
The limiting case as \( b \) goes to 0 is considered and a robust result is found as
follows.

**Theorem 3.** Given the restricted information version of CS game, as \( b(b > 0) \)
goes to 0, the partition equilibrium with the finest partitions \( N(b) \) can not survive.

Proof:
For proving this result, we need to find an infinite sequence of partition equi-
libriums as \( b \) decreases and \( N(b) \) increases such that there is at least one middle
point \( d_{N(b)i} \) for one interval in the partition equilibrium with \( N(b) \) number of
partitions is replaceable (i.e. sender would like to switch from \( d_{N(b)i} \) to other
interval).

It is found that the following inequality holds infinitely often as \( b \) goes to 0:

\[
0 < d_{(N-2)i} - d_{N2} < 2b
\]

where \( d_{ij} \) means the middle point of the \( j \)th interval in the partition equilib-
rium with \( i \) number of partitions. Thus, \( d_{(N-2)i} \) should be read as the midpoint
of the first interval of the partition equilibrium with $N - 2$ number of partitions. And $N = N(b)$ is the largest integer satisfying $2n(n - 1)b < 1$ which is the requirement for the existence of partition equilibriums from the original CS game.

According to Proposition 2, $d_{N2}$ is replaceable, so the finest partition equilibrium can not be put into the set $I$, so the partition equilibrium of size $N$ can not survive.

For showing $0 < d_{(N-2)1} - d_{N2} < 2b$

holds infinitely often for some $b$ as $b$ decreases and $N(b)$ increases. We can show the following.

- According to the original CS model, given $b$, $N(b)$ is calculated from $2n(n - 1)b < 1$. By this inequality, as $b$ decreases, $N(b)$ increases. Moreover, as $b$ goes to 0, $N(b)$ goes to infinite. Therefore, there is an infinite sequence of $N(b)$ as $b$ goes to 0;

- For $N(b) = N$ to be the largest integer satisfying $2n(n - 1)b < 1$, $b$ has to satisfy the following:

$$\frac{1}{2N(N+1)} < b < \frac{1}{2N(N-1)} \quad (1)$$

- Since

$$d_{(N-2)1} = \frac{1}{2(N-2)} - bN + 3b$$

$$d_{N2} = \frac{3}{2N} - 3Nb + 5b$$

in order to obtain the inequality $0 < d_{(N-2)1} - d_{N2} < 2b$, $b$ has to satisfy the following:

$$\frac{1}{2N(N-1)} * \frac{(N-3)}{(N-2)} < b < \frac{1}{2N(N-2)} * \frac{(N-3)}{(N-2)} \quad (2)$$

For large $N$, the following holds.

$$\frac{1}{2N(N+1)} < \frac{1}{2N(N-1)} * \frac{(N-3)}{(N-2)} < \frac{1}{2N(N-2)} * \frac{(N-3)}{(N-2)} < \frac{1}{2N(N-1)} \quad (3)$$

Thus, given any $b$ satisfying (2) satisfies (1). Therefore, there is an infinite sequence of $N$ where

$$0 < d_{(N-2)1} - d_{N2} < 2b$$

occurring infinitely often as $b$ goes to 0. □

As Theorem 3 shows, in the limit case where $b$ goes to 0, the partition equilibrium with the finest partition can not survive for the restricted information model. There is a sharp difference between our model and Crawford and Sobel’s
model in the limit case. According to the original CS model, \( b \) measures how nearly agents’ interests coincide. When \( b \) goes to 0, players interests is coinciding. Therefore, players should follow the partition equilibrium with the finest partitions. Namely, players converge to truth telling. While our result shows that even if truth telling still exists if \( b = 0 \). However, in the process of \( b \) getting close to 0, the partition equilibrium with the finest partitions will never be chosen under the restricted information model. It is hard to tell why would that happen since the restricted information model is far apart from the original CS model. We first develop an ex post version of the CS game. Then restrict sender’s information about \( m \). As a result, some counter intuitive result appears.

4 Conclusion

In this paper, Crawford and Sobel’s information transaction model is explored from the ex post point of view. The dilemma is found while we pile all the ex ante partition equilibriums and all the signals together. There is no ex post informative Nash equilibrium for the ex post CS game. We suggested two approaches to investigate the ex post case and discussed their equilibrium properties.

The first model is to relax the common knowledge assumption about players’ rationalities and build up a nested epistemic model from one player’s point of view. As a result, we showed that as the levels of players’ beliefs go up, only the pooling equilibrium exists.

The second model is to restrict information about the states \( m \) which is given to sender. We found that the number of informative Nash equilibriums depends on the value of \( b \). Moreover, we considered the limit case within the restricted information model when \( b \) goes to 0. It is found that in the limit case, the finest partition equilibrium does not exist as \( b \) goes to but not equal to 0.

References

Abstract. This paper consists of three parts. First, I show that one most plausible reading of the Chisholm set, that raises Chisholm’s Paradox, is not adequately represented in ordering semantics. I present this issue with a new deontic puzzle, the CTD (contrary-to-duty) trilemma. Second, to solve the puzzle it is necessary to acknowledge the presence of the two kinds of ‘ought’-statements in English and to take them into a formal account. Finally, I propose a dynamic approach which accounts for two meanings of ‘ought’s based on Kratzer semantics and show how it solves the CTD trilemma.

1 Introduction

The nature of moral statements such as ‘ought’-sentences and indicative sentences containing moral predicates like “is wrong” has been the central issue in the intersection of moral philosophy and philosophy of language. Despite the ongoing debate on how to draw the distinction, it is widely acknowledged that moral statements can be prescriptive as well as descriptive. However, formal analyses of normative ‘ought’-statements rarely take these two aspects of normativity into account. The importance of taking account of these two aspects of moral statements in formal accounts will emerge as we witness the shortcoming of ordering semantics regarding contrary-to-duty(CTD) obligations. In the following two sections, I will show that although deontic ordering semantics is immune from the classic challenge, Chisholm’s Paradox, it cannot avoid another deontic puzzle with the Chisholm set, which I will call the CTD trilemma. In Sect. 4, I will argue that taking account of the two aspects of normativity is crucial in avoiding the trilemma, and the most natural way of capturing two aspects of normativity in a formal account is a dynamic approach. In Sect 5., I will sketch a dynamic account based on Kratzer semantics with diagrams. In Sec 6., I will show how the CTD trilemma is solved in the dynamic account.

2 The Original Challenge: Chisholm’s Paradox

Chisholm’s paradox is designed to show that Standard Deontic Logic (SDL) has a serious limitation in representing the concept of contrary-to-duty(CTD)obligations. Considering the following set of sentences.
**The Chisholm Set (2)**

(c1) It ought to be that Jones goes to help his neighbors.
(c2) It ought to be that if he does go he tells them he is coming.
(c3) If he does not go then he ought not to tell them he is coming.
(c4) He does not go.

(c3) is the CTD obligation of the Chisholm set, since (c3) is an obligation regarding a sub-ideal situation where the primary obligation, (c1), is violated. It is not hard to imagine the situation where (c1)-(c4) all hold at the same time. However, SDL cannot adequately represent the Chisholm set without inconsistency or redundancy. SDL has the two possible ways of representing conditional ‘ought’-statements like (c2) and (c3): a conditional that is obligatory, “O(p → q),” or a conditional with an obligatory consequence, “p → O(q).” So there are three possible ways of representing the Chisholm set as follows.

<table>
<thead>
<tr>
<th></th>
<th>(c1)</th>
<th>(c2)</th>
<th>(c3)</th>
<th>(c4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O(h)</td>
<td>O(h → t)</td>
<td>¬h → O(¬t)</td>
<td>¬h</td>
</tr>
<tr>
<td>(c1)</td>
<td>O(h)</td>
<td>h → O(t)</td>
<td>¬h → O(¬t)</td>
<td>¬h</td>
</tr>
<tr>
<td>(c2)</td>
<td>O(h → t)</td>
<td>O(h → t)</td>
<td>¬h</td>
<td>¬h</td>
</tr>
<tr>
<td>(c3)</td>
<td>O(¬h → ¬t)</td>
<td>O(¬t)</td>
<td>¬h → O(¬t)</td>
<td>¬h</td>
</tr>
<tr>
<td>(c4)</td>
<td>¬h</td>
<td>¬h</td>
<td>¬h</td>
<td>¬h</td>
</tr>
</tbody>
</table>

In the first column, (c1) entails (c3); in the second column, (c4) entails (c2); in the third column, what is entailed from (c1) and (c2) is incompatible with what is entailed from (c3) and (c4) – “O(t)” and “O(¬t).”

To overcome this classic challenge, Dyadic Deontic Logic (DDL), which takes two-place deontic operators as primitive, has been suggested by von Wright. ([19] and [20]; combined and reprinted in [8]) Ordering semantics such as the Hansson-Lewis system and Kratzer semantics are widely accepted as proper semantic accounts for DDL these days. ([3], [4], [6], [11], [12], [13], [14], [15].)

Unlike SDL, they are considered as immune from Chisholm’s Paradox, since they can take a two-place deontic operator, “O(/)” as primitive and invalidate factual detachment. The dyadic notation, “O(q/p),” is read as the obligation to bring about q in case of p, or as it ought to be that q given that p. An unconditional obligation in a natural language is represented as “O(p/⊤)” in DDL – ⊤ stands for a tautology. Thus, the Chisholm set can be represented in dyadic systems as follows.

(c1) O(h/⊤)
(c2) O(t/h)
(c3) O(¬t/¬h)
(c4) ¬h

To see how to give the truth condition of the dyadic operator, consider Kratzer semantics. Kratzer semantics for conditionals and modals has two contextual parameters, the modal base, f and the deontic ordering source, g. The modal base, f, is a function from a world, i, to a set of propositions, demarcating the set of possible worlds accessible from i. The ordering source, g, is a function from i to a set of propositions, which induces an ordering over the modal base. g(i) for
the deontic modal operator is considered as a set of moral duties and obligations holding in \( i \). Thus, the truth condition of “\( \text{O}(q/p) \)” is given as follows in Kratzer semantics:

\[
\llbracket \text{O}(q/p) \rrbracket^{f,g} = 1 \text{ iff } \{ w : w \in \cap f^+p(i) \land \neg \exists v \in \cap f^+p(i) : v \not\approx_{g(i)} w \} \subseteq \llbracket q \rrbracket,
\]

provided that \( \cap f^+p(i) = \{ w : w \in \cap f(i) \land w \in \llbracket p \rrbracket \} \).

Informally, “\( \text{O}(q/p) \)” means that the best deontic worlds among \( p \)-worlds in \( f(i) \) are \( q \)-worlds given the Limit Assumption. Thus, (c1) means that the deontic ideal worlds are ones in which Jones goes to help his neighbors; (c2) means that the deontic ideal worlds among those in which Jones goes are ones in which Jones tells them he is coming; (c3) means that the deontic ideal worlds among those in which Jones does not go are ones in which Jones does not tell them he is coming.

Given this, none of them implies other sentences of the Chisholm set; there is no redundancy. And, while deontic detachment from “\( \text{O}(q/p) \)” and “\( \text{O}(p) \)” to “\( \text{O}(q) \)” is valid, factual detachment from “\( \text{O}(q/p) \)” and “\( p \)” to “\( \text{O}(q) \)” is not. So while (c1) and (c2) entails “\( \text{O}(t) \)”, (c3) and (c4) do not entail “\( \text{O}(\neg t) \)” Thus, there is no inconsistency in representing the Chisholm set in ordering semantics.

3 The New Challenge: the CTD Trilemma

Ordering semantics for deontic modality easily meets the classic challenge to represent the Chisholm set without inconsistency and redundancy. However, that is only half the victory. They fail to predict that Jones ought not to tell them he is coming when he does not go to help them. In other words, the most plausible reading of the Chisholm set cannot be captured by the account. Consider the following CTD situation in which all sentences of the Chisholm set hold.

Suppose that Jones has three duties as a member of a community. Jones ought to go to help his neighbors in distress; Jones ought to give them a notice that he is coming before he gets there; and Jones ought not to tell them he is coming if he does not go to help them, because that would be a lie. Imagine the situation where his neighbors need Jones’s help, so Jones ought to go to help...

---

1 “\( u \not\approx_{g(i)} v \)” means \( u \) is at least as good as \( v \) given the set of moral principles from the standpoint of \( i \).

2 Consider the dyadic version of factual detachment first. Even when \( p \) is true at \( i \) and all the deontic ideal \( p \)-worlds from \( i \) are \( q \)-worlds, it is possible that \( \text{O}(q) \) is false at \( i \); it is because Weak Centering does not hold for deontic modality. Weak Centering is that the world \( i \) itself is at least as good as deontic ideal worlds as any other possible worlds. Weak Centering is an inappropriate assumption for deontic modality, because our world, \( i \), is not always one of the deontic ideal worlds. For example, killing is morally forbidden, but it still happens at \( i \); in that case, “people ought not to kill other people” holds at \( i \), but \( i \) itself is not a deontic ideal world. On the other hand, the dyadic version of deontic detachment holds. When \( p \) is true at all the deontic ideal worlds from \( i \), and all the deontic ideal \( p \)-worlds are \( q \)-worlds, \( q \) is also true at all the deontic ideal worlds from \( i \).

---
them. Unfortunately, Jones missed the very last train to get back to town. Jones is not going to help them.

We can find several interesting features of the CTD situation depicted by the Chisholm set from this example. Jones violates his duty to go to help his neighbors; that is, the situation where Jones ought to go to help them, but he does not \((O(h) \& \neg h)\). Given his duties, Jones ought not to tell them he is coming \((O(\neg t))\) because that is the best thing he can do in the situation. This is Jones’ new derived obligation, \((c5)\).

\[(c5) \text{ Jones ought not to tell them he is coming.}\]

All of his original duties, \((c1)-(c3)\), and \((c5)\) hold in this CTD situation, but there is no moral conflict for Jones. Thus, this CTD situation is supposed to be represented with no practical tension, let alone any logical tension. In other words, \((c1)-(c5)\) consistently hold in this CTD situation. From these features, we can generalize three requirements an adequate formal representation of the CTD situation has to meet as follows.

(I) The Actual Obligation Requirement
Given \((c3)\) and \((c4)\), what Jones is called upon to do in the CTD situation is not to tell them he is coming; thus, \((c5)\) holds in the CTD situation.

(II) The Violation Requirement
The CTD situation is represented as a violation context, in which Jones violates \((c1)\); thus, \((c1)\) and \((c4)\) hold in the CTD situation.

(III) The Compatibility Requirement
The set of statements holding in the CTD situation is represented without any tension—neither inconsistency nor practical tension.

However, the ordering semantics cannot meet all of these requirements, because (I), (II), and (III) together require a formal system to represent the following set without any tension.

\[
\begin{align*}
(c1) & \quad O(h/\top) \\
(c2) & \quad O(t/h) \\
(c3) & \quad O(\neg t/\neg h) \\
(c4) & \quad \neg h \\
(c5) & \quad O(\neg t)
\end{align*}
\]

This extended Chisholm set is consistently represented in neither SDL nor ordering semantics (with DDL). In ordering semantics, the derivability of \((c5)\) from \((c3)\) and \((c4)\) cannot be explained due to the invalidity of factual detachment. Even if somehow we got \((c5)\), “\(O(\neg t)\)” is incompatible with “\(O(t)\)” entailed by \((c1)\) and \((c2)\) in any system. Therefore, only two of the requirements can be met at the same time; if (I) and (II) are met, (III) cannot be met; to meet (III), either (I) or (II) has to be given up. I will call this new problem around the Chisholm set the CTD trilemma.
4 Two Faces of Normativity: Deontological and Axiological ‘Ought’s

Here’s my diagnosis of the CTD trilemma. In ordering semantics, the four ‘ought’-statements, (c1), (c2), (c3), and (c5), cannot be consistently represented, regardless of (c4). It is simply because what is ideal told by (c1) and (c2) together conflicts with what (c5) says. To avoid this conflict we need to distinguish a derived ‘ought’-statements like (c5) from the ‘ought’-statements primarily given like (c1)-(c3). However, most formal accounts like ordering semantics do not have resources to draw a non-arbitrary distinction between them.

I claim that the relevant, but non-arbitrary, distinction of ‘ought’-statements to solving the CTD trilemma can be drawn with respect to their functions in our normative discourse and different speech acts associated with them.

Normative ‘ought’-statements are used to impose or issue obligations and duties like commands and orders. Sometimes normative ‘ought’-statements are used to describe ideals, what is desirable, and what is the better thing to do relative to a given moral standard. The use of the former kind of ‘ought’-statements is to make the relevant agent to act accordingly, not just to inform her. So they are action-guiding like imperatives. However, the latter kind of ‘ought’-statements are basically descriptive telling or informing what is better than what. I will call the former deontological ‘ought’-statements and the latter axiological ‘ought’-statements.3

For example, consider the sentence, “you ought to go to bed at 9,” used in two different contexts. By uttering “you ought to go to bed at 9” Jane’s mom can command Jane to go to bed at 9. By uttering the same sentence, Jane’s sister can advise Jane when she’d better to go to bed to be a good kid without issuing such a command. While the former use of the ‘ought’-statement has its normative force directly, but the latter has it only indirectly since it describes evaluative facts.4 These two different uses can be considered as performing different speech acts.5 Thus, they have different directions of fit. Jane’s Mom’s utterance of the ‘ought’-statement have the world-to-word direction of fit like imperatives, while Jane’s sister’s utterance of the very same sentence has the word-to-world direction of fit like descriptives.6

3 I borrowed these terms from the deontology and axiology distinction. Roughly speaking, axiology is a theory of value and deontology is a theory of duties. Values and duties are two faces of morality as different sources of the normative force of normative statements.

4 Suppose that Jane’s sister said that because she knows that good kids listen to their parents and that Jane’s mom has ordered Jane to go to bed at 9.

5 Directive and assertive, respectively, according to Searle(17)’s taxonomy.

6 Ninan((16)) provides some linguistic evidence that deontic modal statements with ‘must’ sometimes behaves like imperatives. He concludes that ‘must’-statements have the tendency of being read as having the imperative like force unlike other deontic modal statements. I think that observation does not seem to be exclusive to ‘must’. Most deontic modal statements can be read as having imperative-like force if contexts
With this distinction, let’s reconsider the Chisholm set again. When the ‘ought’-statements of the Chisholm set are taken as axiological statements, there is no violation even when Jones does not go to help his neighbors ((c4)). For (c1) “Jones ought to go to help” as an axiological statement merely describes what is better thing to do for Jones, not generating Jones’ obligation. Also, (c1) as an axiological statement is not true any more in the CTD situation where (c4) holds; what is better thing to do for Jones in the CTD situation is not for him to go to help them, because it is a settled matter that Jones does not go. Thus, this interpretation cannot explain the violation of (c1), how come (c1) holds when (c4) is true and why we still think that somehow (c1) holds when (c4) is true and Jones ought not to tell them he is coming as (c5) says.

The key to solving the CTD trilemma is to read (c1)-(c3) as deontological ‘ought’-statements and (c5) as axiological and to explain how to get (c5) from (c3)-(c4) despite their difference in kinds. In other words, it has to be explained how the normativity of (c5) as an axiological statement piggybacks on the normativity of (c1)-(c3) despite their different kinds.

In the next section, I will outline a formal account geared toward capturing this dynamics between different ‘ought’-statements to solve the CTD trilemma.

5 A Sketch of a Dynamic Approach Based on Kratzer Semantics

The approach I will sketch here is a dynamic one based on Kratzer semantics featuring the two aspects of normativity of ‘ought’-statements. The general idea of dynamic approach is to take the semantic value of a sentence as an operation on a body of information, or an instruction for updating a context. ([10]), [7], [5]), [18]). I will propose a dynamic account with structured informations states, that is, information states with a preference ordering. So it will be represented by a set of ordered possible worlds determined by a modal base and an ordering source.

Before introducing this structured information state and two kinds of ‘ought’-statements we have defined earlier, let me sketch a simple dynamic system. In this dynamic semantics, a model $M$ for a language $L$ is a pair $< W, \llbracket \ast \rrbracket >$, where $W$ is a set of ordered possible worlds, and $\llbracket \ast \rrbracket$ is a valuation function mapping the propositional letters of $L$ to sets of worlds. An update(dynamic) system is a triple $< L, \Sigma, [\ast] >$, where $L$ is a language, $\Sigma$ is a set of relevant information state, $\sigma$, which is any subset of $W$, and $[\ast]$ is an update function. For any $M$, an update function $[\ast]$ is a function from wff of $L$ to functions from information states to information states as follows.

---

properly support them. I think those two possible readings are parallel with the deontological and axiological statement distinction.

7 This approach is different from other attempts to solve Chisholm’s Paradox by introducing different notions of obligations like ideal and actual obligations such as Jones&Pörn([9]) and Carmo&Jones([1]). Given my distinction, their notions of ideal and actual obligations both fall into the category of axiological ‘ought’-statements.
\[\sigma[p] = \sigma \cup \llbracket p \rrbracket\]
\[\sigma[\neg \phi] = \sigma - \sigma[\phi]\]
\[\sigma[\phi \land \psi] = \sigma[\phi][\psi]\]
\[\sigma[\phi \rightarrow \psi] = \{ w \in \sigma : \sigma[\phi] \neq \emptyset \}
\[\sigma[\phi \land \psi] = \{ w \in \sigma : \sigma[\phi][\psi] = \sigma[\phi] \}\]

Here “\(\sigma[\phi]\)” is a result state of updating \(p\) to \(\sigma\). Updating \(p\) to the information state \(\sigma\) has the effect of ruling out all \(\neg p\)-worlds from \(\sigma\). Notice that the functions of epistemic modal statements and conditionals behave like a test; functions like \([\sigma, \phi]\) and \([\phi \rightarrow \psi]\) return either \(\sigma\) or \(\emptyset\). Updating \(\sigma\) with \(\sigma, p\) returns \(\sigma\), if there is at least one \(p\)-world in \(\sigma\); otherwise, \(\emptyset\). Updating \(\sigma\) with \(p \rightarrow q\) returns \(c\), if all \(p\)-worlds are \(q\)-worlds in \(\sigma\); otherwise, \(\emptyset\).

**Dynamic Entailment**

\[\phi_1, \phi_2, \ldots, \phi_n \models \psi \iff \forall \sigma \in \Sigma : \sigma[\phi_1][\phi_2][\ldots][\phi_n] \models \psi.\]

To incorporate two features of normativity of ‘ought’-statements, we need structured information states. The difference between structured information states I am proposing and original information states lies in whether the set of possible worlds representing an information states are ordered or not. Remind that Kratzer semantics has two contextual components: a modal base and an ordering source. Structured information states, \(\sigma(i)\), can be understood as a set of ordered possible worlds defined by \(< f(i), g(i) >\).

First of all, updating a structured information state with factual (declarative) statements amounts to updating a modal base. Since a modal base like \(f(i)\) is just a set of possible words, updating them is just like what we have defined above. The updated modal base \(f(i)\) with a set of propositions holding at \(i\), \(c(i)\), is defined as follows.

\[f^+(i) = \{ w : w \in f(i) \} - \{ w : w \not\in c(i) \} = \{ w : w \in f(i) \cup c(i) \}\].

It’s time to introduce two ‘ought’-statements into this dynamic account: a dyadic operator, “\(D(-/-)\),” for deontological ‘ought’-statements, and “\(A(-/-)\)” for axiological statements. Deontological statements generate duties and obligations; that means, they add new preference orderings. So updating a structured information state with a deontological statement amounts to updating an ordering source with it, and updating an ordering source \(g(i)\) with “\(D(q/p)\)” has the effect ordering \(p\&q\)-worlds over \(p\&q\)-worlds among the possible worlds in \(f(i)\). This update is done indirectly by adding “\(D(q/p)\)” to \(g(i)\), which is set of duties and obligations expressed by deontological statements as follows in this account.

For any \(u, v \in W\) and \(g(i), u \prec_{g(i)} v \iff \exists D(p/q) \in g(i): u, v \in [p] & u \in [q]
\& v \not\in [q] & \neg \exists D(r/s) \in g(i): u, v \in [s] & u \not\in [r] & v \not\in [r]\).\]

According to this definition, \(u\) is strictly better than \(v\) according to the deontological statements in \(g(i)\) iff there is at least one “\(D(q/p)\)” such that both \(u\) and \(v\) are \(p\)-worlds but only \(u\) is a \(q\)-world, and there is no “\(D(r/s)\)” such that
both $u$ and $v$ are $s$-worlds but only $v$ is a $r$-world. Two possible worlds, $u$ and $v$, are incomparable just in case there is no duties or obligations comparing $u$ and $v$ at all, or there are conflicting duties, some ranking $u$ above $v$ and others ranking $v$ above $u$.

Axiological statements describe what is better than what; that is, they are about preference orderings of possible worlds; so they are more similar to conditionals or epistemic possibility modal statements in a dynamic semantics. So updating a structured information state with them amounts to testing whether the structured information is compatible with them.

$$
\sigma(i)[A(q/p)] = \begin{cases} 
\sigma(i), & \text{if } \{w : w \in \cap f^+(i) \land \neg \exists v \in \cap f^+(i) : v \lessdot_{g(i)} w\} \subseteq \llbracket q \rrbracket; \\
\emptyset, & \text{otherwise.}
\end{cases}
$$

Axiological ‘ought’-statements describe either the preference ordering of possible worlds in $f(i)$ of $\sigma$ or that of possible worlds in $f^+(i)$ of $\sigma(i)[c(i)]$ depending on what they focus on. Note that $f^+(i) \subseteq f(i)$. The axiological ‘ought’-statements on $f(i)$ describes the most ideal situation relevant to $i$ regardless of what is happening in $i$, while the axiological ‘ought’-statements on $f^+(i)$ describes the best situation accessible from $i$ given the facts of $i$. So $[A(q/p)]_{i \in M,f,g}^i$ describes the ideality, whereas $[A(q/T)]_{i \in M,f,g}^i$ describes the best accessible possibility. Their static values are defined as follows.

For $M = (W, >_w, [\ast])$ and a set of true propositions at $i, c(i)$,

$$
[A(q/p)]_{i \in M,f,g}^i = 1 \text{ iff } \{w : w \in \cap f^+(i) \land \neg \exists v \in \cap f^+(i) : v \lessdot_{g(i)} w\} \subseteq \llbracket q \rrbracket,
$$

provided that $\cap f^+(i) = \{w : w \in \cap f(i) \text{ and } w \in \llbracket p \rrbracket\}$;

$$
[A(q/T)]_{i \in M,f,g}^i = 1 \text{ iff } \{w : w \in \cap f^+(i) \land \neg \exists v \in \cap f^+(i) : v \lessdot_{g(i)} w\} \subseteq \llbracket q \rrbracket,
$$

provided that $f^+(i) = \{w : w \in \cap (f(i) \cup c(i))\}$.

Informally, the former tells a sort of what ideally ought to be, while the latter tells what actually ought to be.\(^8\) The truth value of $[A(q/p)]_{j \in M,f,g}^i$ is invariant throughout all $j \in f(i)$, while the truth value of $[A(p \top)]_{j \in f^+(i)}$ is sensitive to $c(j)$. To sum up, here are three main ideas of the dynamic account based on Kratzer semantics: (i) an information state in this dynamic account is structured in that it is represented by a pair of a modal base and an ordering, $< f(i), g(i) >$; (ii) updating deontological ‘ought’-statements to a structured information state, $< f(i), g(i) >$, brings about a change of $g(i)$, while axiological statements behave like a test on $< f(i), g(i) >$; and (iii) the static values of axiological statements describe different idealities depending on modal bases—either $f(i)$ or $f^+(i)$. This might sound a bit complicated, but the general idea of this approach can be easily illustrated with diagrams. Here is how diagrams represent the dynamics of three kinds of ‘ought’-statements. In the diagrams, $S$ is a set of relevant possible worlds to the actual world $i$ determined by $f(i)$, which represent the structured information state $\sigma(i)$; circles in $S$ represents non-empty sets of

\(^8\) The latter is unconditional always in a natural language, because all relevant facts in $i$ has already been taken into account in $f^+(i)$.
possible worlds that satisfy the propositions written in the circles; \( g(i) \), a set of duties and obligations expressed by deontological statements, connects the circles in \( S \) with an arrow such that the circle (the set of worlds) on the left is strictly better than the circle on the right; and worlds in a circle are equally good. Unconnected worlds are incomparable with each other. The preference relation represented by an arrow is transitive. \( f^+(i) \), which is a subset of \( f(i) \), puts a dashed line box around the possible worlds compatible with all the factual statements in \( c(i) \) when \( c(i) \) is not an empty set.\(^9\) \( \llbracket A(p/q) \rrbracket_{f^+} \) holds for \( S \) iff the left most possible worlds among \( q \)-worlds in \( S \) are \( p \)-worlds. \( \llbracket A(p/T) \rrbracket_{f^+} \) holds for \( S \) iff the left most possible worlds in a dashed line box in \( S \) are \( p \)-worlds.

Now let's see how deontological statements in \( g(i) \) order possible worlds. Suppose a set of possible worlds with no salient ordering over them, which will look like \( S_0 \).

\[ p \lor \neg p \]

**Fig. 1.** \( S_0 \)

“\( D(p/T) \)” orders possible worlds in \( S \) by connecting \( p \)-worlds and \( \neg p \)-worlds and putting all \( p \)-worlds on the left. A set of possible worlds, \( S_0 \), ordered by a set of a single norm, \( g_1(i) = \{D(p/T)\} \), will look like \( S_1 \).

\[ p \leftrightarrow \neg p \]

**Fig. 2.** \( S_1 = S_0[D(p/T)] \)

\( S_1 \) is equivalent to ordering \( S_0 \) with \( \{D(p/T)\} \). For \( S_1 \), \( \llbracket A(p/T) \rrbracket_{f^+} \) and \( \llbracket A(p/T) \rrbracket_{f^+} \) hold. Next, when \( q \) is not a tautology, \( g_2(i) = \{D(q/p)\} \) connects the \( p \land q \)-worlds and \( p \land \neg q \)-worlds, puts the \( p \land q \)-worlds on the left, and leaves the rest of the worlds unconnected as follows.

\( S_2 \) is ordered by \( g_2(i) = \{D(q/p)\} \). Both \( \llbracket A(q \land p/T) \rrbracket_{f^+} \) and \( \llbracket A(q \land p/T) \rrbracket_{f^+} \) hold of \( S_2 \). A conditional obligation induces a partial ordering. It only takes

\(^9\) When \( c(i) \) is an empty set, that is, \( f(i) = f(i)^+ \), a dashed line box is omitted and \( \llbracket A(p/q) \rrbracket_i^{f^+} = [A(p/q)]_i^{f^+} \).
effect on \(p\)-worlds, and ranks \(q\)-worlds higher than \(\neg q\)-worlds among them. So this ordering is indifferent about \(\neg p \land q\)-worlds and \(\neg p \land \neg q\)-worlds; it means that \(\neg p \land q\)-worlds and \(\neg p \land \neg q\)-worlds are incomparable with other worlds and with each other in \(S_2\). So a non-trivial conditional obligation induces a partial ordering. \(S\) is partially connected by \(g(i)\) iff there are incomparable possible worlds left in \(S\) given the ordering by \(g(i)\).

Without factual statements an actual obligations over a tautology is consonant with an ideal obligation over a tautology. Let’s see how we can take factual statements into the picture. Suppose \(S_3\) ordered by \(g_3(i) = \{D(p), D(q/\neg p)\}\) and \(c_3(i) = \{\neg p\}\).

\[f^{+c_3}(i)\] puts the box around \(\neg p\)-worlds. Now \(\llbracket A(p/T)\rrbracket_i^{f,g}, \llbracket A(q/\neg p)\rrbracket_i^{f,g}\), and \(\llbracket A(q/T)\rrbracket_i^{f^\neg,g}\) hold for \(S_3\). What actually ought to do is determined by the leftmost circle in the box around \(\neg p\)-worlds. These diagrams illustrate how deontological statements order possible worlds, and what kind of axiological statements hold in the given ordered possible worlds.

6 The Chisholm Set in the Dynamic Kratzer Semantics

In Sec. 4, I have pointed out that the CTD trilemma arises when the Chisholm set is read as consisting of three deontological statements and one factual statement. So the Chisholm set gives us \(g_5(i) = \{D(h/T), D(t/h), D(\neg t/\neg h)\}\) and \(c_5(i) = \{\neg h\}\). \(S_5\), defined by \(f_5(i)\), ordered by \(g_5(i)\) and boxed by \(f^{+c_5}(i)\), is illustrated as follows.

\(S_5\) is completely connected by \(g_5(i)\). By taking the factual statement, “\(\neg h\),” into account, we narrow down our focus to the subset of the ordered possible worlds, \(\neg h\)-worlds – those in the blue box in \(S_5\). So \(S_5\) represents the CTD situa-
Fig. 5. $S_5 = S_0 [D(h/\top)] [D(t/h)] [D(\neg t/\neg h)] [\neg h]$

tion of the Chisholm set. $\llbracket A(h/\top) \rrbracket_{i}^{f,g}$, $\llbracket A(h \land t) \rrbracket_{i}^{f,g}$, $\llbracket A(t/h) \rrbracket_{i}^{f,g}$, $\llbracket A(\neg t/\neg h) \rrbracket_{i}^{f,g}$, and $\llbracket A(\neg t/\top) \rrbracket_{i}^{f^+g}$ hold for $S_5$ given $f_5(i)$, $g_5(i)$, and $c_5(i)$.

First, this results meet Requirement (I), since $\llbracket A(\neg t/\top) \rrbracket_{i}^{f,g}$ tells that what actually Jones ought to do given the fact that he does not go is not to tell them he is coming.

Also, it meets Requirement (II). The violation of the norm, (c1) "D(h/\top)," is represented by $\llbracket A(h/\top) \rrbracket_{i}^{f,g} \land \llbracket \neg h \rrbracket_{i}^{f,g}$ of $S_5$.

Last, this result meets Requirement (III). All corresponding axiological and factual statements to (c1) through (c5) hold of $f_5(i)$, $g_5(i)$, and $c_5(i)$: $\llbracket A(h/\top) \rrbracket_{i}^{f,g}$, $\llbracket A(t/h) \rrbracket_{i}^{f,g}$, $\llbracket A(\neg t/\neg h) \rrbracket_{i}^{f,g}$, $\llbracket \neg h \rrbracket_{i}^{f,g}$, and $\llbracket A(\neg t) \rrbracket_{i}^{f^+g}$. However, there is no tension among them, because $\llbracket A(t/\top) \rrbracket_{i}^{f,g}$ and $\llbracket A(\neg t) \rrbracket_{i}^{f^+g}$ describe different idealities in different ranges of the chain.

So far I have sketched a dynamic approach to the semantics of deontic ‘ought’s with the key notions of Kratzer semantics. This dynamic account is not only motivated by two kinds of normativity in a natural language, but also required for resolving the CTD trilemma, generated by the Chisholm set. Also, this account provides us with the story about how two context dependent parameters, $f(i)$ and $g(i)$, interact with and determined by a context, which original Kratzer semantics lacks. Although many details of this account have not been fully stated in this short paper, I believe the proposed dynamic approach with structured information states is the first step toward a more rigorous dynamic system for deontic ‘ought’s integrating the orthodoxy semantics, Kratzer semantics, and dynamic approaches.

Acknowledgement

The author would like to thank the anonymous referees of ESSLLI Student Session for their useful comments and suggestions, as well as the members of the dissertation working group of Cornell Philosophy Department, Lucia Maria Munguia, and Will Starr for their helpful comments and supports.

References

Durative Phrases in Constructions Containing Double Identical Verbs in Mandarin Chinese⋆

Xiaoqian Zhang

Université Paris Diderot-Paris 7, Laboratoire de Linguistique Formelle

Abstract. This paper aims to study the semantics of durative phrases in verb-copying constructions and coordination constructions in Mandarin Chinese. We claim that the existing proposals concerning their semantics in simple sentences are not sufficient to account for their behaviour in these two constructions and that the operation of event identification must be introduced.

1 Introduction

Post-verbal durative phrases in Mandarin Chinese are taken to be quantificational phrases ([5]) or measure phrases ([12]) in the literature because they are non-referential and cannot be preceded by a demonstrative such as na. This is shown by (1)1.

(1) Ta pro3sg lai-le (*na) san tian le. ([12])
   PRO3SG come-PFV DEM three PTCL
   ‘It has been three days since he came.’

Besides simple sentences (henceforth canonical sentences) that contain one VP, as illustrated in (1), durative phrases can also occur in two constructions that contain two identical verbs. The first type is called verb-copying construction in the literature. According to [8], verb-copying constructions refer to a grammatical process in which a verb is “copied” after its direct object when in the presence of certain adverbial elements, such as durative phrases. As is pointed out in the literature ([5] among others), the verb in the first verb constituent (VP1) in this construction does not take any aspectual marker. For example, the accomplishment verb hua ‘draw’ in VP1 is bare in (2). By contrast, the second VP (VP2) contains the same verb modified by the perfective marker verbal -le and a durative phrase liang-ge xingqi ‘two weeks’.

⋆ I would like to thank three anonymous reviewers for helpful comments. Special thanks go to my dissertation advisor Lucia Tovena for numerous discussions and also for comments and suggestions on three earlier versions of this paper.

1 Abbreviations used throughout in the glosses: clf=classifier; de=resultative particle; dem=demonstrative; dur=durative; neg=negation; pfv=perfective; pro=pronoun; prog=progressive; ptcl=particle; sg=singular
Lisi drew four paintings for/in two weeks.

The second type of constructions refer to sentences where both occurrences of the VP are aspectually marked. Consider (3) for instance.

(3) Lisi hua-le si-fu hua, hua-le liang-ge xingqi.
Lisi draw-PFV four-CLF painting draw-PFV two-CLF week
‘Lisi drew four paintings in two weeks.’

The verb hua ‘draw’ is modified by the perfective marker verbal -le in both VP1 and VP2. Furthermore, notice that the first occurrence of the verb can be modified by other aspectual markers. In (4), hua ‘draw’ is modified by the progressive marker zai. We will call this construction coordination construction, following [3].

(4) Lisi zai hua si-fu hua, hua-le liang-ge xingqi.
Lisi PROC draw four-CLF painting draw-PFV two-CLF week
‘Lisi was drawing four paintings. He drew for two weeks.’
‘He finished them in two weeks.’

While the semantics of durative phrases in canonical sentences is largely studied in the literature ([11], [13]), less attention has been paid to their semantics in constructions containing two identical verbs, probably because it is taken for granted that the meaning of durative phrases in canonical sentences can easily extend to constructions containing two identical verbs, cf. [10], [12] and [17].

In this paper, we will argue that the semantics of durative phrases in canonical sentences is not sufficient to account for their behaviour in verb-copying constructions and coordination constructions. An event identification operation is needed in order to explain their distribution. The paper is organized as follows. In Sect.2, we review two main proposals for durative phrases in the literature and point out the problems that arise when they are applied to verb-copying constructions and coordination constructions. In Sect.3, we take into account more data concerning durative phrases in these two constructions and present their semantics. A conclusion will be drawn in Sect.4.

2 Previous Studies

There are mainly two proposals in the literature to account for the semantic behaviour of durative phrases, namely the homogeneity analysis and the ambiguity analysis.

2.1 Homogeneity Approach

The homogeneity approach is due to [10], [11], [12] and [13], who claim that durative phrases modify only homogeneous situations. For example, it is shown
in (5) that liang-ge xingqi ‘two weeks’ is fine with an atelic situation of waiting for that person, but it cannot measure the running time of a telic situation such as drawing four paintings, e.g. (6).

(5) Lisi deng-le na-ge ren liang-ge xingqi.
    Lisi wait-PFV DEM-CLF person two-CLF week
    ‘Lisi waited for that person for two weeks.’

(6) *Lisi hua-le si-fu hua liang-ge xingqi.
    Lisi draw-PFV four-CLF painting two-CLF week

[11] defines the meaning of a durative NP as in (7):

(7) [[Durative NP]] = \lambda R.e/s[R(e/s)\&\text{MU}(\tau(e/s))=\text{Num}\&\text{Hom}(e/s)]

where “e” is a variable ranging over events, “s” is a variable ranging over states, \(\tau\) is a function when applied to an eventuality gives its temporal trace. MU is a measure unit function that yields a number when applied to a temporal trace and Hom is a predicate representing homogeneity. ‘e/s’ means either a variable of “e” or variable “s”. As for the notation ‘R’, it is intended to stand for a variable with a flexible semantic type that involves a sequence of one or more eventuality argument (cf. [14]).”

In what follows, we will examine whether the homogeneity requirement of durative phrases of canonical sentences can be applied to verb-copying and coordination constructions.

If we restrict our attention to the verb constituent where durative phrases occur, the homogeneity claim is plausible since durative phrases modify only homogeneous situations. Specifically, the stative and activity verbs denote a homogeneous eventuality in (8) and (9) respectively.

(8) Lisi xihuan na-ge nvhai xihuan-le shi nian.
    Lisi like DEM-CLF girl like-PFV ten year
    ‘Lisi liked that girl for ten years.’

(9) Ta kan dianhua, zuzu kan-le wu fenzhong.
    PRO3SG look at telephone fully look at-PFV five minute
    ‘She looked at the telephone for a good five minutes.’

As to the accomplishment verb hua ‘draw’ in (2)-(4), it expresses a homogeneous event as well in the absence of an incremental theme.

Sentence (10) indicates that when the achievement verb ting ‘suspend’ is present, the durative phrase liang-ge xiaoshi ‘two hours’ cannot go with it and specify the time that it takes to suspend the classes. But it asserts the duration of the resultant state of suspending the classes. The contrast presents new evidence that durative phrases must modify atelic situations. If the verb is telic, a homogeneous interpretation must be coerced.

(10) Xuexiao ting ke ting-le liang tian.
    school suspend class suspend-PFV two day
‘#The school suspended the classes in two days.’
‘The school suspended the classes for two days.’

However, if we consider the whole construction, we notice that liang-ge xingqi ‘two weeks’ in (2) can measure the duration of the telic situation of drawing four paintings. As a result, we can continue (2) by (11).

(11) Zhongyu hua-wan-le.
finally draw-finish-pfv
‘He finished them finally.’

In addition to the completive reading, the accomplishment can also be coerced into a partitive reading and we can thus continue (2) by (12).

(12) Mei hua-wan.
NEG draw-finish
‘But he didn’t finish them.’

In a word, while the homogeneity analysis can account for the behaviour of durative phrases when combined with eventualities of states, activities and achievements, it cannot predict the completive reading triggered by the combination with the accomplishment event, illustrated by (2).

2.2 Ambiguity Approach

[17] assumes that the semantics of durative phrases is ambiguous: it can either measure an atomic event ([15]) in (13-a) or the sub-event of an atomic event in (14-a). Q is a predicate of events, e is an event variable and \( \mu \) is the measure function.

(13) a. Ta du ≪Hongloumeng≫ du-le ban nian, gang
PRO3SG read Dream of the Red Chamber read-pfv half year just
du-wan. ([17])
read-finish
‘He read Dream of the Red Chamber in half a year and he just finished it.’
b. \( \exists e[\text{Reading}(e, \text{He, Dream of the Red Chamber})&\text{ATOM}(e)&\mu_{\text{year}}(e)=\text{half}&\text{Past}(e)] \)

(14) a. Ta du ≪Hongloumeng≫ du-le ban nian, meiyou
PRO3SG read Dream of the Red Chamber read-pfv half year NEG
du-wan, jiulang-ge xingqi. (modified from [17])
read-finish then give up-pfv
‘He read Dream of the Red Chamber for half a year. But he didn’t finish it. Then he gave it up.’
b. \( \exists e'[\text{Reading}(e, \text{He, Dream of the Red Chamber})&\text{PART}(e', e)&\mu_{\text{year}}(e')=\text{half}&\text{Past}(e')] \)

Although the ambiguity proposal provides appropriate accounts for the ambiguous readings triggered by sentence (2), it is too strong in that it fails to
explain why the ambiguity is impossible in most cases. First of all, notice that in canonical sentences, the durative phrase cannot specify the duration of telic situations, cf. (6). Second, the partitive reading is infelicitous in some coordination constructions, e.g. (3). Lastly, the durative phrase in sentence (10) only specifies the duration of the resultant state ensuing from suspending the classes, but not the time that it takes to succeed in suspending the classes.

3 Durative Phrases in Non-canonical Constructions

In the previous section, we have pointed out that neither the homogeneity proposal nor the ambiguity proposal can directly extend to the semantics of durative phrases in verb-copying constructions and coordination constructions. In order to provide the semantics of durative phrases in these non-canonical constructions, we must take into consideration two aspects.

Firstly, we must explain the different manifestations of durative phrases in canonical constructions and non-canonical constructions, i.e. durative phrases modify only homogeneous situations in canonical sentences, whereas they can sometimes specify the duration of accomplishment situations in non-canonical constructions. We will deal with this aspect in Sect. 3.1 and Sect. 3.2.

Secondly, our proposal must account for the different behaviours of durative phrases within non-canonical constructions. More precisely, we must explain why and how the readings and distributions of durative phrases vary according to different aspectual situation types, different aspectual grammatical markers and different constructions. We will detail this point in Sect. 3.3.

3.1 Event Identification in Non-canonical Constructions

As we have learned above, under most circumstances, durative phrases display a similar behaviour in canonical sentences and non-canonical constructions, i.e. they are subject to the homogeneity restriction. Specifically, in canonical sentences, they are only compatible with predicates expressing homogeneous situations. Likewise, in non-canonical constructions, the verb constituents that contain them must denote homogeneous situations. There is only one exception, which is that when the whole non-canonical construction is considered, durative phrases can often specify the duration of accomplishment situations, because VP1 can contain a quantized object, for example an incremental theme. In order not to blemish the similarity existing between the canonical sentences and non-canonical constructions, we will adopt the viewpoint that durative phrases are only compatible with homogeneous eventualities. Then what is left to discuss is how durative phrases specify the duration of accomplishment situations in non-canonical constructions.

The proposal is as follows. Although each VP in non-canonical sentences expresses separately an eventuality, it is not true that the two eventualities denoted are not related to one another. Our intuition is that the VPs provide two different partial event descriptions of the same event that takes place once
in the real world. In other terms, there is a step of event identification that must be assumed between the eventuality denoted by one constituent and the temporally bounded eventuality denoted by the other constituent containing the durative phrase. Because of the event identification, the running time of the eventuality corresponding to the first event description must be identical to that of the second event description. It is in this way that durative phrases specify the duration of the event denoted by VP2 that contains it and VP1 where the durative phrase is absent. Following the proposals on predicate modification ([4]) and Event Identification ([6], [7]), we claim that the composition rule (15) can be put to a new use and allows us to combine the two eventualities denoted by the two VPs.

(15) If α is a branching node, \{ β, γ \} is the set of α’s daughters, and [[β]] and [[γ]] are both in D_{<e,<v,t>},
then [[α]] = λx ∈ D_e . λe ∈ D_v . [[β]](e)(x)=[[γ]](e)(x) = 1

In the combination rule, x is a variable of type e ranging over individuals, e is a variable of type v ranging over eventualities (states and events). Applying (15) to “draw four paintings draw two weeks”, we get (16).

(16) [[draw four paintings draw two weeks]]
= λx ∈ D_e . λe ∈ D_v . [[draw four paintings]](e)(x)=[[draw two weeks]](e)(x) = 1
= λx ∈ D_e . λe ∈ D_v . x draws four paintings and x draws two weeks

The application of event identification is subject to some constraints. First, the identification is restricted to cases where the verbs in both VPs are the same. For example, although chou and xi both mean smoke in Mandarin Chinese, they cannot occur in VP1 and VP2 respectively in a non-canonical construction. Contrast (17-a) with (17-b).

   Lisi smoke-PFV three-CLF cigarette smoke-PFV twenty minute
b. Lisi chou-le san-gen yan, chou-le ershi fenzhong.
   Lisi smoke-PFV three-CLF cigarette smoke-PFV twenty minute
   ‘Lisi smoked three cigarettes in twenty minutes.’

Second, there must be compatibility between the two partial descriptions of events. In (18), VP1 expresses a plural event that is composed of sub-events of running one hour. The duration of the plural event, namely the sum of the duration of each sub-event must be longer than one hour. This is contradictory with the duration specified by the durative phrase in VP2. Therefore, it follows that event identification does not work.

(18) #Lisi mei tian pao yi-ge xiaoshi pao-le yi-ge xiaoshi.
   Lisi every day run one-CLF hour run-PFV one hour
Third, the description provided by VP2 must supplement the description provided by VP1 by more information, cf. (19-a) versus (19-b). In (19-a), VP1 describes a drawing event and VP2 adds one new piece of information concerning it, which is its duration. With respect to (19-b), VP1 provides the information that there is a drawing event and it lasts two weeks, whereas VP2 only repeats what is expressed by VP1 without adding any new information. The lack of new supplement of information makes the sentence unacceptable.

(19) a. Lisi hua hua hua-le liang-ge xingqi.
   Lisi draw painting draw-PFV two-CLF week
   ‘Lisi drew for two weeks.’

   b. *Lisi hua liang-ge xingqi hua-le hua.
   Lisi draw two-CLF week draw-PFV painting

Furthermore, we notice that event identification also applies to the construction containing more than two identical verbs, as is shown in (20). In this case, the identification operation takes place among three partial event descriptions, namely running three kilometres, running half an hour and running himself tired. The composition rule proposed in (15) can also cover this case in obeying some binary-branching hierarchy among the three partial descriptions.

(20) Lisi pao-(le) san qian mi, pao-le ban xiaoshi, pao de hen lei.
   Lisi run-PFV three thousand meter run-PFV half hour run DE very tired
   ‘Lisi ran three kilometers in half an hour. As a result, he became tired.’

3.2 A Topic-Comment Structure

In the spirit of [1], we propose that a non-canonical construction containing durative phrases is a topic-comment sequence. Specifically, the subject introduces a topic and all the verb phrases that follow it comment on this topic. Because of event identification, all the verbal predicates co-specify a single identified eventuality and anaphorically refer to the topic by means of this identified eventuality.

One question to ask is why do we need to “repeat” the verb to co-specify a single eventuality, if the canonical sentence that contains one VP already describes a single event, e.g. (5).

Putting aside the syntactic motivations, there are some aspectual reasons. First, it is sometimes the lexical aspect that requires the existence of two verb constituents. It has been noticed that in (5) which contains the activity verb deng ‘wait’, the object na-ge ren ‘that person’ and the durative phrase liang-ge xingqi ‘two weeks’ can co-exist in one VP. On the contrary, in (6) where there

2 The non-canonical constructions are also possible, as illustrated by (i).

(i) Lisi deng(-le) na-ge ren deng-le liang-ge xingqi.
   Lisi wait-PFV DEM-CLF person wait-PFV two-CLF week
   ‘Lisi waited for that person for two weeks.’
is the accomplishment verb *hua* ‘draw’, the object *si-fu hua* ‘four paintings’ and the durative phrase *liang-ge xingqi* ‘two weeks’ are incompatible with each other.

Let us compare (5) with (6). In (5), between the object *na-ge ren* ‘that person’ and the durative phrase *liang-ge xingqi* ‘two weeks’, only the durative phrase serves as a situation delimiter, since *deng* ‘wait’ is an atelic verb and the object is a non-incremental theme. By contrast, *hua* ‘draw’ in (6) is telic. Therefore, both the quantized object *si-fu hua* ‘four paintings’ and the durative phrase *liang-ge xingqi* ‘two weeks’ play the role of bounding the event of drawing. In order to avoid that two situation delimiters co-exist in one VP, two VPs must be introduced, as shown in (2) and (3).

A second possible reason is that different aspectual viewpoints also affect the choice of one verb constituent or two verb constituents. If we compare the canonical sentence (21) and the coordination construction (22), we see that both sentences express that there is a temporally bounded event of looking at the telephone for five minutes. However, the coordination construction is more informative than the canonical sentence because of the existence of two verb constituents with different grammatical aspectual markers: while VP1 denotes the event of looking at the telephone which is viewed as being ongoing due to the durative marker -*zhe*, the event of looking at the telephone described by VP2 is viewed as terminated with verbal -*le* being there. The coordination construction makes it possible that a single event that takes place in the real world is partially described by two different viewpoints, while it is impossible for canonical sentences to provide two partial descriptions of the same event described from different viewpoints.

(21) Ta zuzu kan-le dianhua wu fenzhong.
    PRO3SG fully look-at-PFV telephone five minute
    ‘She looked at the telephone for a good five minutes.’

(22) Ta kan-zhe dianhua, zuzu kan-le wu fenzhong.
    PRO3SG look-at-DUR telephone fully look-at-PFV five minute
    ‘She was looking at the telephone. She looked at it for a good five minutes.’

### 3.3 Interpretations and Distributions of Durative Phrases in Verb-copying and Coordination Constructions

In this sub-section, we will provide an account for why durative phrases display different interpretations and distributions in verb-copying constructions and coordination constructions.

Firstly, if we consider the accomplishment verb, it is observed that the verb-copying construction is always ambiguous between a partitive reading and a completive reading, as is illustrated in (2). As for the coordination construction containing the same accomplishment predicate, whether the sentence is ambiguous or not depends on the grammatical markers being used. Specifically, when

---

The sentence is taken from the Center for Chinese Linguistics PKU online corpus: ccl.pku.edu.cn:8080/ccl_corpus/.
the perfective marker verbal -le modifies the verb in (3), only the completive reading is available. However, in the presence of the progressive marker zai in (4), both readings are possible.

We assume that the various readings of durative phrases in non-canonical constructions can be treated from the point of view of aspect. As Mandarin Chinese is an aspect-prominent language, an aspectually bare sentence can sometimes be ambiguous in several ways, cf. (23).

(23) Ta chou yan.  
PRO3SG smoke cigarette  
‘He is smoking.’ (progressive reading)  
‘He is going to smoke.’ (futurate reading)  
‘He smokes.’ (habitual reading)

Coming back to (2), as VP1 is not aspectually marked, the reading triggered can be either a progressive reading or a futurate reading. When the event identification step takes place, either a completive reading or a partitive reading obtains, depending on whether Lisi completely undertook the event of drawing four paintings.

The same reasoning applies to the coordination construction. Specifically, the perfective marker verbal -le in VP1 in (3) requires that the drawing-four-paintings event must be accomplished. As a result, the partitive reading is not felicitous. This is opposed to (4) where the progressive marker zai is present in VP1. The uncertainty of the degree of the completeness makes both the completive and the partitive readings available.

Secondly, we notice that different constructions also affect the distribution of durative phrases in VP1. (24-a) shows that the verb-copying construction whose VP1 contains the durative phrase yi-ge xiaoshi ‘one hour’ is very odd. On the contrary, the same durative phrase can enter into the coordination construction, as shown in (24-b).

   Lisi beat one-CLF hour beat-PFV ten-CLF telephone  
   ‘He finished them in one hour.’

b. Lisi da-le yi-ge xiaoshi da-le shi-ge dianhua.  
   Lisi beat-PFV one-CLF hour beat-PFV ten-CLF telephone  
   ‘Lisi made ten telephone calls in one hour.’

Besides, the same durative phrase yi-ge xiaoshi ‘one hour’ can easily occur in the second verb phrase of the verb-copying construction with both the completive reading and the partitive reading, as (25) illustrates.

   Lisi beat ten-CLF telephone beat-PFV one-CLF hour  
   a. Da-wan-le.  
      beat-finish-PFV  
      ‘He finished them in one hour.’
Comparing (24-a) with (25), the intuition we get is that while it is easy to imagine a scenario where Lisi has a project of making ten phone calls, it is much more difficult to imagine a situation where Lisi plans to call for one hour since the duration of making phone calls is not only determined by Lisi but also determined by the persons that Lisi calls. That is why in the absence of the grammatical marker, *da shi-ge dianhua* ‘make ten phone calls’ can occur in the VP1 more easily than *da yi-ge xiaoshi* ‘call one hour’ with a futurate interpretation.

With regards to the coordination construction in (24-b), the addition of the verbal -le in the VP1 makes clear that making phone calls for one hour is not a project but a temporally bounded event that is completed. The durative phrase can thus easily appear in the VP1.

Lastly, let us discuss cases concerning resultative verbal compounds (RVCs). RVCs in Mandarin Chinese are composed by two verbs, whereby there is a causal relation between the eventuality denoted by the first verb and that denoted by the second verb, cf. [9]. For instance, the RVC *xiu-hao* in (26) consists of the verb *xiu* describing the action of repairing that leads to the telic point of becoming “good” denoted by the second verb.

(26)  
\[
\text{Ta xiu-hao-le shuiguan.} \\
\text{PRO3SG repair-good-PFV pipe} \\
\text{‘He repaired the pipe.’}
\]

It is pointed out by [2] and [12] that all RVCs in Mandarin Chinese are incompatible with durative phrases in verb-copying constructions, cf. (27).

(27)  
\[
\text{*Ta xiu-hao shuiguan yijing xiu-hao-le yi-ge zhongtou le.} \\
\text{PRO3SG repair-good pipe already repair-good-PFV one-CLF hour PTCL} \\
\text{(modified from [12])}
\]

However, we observe that the same RVC can appear in the coordination construction. Consider (28).

(28)  
\[
\text{Ta xiu-hao-le shuiguan, jijing xiu-hao-le yi-ge zhongtou le.} \\
\text{PRO3SG repair-good-PFV pipe already repair-good-PFV one-CLF hour PTCL} \\
\text{‘It has been an hour since he repaired the pipe.’}
\]

Unlike its English counterpart, the verb *xiu* in Mandarin Chinese does not encode a natural endpoint.
We have two questions to ask. Firstly, why cannot the RVCs be coerced into a preparatory process or a resultant state, i.e. a homogeneous eventuality, so as to be allowed in the verb-copying construction? In fact, it is in this way that the accomplishment verbs and some achievement verbs force a homogeneous reading, cf. (2) and (10). Secondly, why can the RVCs enter into the coordination construction by triggering a resultant state reading?

We claim that the homogeneity requirement is the answer to both questions. Specifically, RVCs per se denote an accomplishment situation that consists of a preparatory process and an associated outcome ([16]). Thus, in the verb-copying construction (27), both the VP1 and the VP2 describe an accomplishment of repairing the pipe. However, the semantics of the durative phrase requires that the verb it modifies must express a homogeneous situation. That is why the RVCs are ruled out in verb-copying constructions. By contrast, with regards to the coordination construction (28), as the VP1 is suffixed by the perfective marker verbal -le, it involves a priori a resultant state. The durative phrase in VP2 then modifies the resultant state instead of the accomplishment event.

The fact that RVCs in Mandarin Chinese denote an accomplishment event on their own, makes them different from simple accomplishment verbs on the one hand, and achievement verbs on the other hand, given that the former need to resort to incremental themes or paths to describe an accomplishment situation, whereas the latter encode only a telic point, but not the preparatory process. As a result, both accomplishment verbs and achievement verbs can be coerced into homogeneous situations more easily with respect to RVCs. This is why they are allowed in verb-copying constructions.

4 Conclusion

In this paper, we claim that the semantics of durative phrases in canonical sentences is not sufficient for explaining the behaviour of durative phrases in verb-copying constructions and coordination constructions. We propose that in these two constructions, an event identification step takes place between the eventualities denoted by the verb phrases. Specifically, the verb phrase that contains durative phrases always describes a temporally bounded eventuality. As to the other verb phrase, its denotation varies according to the aspectual nature of verbal predicates and the presence of different grammatical aspectual markers. Specifically, it can express either an atelic or a telic situation viewed as being static, as being in progress or as being terminated/completed. No matter how different the eventuality denoted by the VP1 is from the eventuality denoted by the VP2 from the perspective of aspect, the duration of their running time is always identical. That is why durative phrases in these two constructions can display different distributions and interpretations from their counterparts in canonical sentences.
References

17. Wang, Y.: Shi jian celiang he shiliang jiegou de you yi [event measurement and the semantics of temporal phrases]. Yuyan Jiaoxue yu Yanjiu 4, 66–74 (2013)
A Study of Semi-Supervised Spanish Verb Sense Disambiguation

Cristian Cardellino
Universidad Nacional de Córdoba, Argentina.
ccardellino@famaf.unc.edu.ar

Abstract. We present an exploration of semi-supervised techniques for Verb Sense Disambiguation (VSD) in Spanish. We have started from a small seed of manually annotated examples of verb senses in Spanish, from the SenSem corpus [1]. Then, we apply a semi-supervised bootstrapping technique to add automatically annotated examples to the starting manually annotated corpus. This enhanced corpus is then used to train a classifier for VSD. Augmenting the corpus by bootstrapping techniques improves the performance of the VSD classifier, but has a bias towards most frequent senses, losing coverage in least frequent ones. To gain representativity of least frequent senses, we integrate Active Learning techniques alongside bootstrap. We show that using the most simple Active Learning approach, uncertainty sampling, is useful in some verbs to gain coverage of least frequent senses while maintaining good overall performance.

Keywords: Verb Sense Disambiguation, Semi-supervised Word Sense Disambiguation, Bootstrap

1 Introduction and Motivation

Verb sense disambiguation (VSD) is a crucial task for deep language processing tasks, especially those that could benefit from information about relations between participants provided by subcategorization frames, like machine translation, question answering or information extraction. E.g., in machine translation sense disambiguation is needed to determine the correct translation of a word, and also the transformations needed in the syntactical structure of the sentence. Another example, for question answering over linked data VSD is needed to determine the sense of a verb and the logical relations between the participants to access the adequate nodes and edges in an ontology.

However, VSD is a highly difficult task, that achieves even lower precision rates than word sense disambiguation (WSD) for other categories. This is arguably due to the inherent complexity of Verb senses and the fact that their meaning is more pervasive than that of nouns, thus making it more difficult to discretize [3]. Moreover, most of works in WSD in general and VSD in particular, are focused on English corpora, while Spanish has little or no coverage on VSD.

Find below an example of verb sense ambiguity for the lemma hablar, with five senses in the SenSem lexicon. This can be seen in example 1.
Example 1. [...] una joven realizadora irlandesa que [...] habla castellano con acento malagueño ya que de pequeña vivió tres años en Estepona.

A young Irish producer who speaks Spanish with Malagueño accent because she lived in Estepona for three years when she was a kid.

Example 2. El presidente del Gobierno [...] hablará con las principales autoridades de ambos estados sobre las perspectivas de la Unión Europea [...].

The President of the Government will speak with the main authorities of both States about the prospects for the European Union.

In example 1 hablar is a state with the meaning of “being able to speak a language”, while in example 2 it is a process with the meaning “talk”. The subcategorization frames of each sense are also very different: for the first, we have two themes, while the second has the typical communication frame, with an Agent-Origin, a Theme and a Goal-Receiver.

The bottleneck for automated WSD is the number of annotated examples to learn from. This is because the most useful information for WSD seems to be word forms or lemmas co-occurring with a given sense, alongside syntactic information [12]. Given the Zipfian distribution [19], any number of examples will be small to cover most of the unseen examples, but a small increase in the number of examples can yield an important improvement. In this scenario, semi-supervised approaches to VSD are very promising, because they overcome the limitation of supervised approaches as a means to acquire a big number of annotated examples at a low cost, even if at the expense of some inadequacies.

In this paper we present a study on semi-supervised machine learning techniques to improve VSD for Spanish. Starting from a small annotated corpus, we extend the coverage by adding examples from a larger unannotated corpus, using bootstrapping and active learning techniques.

The rest of the paper is organized as follows. In the next section we describe some relevant work that addresses the problem of VSD and WSD using semi-supervised methods. Then, in section 3 we detail the iterative bootstrapping algorithm we use on our set of experiments. In section 4 we describe the corpora we use and the features to characterize them, and then in 5 we present the experiments and results. Finally, we conclude with some remarks and our lines of current and future work.

2 Relevant Work

This work builds on two main areas of previous work: bootstrapping techniques as applied to word sense disambiguation and verb sense disambiguation.

The landmark work on bootstrapping for word sense disambiguation is the 1995 Yarowsky paper [17]. In his work, Yarowsky builds a disambiguation model based on the words co-occurring with manually labeled examples. Then, this model is applied to unlabeled examples. Examples that can be assigned a sense by the model are then incorporated as training examples, and a new model is trained. This process is iteratively applied until a termination condition is
reached, namely, no new examples can be assigned a sense or the reliability of the evidence found by the model is too low. After each iteration, the resulting model has arguably larger coverage than previous versions. Therefore, this method is useful to build a real-life tool out of a limited number of examples.

Ye and Baldwin [18], use Selectional Preferences (SP) extracted with a Semantic Role Labeler (SRL) for VSD. Their VSD framework is based upon three components: extraction of disambiguating features, selection of the best disambiguating feature with respect to unknown data and the tuning of the machine learner’s parameters. For their study they use a Maximum Entropy algorithm [2]. The VSD features they used include selectional preferences and syntactic features, e.g., bag of words, bag of PoS tags, bag of chunks; parsed tree based features using different levels of the tree as source of information; and non-parse trees based syntactic features, e.g., voice of the verb, quotatives, etc.

Another work on English VSD is the one by Chen and Palmer [3], presenting a high-performance broad-coverage supervised word sense disambiguation (WSD) system for English verbs that uses linguistically motivated features and a smoothed maximum entropy machine learning model. Kawahara and Palmer [6] presented a supervised method for verb sense disambiguation based on VerbNet. Contrary to the most common VSD methods, which create a classifier for each verb that reaches a frequency threshold, they created a single classifier to be applied to rare or unseen verbs in a new text. Their classifier also exploits generalized semantic features of a verb and its modifiers in order to better deal with rare or unseen verbs.

In the SemEval 2007 task for multilevel semantic annotation of Catalan and Spanish [7], Márquez et al. [10] primarily focused on Noun Sense Disambiguation. They used a three way approach: if the word has more than a threshold number of occurrences, it is classified with a SVM classifier; if the word has less occurrences than the threshold it is assigned the most frequent sense (MFS) in the training corpus; if the word is not presented in the training corpus then it is assigned the MFS in WordNet. The SVM classifier features were a bag of words, n-grams of part-of-speech tags and lemmas, and syntactic label and syntactic function of the constituent that has the target noun as head.

Other work in WSD with applications in Spanish is the work of Montoyo et al. [9] where the task of WSD consists in assigning the correct sense to words using an electronic dictionary as the source of word definitions. They present a knowledge-based method and a corpus-based method. In the knowledge-based method the underlying hypothesis is that the higher the similarity between two words, the larger the amount of information shared by two of their concepts. The corpus-based method is based on conditional maximum-entropy models, it was implemented using a supervised learning method that consists of building word-sense classifiers using a semantically annotated corpus. Among the features for the classifier they used word forms, words in a window, part-of-speech tags and grammatical dependencies.

In the line of semi-supervised techniques, active learning has been applied successfully to VSD in Dligach and Palmer [4], where they explore the benefits of
using an unsupervised language model (LM) to select seed examples as a starting corpus for an iterative active learning approach. It involves training a LM on a corpus of unlabeled candidate examples and selecting the examples with low LM probability. This smart starting point seems to provide a better performance for verb senses, where there is a skewed distribution of classes, because it is able to select representative examples of minority classes.

In our approach we explore bootstrap as a basis for further, more sophisticated approaches to VSD, combining the strengths of bootstrapping with those of active learning.

3 Semi-supervised Algorithm for VSD

In this Section we describe the general algorithm for semi-supervised VSD.

First we obtain a manually annotated corpus, and we divide it in the starting training corpus (80%) and testing corpus (20%). The way we create our test corpus is to randomly choose instances from the whole corpus, but we make sure the proportion of instances per class for the test corpus is similar to the one for the whole corpus, e.g., if the whole corpus has 10% occurrences for one class then the test corpus has 10% occurrences for that same class.

The algorithm starts with the instances from the initial corpus that weren’t reserved for testing. This is our initial training set. We train a model for this training set and we use it for the bootstrapping / active learning iterations. In each iteration, we use the model learned from the training corpus to classify a randomly selected sample of instances from the unannotated corpus. We had to use a sample rather than the whole corpus for reasons of hardware limitations. However, we feel the random sampling reflects the behavior of the totality of the unannotated corpus, since these instances are randomly selected in each iteration.

For each instance in this sample of the unannotated corpus, the classifier predicts a class with certain probability. Per bootstrapping, if the confidence of the classifier is over a threshold, that instance is automatically added to the annotated corpus as a new example, with the predicted class as the label. Per active learning, we apply an uncertainty sampling [15] approach and choose the $k$ instances with the least confidence (i.e., the least probability), and give them to an oracle to manually annotate them.

With the newly annotated corpus (the annotated corpus from the last iteration plus the new instances of the current iteration), we get new features to train a new model for the following iteration. Before running the following iteration, we evaluate the new training model using 10-fold cross-validation on the training set accumulated to that moment by the bootstrapping iterations. We get the accuracy for this evaluation and we continue if the model’s accuracy is over a certain threshold. Another stopping criterion is a parameter indicating the maximal number of iterations. If no stopping criterion is met, we use the new model for the following iteration.
When a stopping criterion is met, we classify test instances both with the starting model and with the model from the last iteration. Thus we assess the impact of the semi-supervised approach.

4 Data

4.1 Corpora

The annotated corpus for the initial model is the Spanish portion of SenSem, a manually annotated corpus for both Spanish and Catalan. This corpus provides information on the 250 most common verbs occurring in Spanish. The task of VSD is to disambiguate the different senses for a given verb lemma. Thus, we discard those lemmas with a single sense, roughly 10% of the lemmas. SenSem provides some information on the annotated examples, besides the verbal sense: category, function and role of the constituents of the clause where the verb is occurring. However, we decided to do a preprocess of the corpus using Freeling [11] in order to get some extra information from it such as chunks, syntactic functions and dependency triples. Additionally, this preprocessing is the same that we can carry out in the unannotated corpora, thus making annotated and unannotated examples more comparable.

To avoid the noise introduced by the less frequent classes, the baseline experiment was set only for those verb lemmas with 2 or more senses with more than 10 occurrences each in the labelled corpus. The filtered annotated corpus has over 19,000 examples in 175 lemmas, with a mean of 111.16 examples per lemma. There are 460 senses, with an average of 43.52 examples per sense and 2.62 senses per lemma. For training we used 15,000 examples and the remaining 4,000 for testing.

For the semi-supervised VSD experiments we use the Wikicorpus [14] as an unannotated corpus to take examples for classification and later add to the annotated corpus. Again, we preprocess the Wikicorpus using Freeling to get extra information such as chunks, syntactic function and dependency triples. The corpus already provided information on the lemmas of the words. In the preprocessing of the corpus, we took each automatically analyzed sentence and get the information of all the verbs in them. Those sentences with verb lemmas on our list of verbs to disambiguate were used as unannotated examples.

4.2 Features

Initially we selected as features unigrams, bigrams and trigrams of lemmas. We discarded information about semantic roles and constituents that was provided by the annotated corpus, because the unannotated corpus does not have that kind of information, and annotated and unannotated examples would be incomparable. We then parse the corpora with Freeling to use the chunks, syntactic functions, and dependency triples Freeling returned, besides the bags-of-ngrams the corpora already provided; this way we were able to get a richer set of features.
than just bags-of-ngrams. The selected features for the final experiments were: bags-of-ngrams (unigrams, bigrams and trigrams), chunks, syntactic functions, and dependency triples. Features were also filtered so only those appearing a minimum of 5 times in the annotated corpus were used for the experiments. This provides a reduction of dimensionality that makes the problem more tractable and less prone to overfitting.

5 Experiments

5.1 Supervised Classification

To choose an algorithm for classification we tested two algorithms that are proven to have good results for text categorization: Multinomial Naïve Bayes (MNB) [8] and Support Vector Machines (SVM) [5]. As a baseline, we provided the algorithm that assigns each occurrence of a lemma to the most frequent sense for that lemma. The classification is evaluated using 20% of the corpus, as explained in section 3. As an evaluation metric we use accuracy, which is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

Both methods have good results, with an accuracy mean of over 71%, being both a little under the performance of the baseline. We introduced feature selection with information gain, based on the approach of Yang and Pedersen [16], obtaining a small gain in performance.
Figure 1 shows a boxplot with the accuracy of all the verbs to disambiguate, comparing the 5 different classification methods. All classifiers have an accuracy between 71% and 72%. SVM with feature selection has somewhat better performance than the rest, plus it is the only one to outperform, although slightly, the baseline algorithm. However, doing a matched-pair student’s t-test with R [13] with the null hypothesis $H_0$ being “the mean of the differences between accuracies is equal to zero” (i.e., there is no real gain from using SVM over the baseline), we got a high $p$-value ($\approx 0.82$) so we cannot reject $H_0$.

As the classifier selection does not seem to be very relevant to the final results, we decided to use SVM with feature selection in the rest of our experiments as it is still the classifier with better performance and it is the one that seems less prone to overfitting. Particularly, we didn’t use the baseline classifier for our semi-supervised tasks as it wouldn’t benefit from obtaining more annotated examples.

5.2 Semi-supervised Classification

The following parameters were set in our experiments for bootstrapping.

We set a threshold confidence of 90% to automatically add a new example. This threshold was empirically set after some preliminar experiments as it avoids adding too much error when adding automatically annotated instances. The threshold accuracy to stop the iterations after the 10-fold cross-validation was set on 50%, since roughly half of the lemmas had an initial accuracy between 50 and 60%, to avoid stopping algorithm too early for those lemmas; and the maximum amount of iterations was set to 10 as at this point no new instances were added or the classifier diverge to only add instances of the most frequent sense. For the iterations we run the algorithm in section 3 with the classifier in section 5.1 and the features described in section 4.2. The results were evaluated using the test corpus described in section 3.

Figure 2 shows the accuracy of the classifiers before and after the bootstrap task. As we can see in the Figure, the performance is improved. One of the drawbacks of having bootstrap is its bias towards the most frequent sense, which is common place in skewed distributions. In our experiments, in average, the most frequent classes of each verb were about 65% of the corpus before the bootstrap algorithm took place, and were 75% of the corpus afterwards. However, the bootstrapping approach still outperforms the most frequent sense baseline.

Table 1 shows some of the verbs we worked with and their results before and after the bootstrap iterations. The first 10 rows of the table show those verbs achieving better results after the bootstrap task. These verbs have between 2 and 4 senses per lemma, with many of them having 3, and the average of performance boost for the accuracy is 28 points. The last 5 verbs are the ones with the worst performance after the bootstrap task. In this case the verbs have between 2 and 3 senses per lemma, and the average of points the accuracy drops is 17 points.
Fig. 2: Comparison of accuracies of approaches to VSD: most frequent sense baseline, SVM with the starting annotated corpus only and SVM with bootstrapping.

5.3 Active Learning

Although the bootstrap task proved to improve performance results, it did it with a bias towards the most frequent senses, while minority senses were ignored. Hence, we did an initial exploration on pool-based active learning (AL) with uncertainty sampling to add new manually annotated instances from the unlabelled corpus together with the ones automatically annotated by bootstrap. The intuition driving these experiments was to achieve better balance between senses in the final instances added from the unlabelled corpus. The manually annotated senses were labelled by a human expert who acted as the oracle described in section 3.

Due to time restrictions, we could only label very few verbs. However, results were promising: in many cases we saw a pattern where applying AL gave some of the least frequent senses in the verbs a boost in the performance measured by the $F_1$ score per sense. Also, the population count for those senses did not vanish over each iteration, in contrast with bootstrap tasks where is generally the MFS the only one to grow. Nevertheless, these results are very preliminary and need to be enhanced for a real evaluation.

Figure 3 shows five of the lemmas we worked with: “abrir” (open), “creer” (believe), “hablar” (speak), “pensar” (think) and “ver” (see). The plots compare the $F_1$ score per sense of the lemmas, result of 10-fold cross validation of the training corpus, after each iteration with and without AL. As we can see in the plots on the left, the ones with bootstrap only, the performance tends to improve for one sense only, the MFS, but in the plots on the right it improves for some of the other senses, of which some new instances are added throughout
Table 1: Detailed accuracy obtained for some verbs, before and after applying bootstrapping.

<table>
<thead>
<tr>
<th>Verb</th>
<th>Accuracy Before</th>
<th>Accuracy After</th>
<th>Precision Before</th>
<th>Precision After</th>
<th>Recall Before</th>
<th>Recall After</th>
<th>No. Senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>acercar</td>
<td>61.90</td>
<td>90.48</td>
<td>0.59</td>
<td>0.93</td>
<td>0.62</td>
<td>0.91</td>
<td>3</td>
</tr>
<tr>
<td>conducir</td>
<td>20.83</td>
<td>45.83</td>
<td>0.17</td>
<td>0.45</td>
<td>0.21</td>
<td>0.46</td>
<td>4</td>
</tr>
<tr>
<td>dedicar</td>
<td>56.00</td>
<td>88.00</td>
<td>0.56</td>
<td>0.89</td>
<td>0.56</td>
<td>0.88</td>
<td>3</td>
</tr>
<tr>
<td>encontrar</td>
<td>73.91</td>
<td>100.00</td>
<td>0.76</td>
<td>1.00</td>
<td>0.74</td>
<td>1.00</td>
<td>3</td>
</tr>
<tr>
<td>hallar</td>
<td>52.17</td>
<td>86.96</td>
<td>0.52</td>
<td>0.91</td>
<td>0.52</td>
<td>0.87</td>
<td>3</td>
</tr>
<tr>
<td>interpretar</td>
<td>44.00</td>
<td>80.00</td>
<td>0.42</td>
<td>0.85</td>
<td>0.44</td>
<td>0.80</td>
<td>3</td>
</tr>
<tr>
<td>llambar</td>
<td>58.33</td>
<td>83.33</td>
<td>0.54</td>
<td>0.84</td>
<td>0.58</td>
<td>0.83</td>
<td>4</td>
</tr>
<tr>
<td>renovar</td>
<td>48.00</td>
<td>72.00</td>
<td>0.40</td>
<td>0.67</td>
<td>0.48</td>
<td>0.72</td>
<td>3</td>
</tr>
<tr>
<td>traer</td>
<td>66.67</td>
<td>91.67</td>
<td>0.78</td>
<td>0.93</td>
<td>0.67</td>
<td>0.92</td>
<td>2</td>
</tr>
<tr>
<td>valer</td>
<td>64.00</td>
<td>88.00</td>
<td>0.56</td>
<td>0.78</td>
<td>0.64</td>
<td>0.88</td>
<td>4</td>
</tr>
<tr>
<td>completar</td>
<td>83.33</td>
<td>66.67</td>
<td>0.83</td>
<td>0.44</td>
<td>0.83</td>
<td>0.67</td>
<td>2</td>
</tr>
<tr>
<td>implicar</td>
<td>100.00</td>
<td>84.62</td>
<td>1.00</td>
<td>0.72</td>
<td>1.00</td>
<td>0.85</td>
<td>3</td>
</tr>
<tr>
<td>incluir</td>
<td>92.31</td>
<td>60.23</td>
<td>0.93</td>
<td>0.48</td>
<td>0.92</td>
<td>0.69</td>
<td>2</td>
</tr>
<tr>
<td>repartir</td>
<td>77.27</td>
<td>59.09</td>
<td>0.80</td>
<td>0.77</td>
<td>0.77</td>
<td>0.59</td>
<td>2</td>
</tr>
<tr>
<td>terminar</td>
<td>62.50</td>
<td>50.00</td>
<td>0.53</td>
<td>0.47</td>
<td>0.62</td>
<td>0.50</td>
<td>3</td>
</tr>
</tbody>
</table>

iterations. Exactly how new instances are added is shown in Figure 4, where the charts represent the amount of examples per sense after each iteration, again with and without AL. We can see that, with AL, the least frequent classes also grow throughout iterations for all verbs, but specially in the cases of the lemmas “apuntar”, “hablar” and “ver”.

6 Conclusions and Future Work

In this paper we present our work to assess the impact of different supervised and semi-supervised methods for Spanish VSD. We found that a bootstrapping algorithm to automatically add new examples from a large unlabelled corpus effectively improves the performance of a classical supervised approach. However, bootstrap has a bias towards the most frequent sense when adding new examples. In a preliminary approach, we applied active learning to deal with this flaw, obtaining promising results.

In future work we intend to keep working in the active learning area in the line of Diligach and Palmer [4]. Also, we will include experiments where examples will be characterized by other features, such as synsets of verb constituents and semantic roles. We plan to incorporate richer information from verbal subcategorization frames, in the line of Ye and Baldwin [18].
Fig. 3: F-score obtained for each of the senses for the lemmas “apuntar”, “creer”, “hablar”, “pensar” and “ver”, across iterations, with and without active learning.
Fig. 4: Population count for each of the senses for the lemmas “apuntar”, “creer”, “hablar”, “pensar” and “ver”, across iterations, with and without active learning.
References

Turkish Children’s Acquisition of Nouns and Verbs: A Developmental Study

Fatma Iskin

Middle East Technical University, Informatics Institute, Department of Cognitive Science, 06800, Ankara, Turkey
ifatma@metu.edu.tr

Abstract. The major issue tackled in this article is the early word acquisition of Turkish children. The study particularly deals with children’s lexical diversity. Statistical patterns are investigated in the data, which is a corpus obtained from naturally occurring children’s speech while depicting their paintings in the class environment. Children belong to three major age groups which are: 40-42 months old, 43-49 months old and 55-59 months old. Whether there are any relationships between word types and acquisition order was investigated by means of the Type-Token Ratio (TTR). Results showed that there is a statistically significant developmental effect among age groups in terms of lexical richness. When age increases, the number of acquired words increase, as well. Another finding is that there is a frequency effect in the words used by the youngest group as well as the middle age group. In contrast, frequent word type changed in the oldest group. In addition, according to the TTR results, there is a developmental verb dominance in the data sets.

Keywords: Lexical diversity, types-tokens ratio, noun bias, verb bias, acquisition, age effect, syntax, vocabulary development, lexical richness, frequency of words

1 Introduction and Literature Overview

Language acquisition studies examine various aspects of children’s language production by analyzing statistical measures from spontaneous (naturally occurring) speech. Lexical diversity has an important role in the literature both on normal and atypical language development [5]. In this study, lexical diversity is measured for monitoring the developmental effect among three age groups which are 40-42 months old, 43-49 months old and 55-59 months old. Lexical diversity is a measure of how many different words are used in a text[4]. Thus, if the usage of a particular vocabulary increases in a text, higher lexical diversity occurs. The Speaker or the writer has to use many different words to diversify a text lexically, with little repetition of these words.

Lexical diversity (LD) refers to the range and variety of vocabulary deployed in a text by either a speaker or a writer [7]. Numerous measures of LD have been proposed, however, perhaps the best known measure is the Types-Tokens ratio, or the TTR analysis to measure it statistically [6].
1.1 Types-Tokens Ratio

In this study, for calculating the TTR, the number of different words a child uses in a speech transcript is divided by the total number of words in the transcript in order to determine the degree of variation [4]. The traditional LD measure is the ratio of different words (types) to the total number of words (tokens), the so-called TTR [4]. An obstacle with the TTR measure is that text (or transcript) samples containing large numbers of tokens give lower values for TTR and vice versa. The reason for this is that the number of word tokens can increase infinitely. Although the same is valid for word types, it is often necessary for the writer or speaker to re-use several function words in order to produce one new (lexical) word [4]. This implies that a longer text (or transcript) in general has a lower TTR value than a shorter text/transcript. The most preferable method to avoid such issues is that the data is standardized in terms of the size of the sample and samples of the same numbers of words are compared. For example, taking the first 50 words from all groups could be a solution to avoid this issue. How we dealt with this issue will be provided in the Towards to the Golden Data/Results part.

1.2 Noun-Verb Bias

The noun-verb bias is a continuing debate in the field of word acquisition. According to the Universal Noun Advantage View, children’s early lexicon is made up of nouns, representing concrete objects [2]. Verbs, and verb-like items, are acquired later as they involve more cognitively complex tasks for children to accomplish. Whether nouns are learnt before verbs or vice versa have been argued extensively. Xuan and Dollaghan preferred to use bilingual children to research the noun vs. noun bias because they thought that the studies with monolingual children in different languages vary due to group differences and socio-demographic characteristics [8]. Experiments were conducted among fifty Mandarin-English bilingual children who knew 50-300 words. Data was collected from the parents with the help of inventories method. The mean percentage of nouns among bilingual children for English was 54%. The mean percentage of nouns for Mandarin was 38%. Nouns outnumbered the total number of words in both languages. The list of the most frequent fifty words in English included more nouns and fewer verbs for Mandarin[8] They revealed that features of the input significantly change early lexicons. This finding showed that children’s lexicon for Mandarin includes more verbs. A recent study by Avcu [1] suggests that Turkish tends to be a verb dominant language with verbs. This is attributed to the fact that verbs are mostly located at the end of utterances, and a particularly rich but regular and transparent morphology supports early verb acquisition.

2 Research Questions

In this study, we first aimed to find the lexical diversity among three age levels which are 40-42 months, 43-49 months and 55-59 months. Data was collected
from spontaneous child speech. It was hypothesized that there would be developmental differences between three age groups in terms of the TTRs. Secondly, words among the three groups were examined in terms of their frequencies and lexical types. In the light of the frequency statistics, we investigated whether there are any relationships between word types and acquisition order some authors find.

3 Method

3.1 Sample

An experiment was conducted on 44 children from a kindergarten in Ankara. There are 21 female children and 23 male children in the group. All the subjects have corrected-to-normal vision and normal speech abilities in parallel with their age group. The parents are academic personnel or administrative personnel. The parents socio-economic levels are above average with respect to the levels in Turkey. The subjects are grouped into three in terms of their ages: 40-42 months old, 43-49 months old and 55-59 months old, i.e., they are 3, 4 and 5 years old. The data was collected by a fellow graduate student.

<table>
<thead>
<tr>
<th>Age Period</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>40-42-months-old</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>44-49-months-old</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>55-59-months-old</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1. The age and gender of the subjects used for the study

3.2 Experimental technique and program

A voice recorder was used for recording children’s voices while they were depicting their own paintings. The experimenter also used some notes about the pictures that the children drew. For example, he indicated the names and age of each child. The paintings were stored as pictures. Children were in their natural environment and their teacher was present. The teacher called each child’s name one by one. Each child came to the stage, took their paintings and then explained/described what they drew in this particular picture. After the data collection period, there were three recordings for each age group. Firstly, all the data taken from these recordings were transcribed. Next, the data for children was separated from the experimenter’s and the teacher’s speech, thus the raw children data was obtained. Using a computational program, Hasim Sak Perceptron 2.0, the sentences were morphologically parsed. An example from the raw data is: oynuyor oyna[Verb].
In the first group, children produced 141 words in total. 572 words were produced by the second age group and 517 words were produced by the third age group.

3.3 Towards the Golden Data/Results

A native Turkish speaker in linguistics examined the data one by one, and recognized some mistakes which despite being syntactically correct, are semantically incorrect. For example; "farkında olmak" is a verb phrase, yet the program parsed it as two separate words. To eliminate these types of mistakes and to reach the golden data, the analyzer consulted an expert linguist. A second perceived problem is about the data standardization to minimize the data to the first 141 words, which is the minimum number of words occurring in the first group of children. There are 13 children in the first group and they produced 141 words. There are 14 children in the second group but they produced 572 words. When we just take the first 141 words from this second data, we can just reach 3 or 4 children. Each child described their painting revolving around one topic. For example, one of them talked about funfair throughout the recording, thus used funfair-related words. To sum up, there are more children in the first group so they produced more word types than the other two groups. To avoid these types of problems, percentage was taken for each group, and then they were compared to each other in order to reach the most accurate empirical results in this study.

3.4 Stimuli, Procedure and Location

The major stimuli in the recordings is simple; the teacher asks children to depict their paintings. A single question was asked by the teacher: What did you depict in this picture? Please tell us. She does not use any guidelines for the first two groups (the children in the 40-42 months age group and those in the 43-49 months age group). However, the third and the oldest group cut some objects from magazines and stuck them to their paintings. The experimenter did not give any stimuli to children. He just recorded the children’s speech. The class in which the recordings were made does not contain tables, boards or desks. It only contains mats.

4 Results

To show all the words and their types, Table 2 was formed. There are 10 word types which are nouns, verbs, determiners, conjunctions, adverbs, adjectives, pronouns, postpositions, wh-words (like who which, why) q-particle which is "misin" or "musun". It is a question particle.

Because of the differences in total word counts among age groups, the percentage method is used and provided in Table 3. Table 2 shows that, there are more nouns than verbs in all three groups. A mathematical calculation was carried out for analyzing the noun/verb ratio. Firstly, just the noun/verb ratio for
Table 2. All words with lexical distribution

<table>
<thead>
<tr>
<th>Word Type</th>
<th>40-42 Months</th>
<th>43-49 Months</th>
<th>55-59 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>55</td>
<td>220</td>
<td>186</td>
</tr>
<tr>
<td>Verb</td>
<td>15</td>
<td>71</td>
<td>105</td>
</tr>
<tr>
<td>Determiner</td>
<td>15</td>
<td>63</td>
<td>50</td>
</tr>
<tr>
<td>Conjunction</td>
<td>15</td>
<td>48</td>
<td>54</td>
</tr>
<tr>
<td>Adverb</td>
<td>7</td>
<td>43</td>
<td>17</td>
</tr>
<tr>
<td>Adjective</td>
<td>12</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>Pronoun</td>
<td>19</td>
<td>65</td>
<td>42</td>
</tr>
<tr>
<td>Postposition</td>
<td>0</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Wh-Word</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>q-particle</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Total Word Count</td>
<td>141</td>
<td>572</td>
<td>517</td>
</tr>
</tbody>
</table>

all groups are 55/15=3.6 220/71=3.09 and 186/105= 1.77. According to the ratios, there are age-related effects in acquiring verbs. Hence, when the child grows, the verb ratio increases. The tables showing the TTRs are drawn. In Table 3 the standardization method was used. Thus, the first 141 words were not taken from all three groups.

Table 3. Type-Token Ratio Table with Standardization

<table>
<thead>
<tr>
<th>Age Period</th>
<th>40-42 Months</th>
<th>43-49 Months</th>
<th>55-59 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types Counts</td>
<td>162</td>
<td>60</td>
<td>88</td>
</tr>
<tr>
<td>Tokens Counts</td>
<td>141</td>
<td>141</td>
<td>141</td>
</tr>
<tr>
<td>Types-Tokens Ratio</td>
<td>43.97%</td>
<td>42.55%</td>
<td>62.41%</td>
</tr>
</tbody>
</table>

According to the results taken from Table 3, there is a developmental effect in parallel with age. There is an explicit difference between 40-42 and 55-59 age periods. Although the ratio is 43.97% for the youngest group, 62.41% was found for the oldest group. This result showed that between the 40-42 and 55-59 months old children, there is a significant difference in word acquisition.

Table 4, 5, and 6 were constructed to analyze the most frequent words in the data. For the youngest age group, most frequent words are ben (I), bu (that), yok (not existing/nothing), sey (something), bir (a), which belong to different word types. According to the second groups data, the most frequent words are bu (this), evet (yes), bir (a), köpek (dog), burada (here). As seen, bu (that) is also frequently occurred in the 43-49 months old age period. Lastly, the most frequent words for the third and the oldest group are var (there is/there are), bir (a), evet (yes), sonra (after). The most frequent word for this age group is var (there is/there are). This word is also used frequently in the adults speech. The frequency value for "var olmak" is 97 taken from [3]. This frequent word
Table 4. Frequency table for the 40-42 months old children

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
<th>Word Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ben (I)</td>
<td>12</td>
<td>Pronoun</td>
</tr>
<tr>
<td>Bu (that)</td>
<td>11</td>
<td>Determiner</td>
</tr>
<tr>
<td>Yok (nothing)</td>
<td>10</td>
<td>Verb</td>
</tr>
<tr>
<td>Sey (something)</td>
<td>10</td>
<td>Noun</td>
</tr>
<tr>
<td>Bir (a)</td>
<td>7</td>
<td>Adverb</td>
</tr>
<tr>
<td>Çocuk (child)</td>
<td>3</td>
<td>Noun</td>
</tr>
<tr>
<td>Ama (but)</td>
<td>3</td>
<td>Conjunction</td>
</tr>
<tr>
<td>Çukur (hole)</td>
<td>3</td>
<td>Noun</td>
</tr>
<tr>
<td>Üçgen (triangle)</td>
<td>3</td>
<td>Noun</td>
</tr>
<tr>
<td>Yaptm (I did)</td>
<td>3</td>
<td>Verb</td>
</tr>
</tbody>
</table>

Table 5. Frequency table for the 43-49 months old children

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
<th>Word Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bu (This)</td>
<td>8</td>
<td>Determiner</td>
</tr>
<tr>
<td>Evet (yes)</td>
<td>8</td>
<td>Adverb</td>
</tr>
<tr>
<td>Bir (a)</td>
<td>8</td>
<td>Determiner</td>
</tr>
<tr>
<td>Köpek (dog)</td>
<td>6</td>
<td>Noun</td>
</tr>
<tr>
<td>Burada (here)</td>
<td>4</td>
<td>Pronoun</td>
</tr>
<tr>
<td>Baska (other)</td>
<td>3</td>
<td>Adjective</td>
</tr>
<tr>
<td>Balık (fish)</td>
<td>3</td>
<td>Noun</td>
</tr>
<tr>
<td>Deniz (sea)</td>
<td>3</td>
<td>Noun</td>
</tr>
<tr>
<td>Onlar (they)</td>
<td>3</td>
<td>Pronoun</td>
</tr>
</tbody>
</table>

Table 6. Frequency table for the 55-59 months old children

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
<th>Word Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var (existence)</td>
<td>8</td>
<td>Verb</td>
</tr>
<tr>
<td>Bir (a)</td>
<td>6</td>
<td>Determiner</td>
</tr>
<tr>
<td>Evet (yes)</td>
<td>5</td>
<td>Adverb</td>
</tr>
<tr>
<td>Sonra (after)</td>
<td>3</td>
<td>Postp</td>
</tr>
<tr>
<td>Tas (rock)</td>
<td>2</td>
<td>Noun</td>
</tr>
<tr>
<td>Koklamak (to smell)</td>
<td>2</td>
<td>Verb</td>
</tr>
<tr>
<td>Tane (piece)</td>
<td>2</td>
<td>Noun</td>
</tr>
<tr>
<td>Merkez (center)</td>
<td>2</td>
<td>Noun</td>
</tr>
<tr>
<td>Kalkmak (get up)</td>
<td>2</td>
<td>Verb</td>
</tr>
</tbody>
</table>
is used in a sentence as: There is a dog here. When investigating the data, the researcher realized that, although the oldest group used a completed sentence like There is a dog here (where the child also points at the dog in the painting gesturally), the youngest group just used the dog in here gesturing at the dog in the painting. The youngest group’s sentence, therefore, often lacks the verb. The examples suggest that younger children tended to use incomplete sentences, whereas older children preferred to use completed sentences with verbs. This could be an answer for the debate about the noun bias versus verb bias in Turkish. Our data shows that older children use verbs and nouns correctly, i.e. they have a grasp of the predicate-argument structure. A piece of the raw data for the 43-49 months old children which are so close to the 40-42 months old children in terms of the TTR and 55-59 months old children is shown below:

43-49 months old:

Experimenter: Ela gelir misin? Merhaba Ela, ne çizdin burada anlat bakalım hadi bana. (Ela, can you come here? What did you depict here? Please explain it to us.)
  Ela: Baba, ben ve anne. (Father, me and mum.)
  Experimenter: Bu abla mı? (Is she the sister?)
  Ela: Ben. (Me)
  Experimenter: Kim? (Who?)
  Ela: Ben. (Me)
  Experimenter: Sen misin bu? Hmm, anladım... Başka neler çizdin anlat bakayım. (Is this you? Hmm, I got it, What else did you depict? Please explain.)
  Ela: Ağaç, çimen ve ... bir de güneş. (Tree, grass and sun.)
  Experimenter: Bir de güneş mi var? Peki neresi burası? (Is there also the sun? Where is this place?)
  Ela: Evimiz. (Our house.)
  Ela: Oyuncakların. (Their toys.)
  Experimenter: Bursanın bahçeniz mi sizin? (Is it your garden?)
  Ela: Evet. (Yes.)
  Experimenter: Evinizin bahçesi... Peki çok teşekkür ederim gidebilirsin. (It is your houses garden. Well, thank you.)

55-59 months old:

Experimenter: Rüzgar gelsin. Rüzgar ne kesmişti? Bir tane oda var. Askıda bir şeyler asılı. Rafflar var ve bir oturma odası yada salonun resmi var. Bakalım Rüzgar bunlarla neler yapmış? (Rüzgar, come here. What did Rüzgar cut out? There is something in the coat hanger. There was something in hanger. There is the picture of the living room or the drawing room.)
Rüzgar: Çocukları beğenmemiş. Ağaca yakın bir ev yapmaya karar vermiş. Henüz ağaca yaklaşılamamış. Merdiven de yokmuş. Hava kararmış. Yerde uyumuş. (He did not like the children. He decided to do a house from trees. He did not reach to tree yet. There was no staircase either. It was dark. He slept on the ground.)

Experimenter: Dışarda mı kalmış? Evi neden beğenmemiş peki? (Did he stay outside? Why didn’t he like the house?)

Rüzgar: Çünkü Çünkü . Masasının deliği açık. (Because because . Her tables hole was open.)

Experimenter: Hm o ağaca mı yapmak istiyor evini? (Hm did he want to build his home to these trees?)

Rüzgar: Evet. (Yes.)

As we indicated in Table 1, there are developmental effects in terms of verb counts. The examples above show that older children construct semantic relations between words, showing an awareness of the predicate-argument structure of verbs (e.g. like the children, do a house, sleep on the ground) with proper word order (SOV).

5 Discussion and Conclusion

The results of this study revealed three main findings. The first one is that there is a developmental effect in terms of lexical diversity between three age groups which are 40-42 months old, 43-49 months old and 55-59 months old. Results obtained from the TTR calculations are: 43.97% for the youngest age group, 42.55% for the middle age group and 62.41% for the oldest group. An explicit difference is found between 40-42 months old and 55-59 months old. When children grow, their lexical richness increases in the Turkish corpus. The second finding is that the verb ratio is increasing in parallel with age. Thus, children construct the relationship between verbs and nouns, and showing the syntactic relations between the constituents. The third finding is about the frequency of the words, which is also related with the second and the first findings. As the verbs and lexical diversity increase, the most frequent word in the data change to the verb lexical type. What is the mystery behind this rapid improvement of verbs? Is it just related with the linguistic ability or are there any other mechanisms inducing this change? Further studies would give more definitive answers by including more data. One could also investigate the influence of other cognitive effects such as environment on the development of verbs in children at 55 months age and older.

Acknowledgement I take this opportunity to convey my sincere thanks to Deniz Zeyrek. The idea of this study was conceived during her Language Acquisition course. She also checked the parsed data and eliminated mistakes in terms of syntactic structure. I appreciate her valuable feedback also in the process of writing the article. I extend my sincere thanks to Gökhan Gönül, who collected the data and transcribed half of the recorded data. Lastly, I would like to thank Murathan Kurfah for his help in data parsing. All remaining errors are mine.
References

Abstract. In this work, we present a distributional semantic model for recipes by building a representation of food ingredients and recipes in terms of co-occurrence vectors, recording their distributional pattern in a recipe corpus. We build a space for ingredients and two spaces for recipes. We show that by composing ingredients in a recursive manner we get a more accurate approximation of recipes than by a more basic method. This idea can be exploited in order to generate new recipes using existing ingredients by looking at similar vectors. Alternatively, it can be used to build an app which might be used to get inspiring ideas for what to cook while having a list of ingredients.

1 Introduction

When explaining the concept of an algorithm to someone unfamiliar with it, an often used example for real-world applications of “algorithms” is recipes. Recipes are (somewhat) unambiguous instructions on when to add what ingredient in what way and how to process it to result in a certain meal. Algorithms are written in some kind of language whose interpretation gives the interpreter (whether human or computer), some clue on how to perform it.

Using an online database of recipes, we therefore collect co-occurrence counts for ingredients, and then use them to compose them back into recipes. We find our ingredient space gives interesting results when queried for nearest neighbors of ingredients. The composed recipes, unsurprisingly, are close to original recipes with similar ingredients.

Like [7], we see food as a language, more specifically, recipes. In the tradition of distributional semantics [15], we assume that the “meaning” of a word or ingredient is given by the context in which it occurs. The distributional (gastronomical) hypothesis states that similar ingredients should appear in similar contexts.

2 Related Work

[16] build recipe retrieval and creation systems leveraging the relations between food-related words in WordNet. Users can enter a list of ingredients they have, and get presented with suggested recipes that best match the input query. To do that, they also develop a similarity measure between two lists of ingredients. Furthermore, their system can accurately predict a type of dish and a type of cuisine.
build and examine ingredient-ingredient networks of complements and substitutes of ingredients. Networks were built using PMI-weighted co-occurrences of ingredients in recipes, and suggestions for ingredient modifications in recipes, respectively. Ratings of recipes were predicted well, combining features from both models.

[11] build a corpus combining high-quality video recordings of cooking tasks with aligned textual descriptions, who in turn are aligned to pre-defined activity descriptions. While it also includes (a few) cooking ingredients, its focus is not on lexical and distributional knowledge of them, but on the cooking process itself, and the actions and tools involved, plus linking it with visual information. Similar work in semantic parsing of instructional cooking videos was done by [8].

In a similar direction goes [13], who develop a formal instruction language called MILK that recipes can be written in, and develop a small corpus of 350 recipes translated (by annotators) into MILK. In contrast, we build on a large and inconsistent web corpus and therefore don’t model the cooking process.

[1] test Rozin’s flavor principle [12] at the molecular level across the world, and find that Western European and North American recipes tend to pair ingredients that share the same flavor, whereas the East Asian recipes combine ingredients that do not have overlapping compounds. This difference might characterize a cuisine. Recently, the “food pairing hypothesis” was tested for Indian cuisine too [5]. It was found that the flavor sharing was significantly less than expected by chance and that the spices are ingredients that contribute the most to the negative food pairing.

3 Methods

We view recipes as sentences of a language, where a word is an ingredient. We further simplify our model by saying that the order of words or ingredients doesn’t matter, which is motivated by the fact that ingredients in our corpus aren’t always sorted by order of application.

**Open Recipes** 

is a database of recipes, automatically collected by crawler scripts, in a unified JSON format. As our input corpus, we used the latest Open Recipes dump (accessed on 2014–03-20). For practical reasons, we ignore a large part of the information present in the corpus: Neither instructional information on the cooking process is used, nor the amounts of ingredients.

12 whole Dinner Rolls Or Small Sandwich Buns (I Used Whole Wheat)
1 pound Thinly Shaved Roast Beef Or Ham (or Both!)
10 ounces carrots, shredded on a box grater or sliced whisper thin on a mandolin
optional: whole pomegranate seeds or fresh/dried rose petals, a bit of bee pollen

**Fig. 1.** Examples of uncleaned “ingredients” in the corpus

[1] https://github.com/fictivekin/openrecipes
Co-occurrences between ingredients were then collected. As a first step, the ingredient names were cleaned. Because of noisy data entered by users (some examples in Figure 1) and amounts and units being part of the ingredients in the corpus, there was a total of 412,858 different strings for ingredient names. Using a number of heuristics, that included cutting away anything that looked like a unit, a number, parts in parantheses, after commata, a naive stemming procedure, and more, we reduced that number:

No existing stemmer was used, due to the specific nature of our data: The cleaning wasn’t in the scope of a stemmer, since we actually want to get rid of a lot of words, not just normalize them. Instead, the cleaning/stemming was done using a number of regular expressions. First, substrings in parantheses were removed, then everything after a potential first comma, all digits (including symbols for fractions commonly used). A few other replacements and prunings were made, like splitting the text at “for” and only using the first part (which turns e.g. “tomatoes from spain” to “tomatoes”). Most importantly, using a hand-compiled list, units and their abbreviations were removed, since we don’t want to handle “1 tsp sugar” and “2 g sugar” differently.

<table>
<thead>
<tr>
<th>Number of recipes</th>
<th>172893</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ingredients (tokens in corpus)</td>
<td>1689892</td>
</tr>
<tr>
<td>Number of ingredients (types in corpus)</td>
<td>412858</td>
</tr>
<tr>
<td>Number of ingredients (types after unifying)</td>
<td>6514</td>
</tr>
</tbody>
</table>

Table 1. Information about the corpus

2 There are many different strings that could be reduced to the same ingredient. Sources for noise are comments by users (see the first two lines of Figure 1), several distinct ingredients listed as one (see the last line), or descriptions of how to handle the ingredient; in general everything that leads to inconsistent entries for the same ingredient.
In the end, the number of different ingredients was 6514 (see Table 1).

The resulting list was then weighted by occurrences. We considered only ingredients used more than ten times to filter out further noise, which brought us down to a reasonable number of 6,326 relevant ingredients. We considered purging even more, after all, this list still contains almost 150 distinct elements containing the string “onion”, but we assumed we would lose information by removing adjectives, at least for some ingredients. A “large onion”, for instance, could be used in very different contexts than a “small onion”.

$$\text{PPMI}(a, b) = \max\left(0, \log \frac{P(a, b)}{P(a)P(b)}\right)$$  \hspace{1cm} (1)

A large co-occurrence matrix of ingredients was then automatically created. For handling the vector space operations, we utilized the DISSECT toolkit [2]. The matrix was imported as a vector space, weighted with the Positive Pointwise Mutual Information (PPMI, see (1)) measure and reduced to 20 dimensions (after experimenting on 200, and 50 dimensions too), using Singular Value Decomposition. We chose such a low number in the hope that similar features would merge together and our results would eventually not just be reproducing the original recipes when ingredients were composed, but also finding similar recipes that don’t necessarily contain the ingredients of the original composition.

We then built recipe spaces using two different methods: In our BasicRecipes space we obtained recipe vectors by just setting the co-occurrence of each recipe with all of its ingredients to 1.

The ComposedRecipes space was created by summing the vectors of all of its ingredients using a recursive (see Figure 2) Weighted Additive function [10] with $\alpha$ and $\beta$ set to 1: For a given list of $n$ ingredients, if $n$ is 1, return the first element of the list as the composed vector. Otherwise, take the last two vectors from the list, compose them using the Weighted Additive function, and put them back on the list. This way, even recipes with a large amount of ingredients could be reduced to the same simple vector as ones with fewer ingredients.

4 Results

<table>
<thead>
<tr>
<th>Input ingredient</th>
<th>Nearest neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>flour</td>
<td>warm water, egg yolk, nutmeg</td>
</tr>
<tr>
<td>milk</td>
<td>melted butter, butter or margarine, eggs</td>
</tr>
<tr>
<td>mozzarella</td>
<td>basil, pasta, freshly grated parmesan</td>
</tr>
<tr>
<td>blueberries</td>
<td>frozen mixed berries, peaches, strawberries</td>
</tr>
</tbody>
</table>

Table 2. Nearest neighbors (cosine distance) of some ingredients
We manually and subjectively evaluated our ingredient space by giving it an input ingredient and judging whether we consider the most similar vectors to be related gastronomically, where as a similarity measure we used the Cosine Distance, defined as \( \cos(u,v) = \frac{u \cdot v}{\|u\| \|v\|} \). For the most part, this works very well. When presented with an ingredient, the nearest neighbors seemed to “fit well” to it (Table 2).

Next, we manually looked at the nearest neighbors of composed recipes with recipes in the two recipe spaces. Our investigations seem to confirm our intuition: The ComposedRecipes space outperforms the primitive BasicRecipes space. When given, for example, “water”, “flour”, “yeast”, and “salt”, composing them leads to nearest neighbors such as “home made mini bread”, “Italian bread”, and “french baguettes” in ComposedRecipes, but “pina colada baklava”, “huckleberry cheesecake”, and “honey-mint yogurt smoothie” in BasicRecipes.

Quantitative evaluation of our work proved challenging. To automatically evaluate the quality of our vectors, we took all of our ingredients that are found in WordNet [9], calculated the Cosine Distance between their vectors, and compared it with the JCN distance [6, see (3)\(^3\)] in WordNet, using the Brown corpus [3] as a seed corpus. As a comparison, we also compared it with the path distance in WordNet (see (2)).

\[
\text{Sim}_{\text{path}}(\text{ing}_1, \text{ing}_2) = \frac{1}{\text{minpath}(\text{ing}_1, \text{ing}_2)} \tag{2}
\]

\[
\text{Sim}_{\text{JCN}}(\text{ing}_1, \text{ing}_2) = \frac{1}{\text{IC}(\text{ing}_1) + \text{IC}(\text{ing}_2) - 2 \times \text{IC}(\text{LCS}(\text{ing}_1, \text{ing}_2))} \tag{3}
\]

<table>
<thead>
<tr>
<th>Cutoff at n</th>
<th>Path distance</th>
<th>JCN distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.63</td>
<td>0.53</td>
</tr>
<tr>
<td>20</td>
<td>0.30</td>
<td>0.11</td>
</tr>
<tr>
<td>50</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>100</td>
<td>0.04</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 3. Spearman’s \( \rho \) correlation between Cosine similarity and Path distance or JCN, respectively, at various cutoffs.

The Spearman correlation values for various cutoffs (using the most common \( n \) ingredients) can be seen in Table 3. None of them were significant with \( p < 0.05 \), but that is not too surprising: In WordNet, words are arranged according to their actual hierarchy, while our space attempts to represent how they are used in cooking, which could be (and is assumed to be) very different. For example, if butter and olive oil were used interchangeably in recipes, they would be very different.

\(^3\) IC is the information content of a concept based on some corpus, LCS is the least common subsumer, i.e. the deepest node that subsumes both ingredients.
close in our space, but not in WordNet, because olive oil is in fact something very different from butter.

![2D projection of ingredients](image)

**Fig. 3.** 2D projection of ingredients

A 2D projection of some top ingredients from our space (plotted using the R library LSAfun [4] with the Multidimensional Scaling (MDS) method), can be seen in Figure 3. It can be seen that some similar ingredients cluster together: The spices “paprika” and “pepper” together on the top, vegetables on the middle right, and mostly sweet or neutral items on the left side of the plot (“sugar”, “honey”, “raisins”).

In Figure 4, the nearest neighbors of “grapes” were plotted using the same technique. It can be seen that all grape-related vectors are fairly close to each other, followed by other fruits and a fruit juice.

## 5 Conclusion & Future Work

The presented methods are just a first step in what we expect to be a long future of distributional gastronomics. We took only the most obvious steps and expect others (and ourselves) to come up with more advanced techniques in the near future.

We composed solely using untrained Weighted Additive functions, and Cosine distance as a similarity measure. Future work could explore other composition methods and similarity measures. It would also be interesting to see whether giving up on some of our simplifications can produce more interesting results:
If the ingredient list is expected to be sorted in some way, order of ingredients could play a role in the weighting of composing vectors. A different improvement could lie in harnessing the supplied amounts (and units) of ingredients to see whether an ingredient plays a major role in a recipe or is just a small addition. An obvious improvement to our data would have been a cleaner corpus with standardized, fixed ingredient names, clearly separated by amounts and units.

We completely ignore the cooking instructions, by which we obviously lose a lot of information, exemplified by the fact that there are many ways to combine a set of ingredients. Integrating it — maybe in a manner similar to [14] — into the vector model should yield an improvement. A different source corpus as a start could partially solve many of the problems mentioned, but would likely come with a loss in size.

Better methods of automatic evaluation would be a welcome addition to our work, alternatively, evaluating it against a large-scale human gold standard: One could use a platform like Amazon Mechanical Turk\(^4\) to have annotators assess similarity judgements made by our system.

Very interesting to see would be applications built on distributional gastronomy. As proposed in our abstract, such an application could consist of making recipe proposals to a user who enters a list of ingredients they have. A very simple proof-of-concept of such an application was written as a small script working on our ingredient and ComposedRecipes space.

As shown in [1] (and recently validated for Indian cuisine in [5]), recipes from different parts of the world behave very different: While the flavor in western cuisine is mostly similar, East-Asian cuisine pairs contrasting flavors in typical recipes. Given that insight, vector spaces for different cuisines will look very different and localized spaces might gain informativeness over general ones.

\(^4\) http://mturk.com
References

Gender Identification for Adversarial Writing

Yunita Sari
Department of Computer Science, University of Sheffield
Regent Court, 211 Portobello
Sheffield S1 4DP, United Kingdom
y.sari@sheffield.ac.uk

Abstract. Past research in gender identification has explored various features and techniques such as stylistic features, function words, gender-linked cues, content words and more advance linguistic features. Although gender identification methods have achieved some degree of success in many literary and forensic applications, no studies have been undertaken specifically for adversarial writing. In this paper, we presented our experiment on gender identification in adversarial writing. We employed several feature sets including stylometric features, content words and function words. Our experiments show that standard gender identification task is more resistant to imitation attacks rather than obfuscation attacks. Among others, combination of function words and stylometric features appeared as the most stable features, producing the best accuracy of 75.56% on imitated document, and this figure only dropped slightly, to 62.22% and 72.45%, for obfuscated and original documents.

Keywords: Gender identification, Author Profiling, Adversarial writing

1 Introduction

The main challenge of research in adversarial stylometry is to identify the real author of an adversarial document given only some samples of original writing from several authors. Only a few studies [2, 4, 5, 9] tried to focus in this area and mostly failed to achieve high accuracy. Adversarial writing can be described as a type of writing where the original authors intentionally change their writing style in order to hide their true identity. Different cases such as plagiarism and sexual predator conversations can be considered as part of adversarial writing too.

In this paper, we are particularly interested to observe the effect of the adversarial attacks to the gender identification’s performance. Previous works have found that there is a strong correlation between gender of the author and his/her writing style. Gender identification methods have achieved some degree of success in many literary and forensic applications [6, 10]. However, no studies have been undertaken specifically for adversarial writing. We believe, profile of the author such as gender can be used as the clue to identify the true author of an adversarial document.
In this experiment, we explore several stylometric and contextual features that have been known to be effective to identify gender. We applied standard gender identification methods trained on original writings and tested on adversarial documents. In this experiment, we used Extended Brennan-Greenstadt adversarial corpus [4]. The obtained results show certain features (i.e. function words) are relatively resistant to the adversarial attacks.

The remainder of the paper is organized as follows. Section 2 provides a review of related work on gender identification. Explanation of adversarial writing is presented in section 3. Section 4 describes data set, features, classifier and result of our experiments. Discussion of experimental results are provided in section 5. Finally, the conclusion is presented in section 6.

2 Gender Identification

Research in gender identification has applied various features and techniques to different types of data set. For example, Koppel et al. [10] used a large number of content-independent features consist of 405 function words and a list of n-grams of part-of-speech to identify gender in 920 English documents of BNC corpus. Optimal performance (82.6% on accuracy) was obtained when a combination of function words and part-of-speech n-grams applied to non-fiction documents. In addition, Koppel et al. found an interesting fact that there is a strong difference in usage of determiners, negation, pronoun, conjunction and preposition between male and females either in fiction or non-fiction documents.

Another work by Schler et al. [15] studied the effects of age and gender on blogging. In contrast to Koppel et al., they used content-based features along with style-based features to identify author’s gender and age on 1,405,209 blog entries. Schler et al. used simple content words and LIWC’s special classes words [13]. By analysing content-words which have high frequency and information gain, they concluded that male bloggers tend to write more about politics, technology and money while female bloggers prefer to discuss about their personal life. However, both Koppel and Schler have agreed that stylometric features provide more information about gender of the author, despite the fact that there is big difference in content between male and female.

Gender-linked features are also useful for identifying gender. Work by Cheng et al. [6] applied nine gender-linked cues including affective adjective, exclamation, expletives, hedges, intensive adverbs, judgmental adjectives and uncertainty verbs. Past studies [8,12] showed that different gender of authors has their own preferences for using gender-linked cues. For example, compared to female, male authors rarely use emotionally intensive adverbs and affective adjective in their writing. Men often use first-person singular pronoun to express independence and assertions.

Since 2013, PAN (a periodic event on uncovering plagiarism, authorship and social software misuse) has included author profiling as the main task. Given a set of unlabeled document, participants need to identify the age and gender of the authors. In PAN-2014, there were 10 submissions to the task of author
profiling [14]. Among the submissions, participants who used second order representation based on relationships between documents, term, profiles and sub-profiles [11] successfully obtained the overall best performance on average for English and Spanish tasks. This result outperformed the vast majority of the teams who applied common author profiling features such as style and content-based features.

Past research provides evidence that gender identification with standard approach could obtain relatively high accuracy on non-adversarial writing. We are interested to apply the similar method to adversarial writing and observe the effect of the attacks to the gender identification performance.

3 Adversarial Writing

Most previous studies in authorship attribution assumed that the authors write in their original writing style without any intention to modify it. It has been shown that current authorship attribution methods could achieve relatively high accuracy in identifying authors [1]. Advances in authorship attribution have raised concerns about applying attribution methods for adversarial writing and has led to the new problem area called adversarial stylometry. According to Brennan et.al. [4], adversarial writing can be in the form of obfuscated writing, where the subject tries to modify his original writing style; imitated writing, where the subject is copying another person’s writing style; and translated writing, where machine translation is used to modify the original text. The main goal of adversarial writing is to hide the true author’s writing style.

Adversarial stylometry can be considered as new research in the authorship attribution fields. Only a few works have been addressed this area. Most of them studied how to develop an attribution method which robust to any adversarial attacks, including obfuscation, imitation and translation. However, previous work [2,4,9] show standard authorship attribution is not resistant to the adversarial attack. As the result, the attribution methods achieved very low accuracy that is sometimes even below the standard baseline performance.

4 Approach

In this experiment, we applied several stylometric and content-based features to capture the author’s writing style. Three different classifiers including SVM, K-NN and Naive Bayes were used. Each of the classifier were trained only on non-adversarial documents and tested on both original and adversarial documents.

4.1 Features and classifiers

Features
We applied three standard types of author profiling features in our experiment, including: stylometric features (SF), function words (FW) and content-based
features (CW). In addition, we also tried two different combinations of those features which are C1 (combination of SF and FW) and C2 (combination of SF, FW and CW). A list of function words was adapted from Cheng et al. work [6] with slight modification. Totally there were 201 function words used in this experiment. For content-based features, we used 1000 most frequent content words in the training data. Stylometric features contain 20 different types of statistical profile of the document as described in Table 1.

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Number of features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of words</td>
<td>1</td>
<td>all words used in a document (punctuations are excluded)</td>
</tr>
<tr>
<td>Total number of character</td>
<td>1</td>
<td>all character used in a document</td>
</tr>
<tr>
<td>Average word length</td>
<td>1</td>
<td>ratio of total number of characters in a document to total number of words in a document</td>
</tr>
<tr>
<td>Lexical diversity</td>
<td>1</td>
<td>ratio of total number of unique words to total number of words in a document</td>
</tr>
<tr>
<td>Hapax-legomena ratio</td>
<td>1</td>
<td>total number of words that occur only once in a document</td>
</tr>
<tr>
<td>Dis-legomena ratio</td>
<td>1</td>
<td>total number of words that occur twice in a document</td>
</tr>
<tr>
<td>Average number of words per sentence</td>
<td>1</td>
<td>ratio of total number of words in a document to total number of sentences in a document</td>
</tr>
<tr>
<td>Percentage uppercase</td>
<td>1</td>
<td>ratio total number of uppercase characters to total number of characters in a document</td>
</tr>
<tr>
<td>Total number of letters</td>
<td>1</td>
<td>total number of letters in a document</td>
</tr>
<tr>
<td>Total number of punctuations</td>
<td>1</td>
<td>total number of punctuations used in a document</td>
</tr>
<tr>
<td>Total number of specific part-of-speech tagset</td>
<td>10</td>
<td>total appearances of part-of-speech tagsets including ADJ(adjective), ADV(adverb), DET(determiner, article), CONJ(conjunction), PRO(pronoun), VERB(verb), NOUN(noun), NUM(number), ADP(adposition), X(other, for example: dunno, gr8, univeristy)</td>
</tr>
</tbody>
</table>

Classifiers
Several machine learning algorithms, including the WEKA’s implementations of SVM (with RBF kernel), K-NN and Naive Bayes [7] were used as the classifier.
SVM and K-NN hyperparameters were optimized based on accuracy, using two meta-classifiers: GridSearch and CVParameterSelection.

4.2 Data

In all experiments, an adversarial corpus called the Extended Brennan-Greenstadt corpus [4] was used. This corpus was created for the adversarial stylometry task using a survey conducted at Drexel University. Amazon’s Mechanical Turk (AMT) platform was used to create a corpus which includes text that has been obfuscated to disguise the author. This corpus is freely available and can be downloaded from the website of Drexel University’s Privacy, Security and Automation Lab\(^1\). It contains 699 original documents, 45 obfuscated documents and 49 imitated documents from 45 authors. Each document contains approximately 500 words. In addition, the corpus also contains demographic information about the author including, gender, age, language and education background. Details of the corpus can be found in their publications [4,5].

4.3 Experiment

In our experiment, we split the data into training and test sets. We set 654 original documents as training data and 45 documents for each type of texts as test data. Table 2 provides the distribution of the data set we used in our experiments. We conducted different experiments with four different sets of features. To evaluate our method, we used accuracy as the measure of prediction that can be defined as the number of documents whose author gender was correctly identified divided by the total number of documents in the testing set.

<table>
<thead>
<tr>
<th>Data set</th>
<th>male</th>
<th>female</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train data</td>
<td>358</td>
<td>296</td>
<td>654</td>
</tr>
<tr>
<td>Test data</td>
<td>24</td>
<td>21</td>
<td>45</td>
</tr>
</tbody>
</table>

5 Result and Discussion

Table 3 shows the accuracy of gender identification task trained on original documents and tested on original, obfuscated and imitated documents. We trained the classifiers only on original documents to imitate the real situation where no adversarial texts are available. Each result described in that table is an average of accuracy of five different experiments with different subsets of training data.

\(^1\) https://psal.cs.drexel.edu/index.php/Main_Page
This experiment setting is similar to 5-fold cross validation. Except for experiments on obfuscation and imitation documents, due to the limited number of the data, we used the same set of obfuscation/imitation documents as the test data set.

The best identification result produced by the SVM classifier with content words features of original documents, which implies the fact that the gender of the author of obfuscated/imitated documents are more difficult to identify than the original documents. Our identification results for original documents are roughly comparable to the results of previous gender identification work [3,6] which used considerably much larger data set. In most of the experiments, the SVM classifier performed better than the Naive Bayes and K-NN classifiers. The highlighted cells show the highest accuracy for each of the document type.

Table 3: Accuracy of gender identification on original, obfuscated and imitated documents using different feature subset

<table>
<thead>
<tr>
<th>Feature</th>
<th>original (%)</th>
<th>obfuscated (%)</th>
<th>imitated (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NB SVM K-NN</td>
<td>NB SVM K-NN</td>
<td>NB SVM K-NN</td>
</tr>
<tr>
<td>SF</td>
<td>62.22 70.22</td>
<td>69.78 48.89</td>
<td>54.67 48.89</td>
</tr>
<tr>
<td>FW</td>
<td>65.33 71.11</td>
<td>69.78 47.11</td>
<td>60.00 47.56</td>
</tr>
<tr>
<td>CW</td>
<td>80.88 82.67</td>
<td>72.44 44.44</td>
<td>37.78 57.33</td>
</tr>
<tr>
<td>C1</td>
<td>67.54 72.45</td>
<td>66.67 55.56</td>
<td>62.22 50.22</td>
</tr>
<tr>
<td>C2</td>
<td>60.00 69.33</td>
<td>57.78 48.89</td>
<td>56.44 47.56</td>
</tr>
</tbody>
</table>

In order to find out whether the differences in gender identification accuracy between document types are significant, a pairwise comparison test was performed using a one-way ANOVA with post-hoc Tukey HSD. In this test, we only used the result from the SVM classifier. The null hypothesis is that there is no difference in the performance of gender identification on two document types. We consider the performances to be significantly different from each other if the probability of accepting the null hypothesis is \( p < 0.05 \).

From the analysis, we concluded that there is a significant difference between the performance of gender identification task in original and obfuscated documents. Some combinations of features and classifiers were very effective for the original documents. For example, SVM with content words successfully identified the author’s gender with accuracy of 82.67%.

However, for obfuscated documents most of those feature-classifier combinations performed just slightly better than the baseline performance. The best accuracy of 62.22% was obtained when SVM and C1 feature combination were applied. On the other hand, result of the pairwise test also demonstrated that in imitated documents, gender identification task worked as well as in the original documents, particularly when function words (FW) or C2 feature combinations...
were used. Result of pairwise comparison test between document types is presented in Table 4.

<table>
<thead>
<tr>
<th>Pairwise</th>
<th>SF</th>
<th>FW</th>
<th>CW</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ori</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Obf</td>
<td>-</td>
<td>Δ</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

We also performed another pairwise comparison test to observe the significance difference between features. The result shown in Table 5 indicates that for each type of documents, each feature performed differently. For original documents, content words performed significantly better than other features. This is caused by the nature of the dataset. During the construction of Extended Brennan-Greenstadt corpus, particularly for original writing, the authors were free to choose any topics as long as it still considers to be formal writing. However, for obfuscated and imitated documents they were provided with a specific topic which makes these documents appear similar when using content words (CW) features.

The result also proved that function words (FW) are effective features for gender identification, particularly in imitated documents. Combining FW with other features also improves accuracy. As can be seen in Table 3 the best accuracy (81.33%) was obtained by applying a Naive Bayes classifier with C2 features. However, we believed that the high performance was also affected by the nature of the dataset itself. Imitated documents were created by asking the participants to write a narrative passage about their own daily activities from a third-person perspective. Thus, the frequent used of third-person pronouns (such as he, she, him and her) form a strong clue for identifying the author’s gender.

In contrast, in obfuscated documents, we found that there was no standout feature. Results indicated that the gender identification task seems not resistant to obfuscation attack, even though the accuracy is still slightly better over random chance.

Overall, the combination of function words and stylometric features (C1) appeared as the most stable features, producing best accuracy of 75.56% on imitated document and this figure only dropped slightly, to 62.22% and 72.45%, for obfuscated and original documents.

6 Conclusion

We conclude that identifying author’s gender in adversarial documents is more difficult than identifying gender in the original documents. However, our exper-
Table 5: Pairwise comparison test of gender identification accuracy between features. Significant differences are marked with asterisk (*) (p < 0.05), while ∆ means the difference is not significant (p > 0.05).

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Obfuscated</th>
<th>Imitated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FW</td>
<td>CW</td>
<td>C1</td>
</tr>
<tr>
<td>SF</td>
<td>∆</td>
<td>*</td>
<td>∆</td>
</tr>
<tr>
<td>FW</td>
<td>–</td>
<td>*</td>
<td>∆</td>
</tr>
<tr>
<td>CW</td>
<td>–</td>
<td>–</td>
<td>*</td>
</tr>
<tr>
<td>C1</td>
<td>–</td>
<td>∆</td>
<td>–</td>
</tr>
</tbody>
</table>

Experiments demonstrated that standard gender identification task is more resistant to imitation attacks than obfuscation attacks. The result shows that there is no significant difference between the performance of the identification task in original and imitated documents. In addition, the accuracy of identification task applied to both of adversarial writings is better than the baseline performance. This outcome is rather different to the result of the authorship attribution task which seems more vulnerable to adversarial attacks. As explained in section 3, previous authorship attribution applied on adversarial text achieved very low accuracy which is no better than random guessing. In terms of features, function words proved to be effective for identifying gender in this data set. Yet, the choice of features should also consider the nature of the dataset since this may affect performance.

Acknowledgments The author would like to thank Mark Stevenson for valuable discussions and helpful suggestions, and the anonymous ESSLII Student Session 2015 reviewers for their comments.

References

An Event-Based Ontology for
Sign Language Phonology

Ayça Müge Sevinç *

Cognitive Science, Informatics Institute,
Middle East Technical University, Turkey

Abstract. This paper presents an ontology which analyzes the domain of sign language phonology based on an autosegmental model, namely, Movement-Hold model. We re-use the event-based ontology of Bird and Klein, which constructs a common understanding of the structures in “autosegmental phonology”. By extending this temporal ontology to sign language, we assign appropriate meanings to autosegmental representations, and capture both simultaneous and sequential aspects of sign language phonology via temporal relations. Overall, we construct a mechanism that checks well-formedness of signs. This event-based ontology for sign language phonology and its knowledge base are implemented by using Protégé editor.

1 Introduction

This paper presents a domain ontology which specifies how signs are composed of meaningless units, namely phonemes. Based on evidence from minimal pairs, Stokoe states that handshape, orientation, location and movement are phonemes of a sign [17]. Klima and Bellugi claim that organization of phonemes are simultaneous rather than sequential [11]. Liddell and Johnson argue that there is also sequentiality in addition to simultaneity [13]. Sequentiality is observed during a change in handshape, orientation or non-manual markings. These simultaneous and sequential aspects of phonological representations are modeled by temporal relations in our ontology.

The general purpose of developing such an ontology is to share the common understanding of the structure of phonological representations among researchers from different fields such as computational linguistics, machine translation, lexicography and sign linguistics, and to make this information accessible to software agents for computational processing and analysis.

Sign language data are mostly recorded and stored in video format. For computational analysis of recorded signs, a notation system is needed. One of the

* I thank my supervisor, Cem Bozşahin, for suggesting one level phonology for SL. I express my thanks to Orkan Bayer for his help with Protégé and proof-reading the paper. Special thanks goes to Aziz Zambak for financial support, and all members of Laboratory for Computational Ontology who come and listen to my draft talk, especially to Sezen Altuğ for reading the final paper. I also express my deep gratitude to my reviewers for constructive suggestions and helpful comments.
most popular notation systems for sign language is Hamburg Notation System (HamNoSys)\(^1\) [8]. It is used by sign language generation systems as input to animation of signing avatars [5, 14].

HamNoSys is a phonetic transcription system which is similar to International Phonetic Alphabet (IPA). Its alphabet is based on Stokoe’s notation [17] in which the symbols represent handshapes, hand positions, locations, non-manuals, orientations and movements. HamNoSys is a linear representation of these symbols in which the symbols are accepted as simultaneous. Besides, HamNoSys reflects sequentiality taking place in handshape, orientation and location change. Even if HamNoSys has sequential and simultaneous flavor, HamNoSys is not a detailed representation that shows how these components of the signs are temporally related. Exact timing information of how these components come together to make up a sign, makes an enhancement in the performance of signing avatars which generate animation from HamNoSys representations [7].

Sign phonology models such as Movement-Hold model [13] and Hand-Tier model [16] are based on the “autosegmental phonology” [6]. Both models have a multi-tiered representation and suggest that there is a segmental tier consisting of segments which are associated with some phonological features located on the other tiers. They serve as better representatives of simultaneity and sequentiality.

We construct an event-based ontology [2] based on MH model [13], in which autosegmental representations are given appropriate meanings. This temporal ontology for sign language is based on time intervals, and it has the power of interpreting the relations between the autosegments as temporal relations. Having an ontology based on a phonological model, it is possible to test the assumptions made by this model, and to check if these assumptions are consistent. The purpose of this ontology is checking for well-formedness of signs. In the following sections, MH model [13] and event-based ontology [2] are presented, and implementation details of our ontology are explained.

2 Sign Language Phonology

Sign phonology models such as Movement-Hold model [13] and Hand-Tier model [16] are based on “autosegmental phonology” [6] framework from which these two models get their explanatory power for representing both sequentiality and simultaneity. Autosegmental phonology [6] is originally developed for modeling phonology of tonal languages. It is a multi-tiered phonological model in which similar phonological features are placed on the same tier and different features are placed on different tiers. The autosegmental tiers are linear sequences of autosegments. For example, in Ngbaka, a tonal language, tense is indicated by the specific tones which are placed in a different layer than the root morpheme ‘kpolo’ [10]. In Fig. 1, tones such as high pitch ‘H’, low pitch ‘L’ or medium pitch ‘M’ are autosegments on the tone tier, whereas tone bearing units are located in the root tier.

\(^{1}\) http://www.signlang.uni-hamburg.de/projects/hamnosys.html
Temporal relations between autosegments in different tiers are organized with association lines. The association of one autosegment in a tier with two or more elements in another tier is called spreading. A chart is a pair of tiers linked by association lines. In [13], Goldsmith defines the well-formedness condition for charts as follows:

**Definition 1. Well-formedness Condition:**

a. All vowels are associated with at least one tone.
b. All tones are associated with at least one vowel.
c. Association lines do not cross.

When features which are employed in autosegmental representations encode information about manner or place of articulation, this framework also explains “how articulators such as tongue, lips, velum and larynx are coordinated” to form a spoken word [6]. In the case of sign language models, features carry the information of how articulators such as fingers, palm, thumb, wrist and forearm work together to build a sign from its components.

Following Goldsmith [6], Movement-Hold [13] and Hand-Tier [16] models suggest autosegmental phonological representations for signs. Both models inherit the idea of timing skeletons [15]. MH model has a segmental tier consisting of movements (Ms) and holds (Hs) as its segments [13], this skeleton tier codes relative timing of the segments. Movement is defined as a time interval in which some articulatory features change. Hold is a segment in which the articulatory features are in a static state.

A segment has five main entries: a major class, a contour of movement, a local movement which is a rapidly repeated movement, a quality feature describing temporal or physical quality of the segment, and a contour plane upon which the hand moves. The values of these properties are given in Table 1.

In MH model [13], there are four articulatory tiers in addition to the segmental tier: hand configuration, place of contact, orientation and facing tiers. Hand configuration tier describes handshape, it is composed of following three articulatory features: forearm involvement, handshape and configuration of fingers. Place of contact tier describes hand’s contact with the body by employing four kinds of features: the handpart which touches the body, the location on the body which is contacted, spatial relation between the handpart and the location, and
Table 1. The properties that segments carry in Movement-Hold model [12]

<table>
<thead>
<tr>
<th>Major class</th>
<th>hold, movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contour of movement</td>
<td>straight, round, seven contour, arc , circle</td>
</tr>
<tr>
<td>Local movement</td>
<td>wiggling, circling, rubbing, hooking, twisting, flattening, releasing</td>
</tr>
<tr>
<td>Quality features</td>
<td>prolonged, shortened, accelerating, tense, reduced path, enlarged path and contacting</td>
</tr>
<tr>
<td>Contour plane</td>
<td>horizontal, vertical, surface, midsagittal, oblique</td>
</tr>
</tbody>
</table>

the *proximity* which describes how proximate the handpart is to the location. Facing is composed of a *handpart* and a *location* where the handpart points to the location. Orientation is made up of a *handpart* and a *plane*, and it indicates which handpart looks towards the ground (HP). Articulatory features are shown in Table 2.

Table 2. Articulatory features in Movement-Hold model [12]

| handshape | A, S, 1, !, I, Y, =, >, H, V, K, D, R, r, W, 6, 7 ,8 , F, 9, B, 4, T, N, M |
| finger configuration | o (open), op(closed)," (hooked), "(flattened) |
| handpart of contact | RAFI (radial side of the finger(s)), TIFI (tips of finger(s)), PDFI (pads of finger(s)) |
| proximity | p (proximal), m (medial), d (distal), c (contact) |
| spatial relation | ipsi, contra, over, under, behind, tipward, baseward, toward ulnar side, toward radial side, palmward, backward |
| location | BH (back of head), CN (chin), TH (top of head), NK (neck), FH (forehead), SH (shoulder), SF (side of forehead), LG (Leg), ST (sternum), NS (nose), CH (chest), CK (cheek), TR (trunk), ER (ear), UA (upper arm), MO (mouth), FA (forearm), LP (Lip), AB (abdomen), JW (jaw) |
| handpart | IN (inside), PD (pad), BK (back), RA (radial), UL (ulnar), TI (tips), KN (knuckle), BA (base), HL (heel), WB (web), PA (palm) |
| plane | HP (horizontal plane), VP (vertical plane), SP (surface plane), MP (midsagittal plane), OP (oblique plane) |

Figure 2 presents how segmental and articulatory features are associated to the timeline of the video recording of ASL sign LIKE and shows how temporal overlap between the elements of the tiers takes place in real time. This annotation is created by ELAN [9] which is perhaps the most frequently used tool to create, edit, visualize and search annotations for video recordings of sign languages.

Figure 3 demonstrates the corresponding autosegmental representation of LIKE in MH model [12]. This sign has HMH sequencing of segments. There is
a handshape change which is from open 8 handshape to closed 8 handshape. 8 is a handshape which has all but middle finger open and spread. The location of the sign is the chest. The sign starts with a contact (c) of pads of fingers (PDFI) to sternum (ST) and moves proximal (p) to ST and ahead of it. While moving ahead of the sternum, the tips of the selected fingers (TIFI) touch each other. The palm (PA) faces the surface plane (SP), in other words, the body plane. Orientation is defined as the relation between ulnar (UL) part of the hand and the ground (HP). There is no orientation change. There are no non-manuals accompanying the sign, hence non-manual tier is empty. As seen Fig.3, all elements in articulatory tiers are associated with segments in segmental tier.

Fig. 2. Annotation of the phonological representation of ASL sign LIKE with respect to Movement-Hold model [12]. It is annotated by using ELAN [9]. The video is taken from the web page: http://media.spreadthesign.com/video/ogv/13/48560.ogv

3 Event Based Ontology

Bird and Klein [2] design an ontology which constructs a common understanding of the structures in autosegmental phonology [6]. The autosegmental representations, which are transformed into state-labeled automata by one-level phonology [1], are assigned a formal temporal semantics [2]. In this temporal semantics,

2 The tier organization in Fig. 3 is slightly different than the tier organization in [13] where there are four articulatory tiers, and point of contact tier consists also the information about proximity, spatial relation, and location.
autosegments are accepted as *extended time intervals* associated with phonological properties (features or gestures). Event, since it is the interpretation of an autosegment, is the basic unit. Hence, this semantics is also called “*event-based ontology*” [2].

An autosegmental tier which has a set of autosegments in a linear order is interpreted as a *melody* which has an ordered set of events. The linear order among autosegments is interpreted as *temporal precedence* relation among the intervals of the events. Association lines between autosegments on different tiers are interpreted as *temporal overlap* relation [2]. *Phonological event structure* is the interpretation of an autosegmental representation which consists of multiple tiers. It has a set of melodies and two sets of pairs of intervals to indicate precedence and overlap relations.

In the rest of the paper, we explain the class hierarchy in Fig. 4^3^ and introduce properties of these classes, relations among them, and axioms defined over these relations. This ontology is developed in Web Ontology Language (OWL) and Semantic Web Rule Language (SWRL) by using Protégé editor[^4^]. In Table 3^5^, domain, range, OWL object property characteristics like functionality, inverse functionality, transitivity, symmetry, asymmetry, reflexivity, and irreflexivity are listed for each of the relations in Fig. 4. We organize the following sections as to explain the main classes: events, melodies, and phonological event structures.

---

[^3^]: ‘sc’ stands for subClassOf. Enumerated relations are explained in Table 3.
[^5^]: PES stands for PhonologicalEventStructure, xImmPrecedes is shortened form of xImmediatelyPrecedes.

---

218
Fig. 4. Class hierarchy for event-based ontology [2].

Table 3. Relations in Class Hierarchy

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>eHasProperty</td>
<td>Event</td>
<td>Property</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>eHasInterval</td>
<td>Event</td>
<td>Interval</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>iIntervalOf</td>
<td>Interval</td>
<td>Event</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>4</td>
<td>iImmPrecedes</td>
<td>Interval</td>
<td>Event</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>iPrecedes</td>
<td>Interval</td>
<td>Interval</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>iOverlaps</td>
<td>Interval</td>
<td>Interval</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>7</td>
<td>eImmPrecedes</td>
<td>Event</td>
<td>Event</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>8</td>
<td>nHasEvent</td>
<td>EventNode</td>
<td>Event</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>9</td>
<td>eEventOf</td>
<td>Event</td>
<td>EventNode</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>10</td>
<td>nImmPrecedes</td>
<td>EventNode</td>
<td>EventNode</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>11</td>
<td>nHasNextNode</td>
<td>EventNode</td>
<td>EventNode</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>12</td>
<td>mHasEventSet</td>
<td>Melody</td>
<td>EventNode</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>13</td>
<td>pesHasMelody</td>
<td>PES</td>
<td>Melody</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>14</td>
<td>tTypeOf</td>
<td>EType</td>
<td>Property</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>15</td>
<td>pHasType</td>
<td>Property</td>
<td>EType</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>16</td>
<td>eHasType</td>
<td>Event</td>
<td>EType</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>17</td>
<td>nHasType</td>
<td>EventNode</td>
<td>EventNode</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>18</td>
<td>mHasType</td>
<td>Melody</td>
<td>EType</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>19</td>
<td>pesHasType</td>
<td>PES</td>
<td>EType</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Based on MH model [13], we construct a knowledge base by deciding what an event, a melody or a phonological event structure is in this model, and what kind of phonological properties basic events have and lastly which basic types these properties have.
3.1 Events

Event is the interpretation of an autosegment. Bird and Klein [2] define an event \((E_\theta)\) as “an ordered pair \(<i, \pi_\theta >\theta\) consisting of an interval \((i)\) and a property \((\pi_\theta)\)”. The type of event is same as the type of its property, \(\theta\).

We define \(eHasInterval\) relation to attach events an extended time interval and \(eHasProperty\) to associate events with some phonological property. \(iIntervalOf\) is inverse of \(eHasInterval\). Self loops on Interval node in Fig. 4 represent \(iImmediatelyPrecedes\), \(iPrecedes\) and \(iOverlaps\) relations.

We include three subclasses under Property class: BasicProperty, Melody and PhonologicalEventStructure. Basic properties correspond to features or gestures; i.e., for spoken languages, they are things like phonemes, tones, demisyllables and articulatory features. Based on MH model [13, 12], we define 14 subclasses of BasicProperty class as in Table 4. We add individuals to each subclass by using the values of segmental and articulatory features in Table 1 and Table 2.

Atomic event is an event with a basic property. AtomicEvent is defined as a subclass of Event as follows: “Event that \(eHasProperty\) exactly 1 BasicProperty”. Some examples of atomic events are \(E_1=<i_1, 8hs >hs\), \(E_2=<i_2, ofc >fc\). \(E_1\) is an atomic event which has the property ‘8 handshape’ which is associated with interval \(i_1\). \(E_2\) has the property ‘open finger configuration’ which is realized in interval \(i_2\).

Melody class has an ordered set of events (\(mHasEventSet\)) and PhonologicalEventStructure has a set of melodies (\(pesHasMelody\)). EventSet is implemented as a linked list whose elements are represented by the EventNode class. Each EventNode has a content, namely an event (\(nHasEvent\)), and a next node (\(nHasNextNode\)). Definition of event is recursive (Fig. 4), but it is not cyclic [2].

This ontology is all typed. Property has a type (\(pHasType\)) and Etype is the type of a property (\(tTypeOf\)). The type of the property may be a basic type, or a set of basic types or a set of set of basic types [2]. The set of basic types is determined on the basis of the phonological properties that the language has. For our ontology based on MH model, we insert 13 individuals of BasicType class all of which are defined to be the types of basic properties (Table 4). We relate each basic property with a basic type by defining it as an equivalent class of “BasicProperty and pHasType value X”, where X is one of the basic type values in Table 4.

The type of event is same as its property’s type, \(eHasType\) is defined as a composite relation “\(eHasProperty o pHasType\)”. Type of melody is same as the type of its events, and type of phonological event structure is the set containing the types of its melodies. \(mHasType\) and \(pesHasType\) are also composite relations.

3.2 Melodies

Melody is the interpretation of a tier in an autosegmental representation. The order among autosegments in a tier is defined in terms of “temporal precedence” relation. Formally, Bird and Klein [2] define a melody \(\tau\) as “a pair \(<E, <^o>_\{\theta\}\)
Table 4. BasicProperty subclasses and their corresponding BasicType individuals

<table>
<thead>
<tr>
<th>BasicProperty</th>
<th>BasicType</th>
<th>BasicProperty</th>
<th>BasicType</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC-HandShape</td>
<td>hs</td>
<td>NMM-LIPS</td>
<td>nmm</td>
</tr>
<tr>
<td>HC-FingerConfiguration</td>
<td>fc</td>
<td>HandPart</td>
<td>hdp</td>
</tr>
<tr>
<td>POC-Proximity</td>
<td>cm</td>
<td>Plane</td>
<td>pl</td>
</tr>
<tr>
<td>POC-SpatialRelation</td>
<td>srel</td>
<td>SEG-LocalMovement</td>
<td>lm</td>
</tr>
<tr>
<td>POC-HandPart</td>
<td>chdp</td>
<td>SEG-Quality</td>
<td>qm</td>
</tr>
<tr>
<td>LOC-MajorBodyLocation</td>
<td>loc</td>
<td>SEG-ContourOfMovement</td>
<td>cm</td>
</tr>
<tr>
<td>LOC-SigningSpaceLocation</td>
<td>loc</td>
<td>SEG-MajorClass</td>
<td>mc</td>
</tr>
</tbody>
</table>

Consisting of an event set \( (E) \) and immediate precedence relation \( (\prec_o) \). All the elements of the event set have the same type \( \theta \). \( \prec_o \) is defined over the event set \( E \) as follows:

For any events \( x, y \in E \), \( x \) immediately precedes \( y \) \((x \prec_o y)\) iff \( x < y \) and there is no \( z \in E \) such that \( x < z < y \). This relation is

(i) asymmetric: \( \forall x y. x \prec_o y \leftrightarrow \neg(y \prec_o x) \)
(ii) irreflexive: \( \forall x. \neg(x \prec_o x) \)
(iii) intransitive: \( \forall x. (x \prec_o y \land y \prec_o z) \rightarrow \neg(x \prec_o z) \).

In our ontology, \( \text{iImmediatelyPrecedes} \) is defined on Interval class and marked as asymmetric and irreflexive as in Table 3. \( \text{iPrecedes} \) is defined as a super property of \( \text{iImmediatelyPrecedes} \) in the object property hierarchy. We define temporal relations among events (\( \text{eImmediatelyPrecedes}, \text{ePrecedes} \) and \( \text{iOverlaps} \)) based on their counterparts defined on intervals. Similarly, these temporal relations between event nodes in the melody are defined based on relations of their events.

Bird and Klein’s definition is equivalent to an ordered list of events. Rather than having a set, we prefer to define melody as a list. In OWL, there is “no built in support specifically for sequences or ordering” [4]. Hence, we implement melody as a linked list following [4]. Melody has the relation \( \text{mHasEventSet} \) which directs it to first EventNode in the linked list. We define EventNode class as having two main relations: \( \text{nHasEvent} \) and \( \text{nHasNextNode} \). The last node is marked with a subclass of EventNode class, MaxNode, which is defined to have no next node. The next node is defined to have the same type as current node in the SWRL rule: “\( \text{nHasNextNode(?en1, ?en2)}, \text{nHasType(?en1, ?t1)}, \text{nHasType(?en2, ?t2)} \rightarrow \text{SameAs (?t1, ?t2)} \)”.

Melody may have events, or may be empty. EmptyMelody is equivalent to Melody class with a maximum cardinality restriction of having zero event: “Melody and (\( \text{mHasEventSet} \) exactly 0 EventNode)”. We define AtomicMelody as a melody whose events are all atomic: “Melody and (\( \text{mHasEventSet} \) exactly some (EventNode and (\( \text{nHasEvent} \) only AtomicEvent)))”. An example of an atomic melody is the interpretation of the aperture change from open (o) to closed (op) finger configuration in Fig. 3. \( \tau = \langle \tau_1 : \text{o}fc \rangle, \langle \tau_2 : \text{op}fc \rangle, \langle \text{<o} \rangle \{fc\} \), where \( \text{<o} = \{\langle \tau_1, \tau_2 \rangle\} \).
In an event-based ontology based on Movement-Hold model [12], some other examples of possible melodies are the interpretations of H, M, HM, HM1, M2, M3H on segmental tier, where Ms and Hs are segments. For instance, a simplified interpretation of HMH is $\tau = \langle \langle \iota_1 : H_{mc} \rangle, \langle \iota_2 : M_{mc} \rangle, \langle \iota_3 : H_{mc} \rangle \rangle$, where $\prec = \{ \langle \iota_1, \iota_2 \rangle, \langle \iota_2, \iota_3 \rangle \}$ and $mc$ is type of SEG-MajorClass property.

Melodies are not always atomic in Movement-Hold model, subclasses of Melody such as SegmentalMelody, HandConfigurationMelody, FacingMelody, PointOfContactMelody and OrientationMelody are complex in the sense that they have an event set whose members have phonological event structures as their properties.

### 3.3 Phonological Event Structure

Autosegmental representations are interpreted as phonological event structures. "A phonological event structure $R$ is a triple $\{ \tau_0, \ldots, \tau_n \}, \prec, o \}$ where $\{ \tau_0, \ldots, \tau_n \}$ is the set of melodies in the event structure", $\prec$ and $o$ are precedence and overlap relations [2]. Melodies in an event structure have different types. A melody in an event structure may be empty. The association lines between the tiers are represented by overlap relation.

We define two subclasses of PhonologicalEventStructure class: SegmentCProperty and ArticulatoryCProperty. SegmentCProperty defines the properties of segmental events. ArticulatoryCProperty is either FacingCProperty, HandConfigurationCProperty, OrientationCProperty, or PointOfContactCProperty. These four classes define the properties of the events in the four main melodies described by Movement-Hold model[13]. For example, events on HandConfiguration melody are defined to have two basic properties: handshape (hs) and finger configuration (fc). We define HandConfigurationCProperty as 

## Axioms

The axioms in [A1]-[A5] are the minimal set of axioms which hold for overlap ($o$) and precedence ($\prec$) relations.

A1. $\forall x. xox$
A2. $\forall xy. xoy \rightarrow yox$
A3. $\forall xy. x \prec y \rightarrow \neg (y \prec x)$
A4. $\forall xy. x \prec y \rightarrow \neg (xoy)$
A5. $\forall wxyz. (w \prec x) \wedge (xoy) \wedge (y \prec z) \rightarrow (w \prec z)$

iOverlaps is a reflexive, non-transitive and symmetric relation. We declare reflexivity for Interval class as equivalent to ‘iOverlaps some Self’ (A1). We check only the box of symmetry (A2) for iOverlaps as in Table 3. iPrecedes is a transitive, asymmetric and irreflexive relation, however we only check the boxes of asymmetry (A3) and irreflexivity as in Table 3, because OWL has the limitation of not having transitive property with irreflexive and asymmetric property. We satisfy transitivity by inserting the SWRL rule “iPrecedes(?x, ?y),
iPrecedes(?y, ?z) → iPrecedes(?x, ?z). For irreflexivity of iPrecedes, we insert the SWRL rule “iPrecedes(?x, ?y) → DifferentFrom (?x, ?y)” (A3). iPrecedes and iOverlaps are defined to be disjoint (A4). In a similar manner, we define temporal relations eOverlaps, ePrecedes and eImmediatelyPrecedes. A5 defines well-formedness condition which states that association lines do not cross. We rephrase well-formedness condition in Defn. 1 to make it usable for signs and compatible with the terminology of Hold-Movement model [13].

**Definition 2.** Well-formedness condition for Signs

a. All articulatory events are associated with at least one segment.

b. All segments are associated with at least one articulatory event.

c. Association lines do not cross.

For first condition in Defn. 2, we define ArticulatoryEvent as “Event that (eHasProperty some ArticulatoryCProperty) and (eOverlaps min 1 Segment)” and add a SWRL rule: “ArticulatoryEvent(?e), Event(?e2), eOverlaps(?e, ?e2), DifferentFrom (?e, ?e2) → Segment(?e2)”. Condition 2 is handled in the same manner. We define Segment as “Event that (eHasProperty some SegmentCProperty) and (eOverlaps min 1 ArticulatoryEvent)”, and insert the rule: “Segment(?e), Event(?e2), eOverlaps(?e, ?e2), DifferentFrom (?e, ?e2) → ArticulatoryEvent(?e2)”. For third condition, we add the rule: “ePrecedes(?w, ?x), eOverlaps(?x, ?y), ePrecedes(?y, ?z), DifferentFrom (?x, ?y), DifferentFrom (?y, ?z), DifferentFrom (?z, ?w) → ePrecedes(?w, ?z)”.

4 Conclusion and Future Remarks

A computational ontology which covers the terminology of MH model [12], an autosegmental phonology model for sign languages, is developed. We re-use and extend the event-based ontology\(^6\) to cover autosegmental representations defined by MH model [13, 12] and show how the terms of MH model are to be expressed within an event-based ontology. Our ontology makes the domain knowledge of sign language phonology accessible to machines and makes it possible to test the assumptions made for this domain automatically. Our ontology is useful for making inferences about the phonological representations of signs and for checking the well-formedness of signs.

As future work, this ontology may be integrated to ELAN to make annotators use the same restricted terminology based on a specific sign language phonology model. As seen in Fig. 2, ELAN allows multiple tiers of information to be attached to the timeline of media, and these tiers may be subdivided into time segments which are linked to the annotations. This annotation strategy is very compatible with the autosegmental structure of MH model [12]. Ontology-based annotation is an idea suggested by ONTO-ELAN project [3], however their terminology is not relevant to the topic of phonology.

---

\(^6\) We re-implement their ontology based on the definitions in [2]. We use Protégé which is a free, open source ontology editor for the implementation.
References

Understanding belief bias by measuring prior beliefs for a Bayesian model of syllogistic reasoning

Michael Henry Tessler
Department of Psychology, Stanford University
mtessler@stanford.edu

Abstract. The phenomenon of belief bias in syllogistic reasoning occurs when the a priori believability of a conclusion influences the intuitive acceptability of that conclusion. Prior beliefs about the world can be formalized into a probabilistic generative model of situations. Tessler & Goodman proposed that this very idea can account for the range of acceptabilities of conclusions from categorical syllogisms with abstract content [15]. Here, I generalize their model to accommodate syllogistic reasoning data where content effects are observed. I collect data about the prior plausibility of various properties co-occurring, and use this data to predict syllogistic reasoning behavior in a separate experiment. I compare models with different types of assumptions concerning the prior and discuss open questions for this approach.

Your logic-chopping friend is in a room with a number of lamps and lightbulbs; you are in a different room and cannot see what she sees. She gives you the following logic puzzle:

All of the lightbulbs that are hot are bright.
Some of the lightbulbs that are bright are not on.

Are some of the hot lightbulbs not on? Are any of the hot ones on?

Prior beliefs about the world guide our actions, thoughts, and reasoning in new situations. It can be helpful, for example, to know how fast a particular kind of animal can run, if you are also thinking about if that animal can eat you. Similarly, humans can use prior beliefs in a domain (e.g. life expectancies) to reason accurately about everyday contexts (e.g. guessing how long someone will live) [10]. Finally, it has been argued that prior beliefs influence the very meaning of words [7]. It is odd then that so little formal theory has gone into understanding prior beliefs in classic reasoning tasks, but cf. [13,3].

Bayesian approaches to cognitive science have a natural way of accounting for prior beliefs in reasoning. Tessler & Goodman described a generative model of argument strength that uses a truth-functional semantics applied to idealized situations composed of objects with properties [15]. This model accounted for much of the variability in [1]'s meta-analysis data of categorical syllogistic
reasoning. This work was done with respect to meta-analysis data that differed largely in the materials used, and for which prior beliefs about the materials were not expected to have a substantial effect. However, it is known that prior expectations about the categories and properties at stake in a syllogism influence the acceptability of a conclusion [5, 2, 4].

Here, I generalize [15]'s model of argument strength to capture phenomena associated with belief bias. This is done by empirically measuring prior beliefs about real-world content, deriving model predictions based on those measured beliefs, and testing the probabilistic model of argument strength against behavioral data obtained in a separate experiment of syllogistic reasoning. A secondary, primarily methodological concern is about the granularity of information needed to capture these syllogistic reasoning phenomena.

To foreshadow the results, empirically measured priors (Expt. 1) coupled with a Bayesian model of argument strength accounts for much of the syllogistic reasoning data (Expt. 2), including qualitative effects of content. The predictions of the model, which has no parameters, are as good as those of a model parametrized by 12 variables. The most likely values (conditioned on the data of Expt. 2) of these 12 variables correspond roughly with the marginal distributions of the priors elicited in Expt. 1. This interesting correspondence suggests the syllogistic reasoning task is too coarse-grained to disambiguate models of reasoning about correlated properties from models where independence is assumed.

1 Bayesian argument strength in syllogistic reasoning

A formal account of gradience in syllogistic reasoning was presented by [15]. The computational model is a Bayesian model; as such, it is important to understand the implications of the prior for syllogistic reasoning. I review the model, highlighting along the way how I generalize the model to consider content effects.

1.1 Ontology

The model is based on an ontology of situations composed of objects with properties, similar to mental models [11]. A situation \( s \in S \) is composed of \( n \) objects: \( s = \{o_1, o_2, ..., o_n\} \), each of which can have 3 properties:

\[
s = \{\{A_{o_1}, B_{o_1}, C_{o_1}\}, \{A_{o_2}, B_{o_2}, C_{o_2}\}, ..., \{A_{o_n}, B_{o_n}, C_{o_n}\}\}
\]

Properties \( A, B, \) and \( C \) of these objects are stochastic and assumed to be Boolean for simplicity. Properties across objects are assumed to be independent and identically distributed (iid); hence,

\[
P(s) = \prod_{1 \leq i \leq n} P(A_{o_i}, B_{o_i}, C_{o_i}) = (P(a, b, c))^n
\]

To account for syllogistic reasoning in [1]'s meta-analysis of 5 studies, which differed with respect to the materials used, the model assumed no a priori
information about the meaning of the properties; thus, properties within objects were determined independently and identically (i.i.d.): 

\[ P(A_{oi}, B_{oi}, C_{oi}) = P(A_{oi}) \cdot P(B_{oi}) \cdot P(C_{oi}) = (P(p))^3, \]

with \( p \sim \text{Bernoulli}(\theta) \).

The number of objects in a situation \( n \) is a parameter of the model, as is the base rate \( \theta \) of properties. In fitting the model to meta-analysis data, [15] found \( \theta \approx 0.25 \), consistent with the “rarity assumption”—properties are relatively rare of objects—first used by [14]. The best fitting \( n \) was around 5, also consistent with the “minimal model assumption” of the Mental Models framework [11].

### 1.2 A generative model of argument strength

The generative model of situations can be turned into a generative model of syllogistic reasoning by providing a semantics for the quantifier sentences of a syllogism. The model uses the interpretation of quantifier sentences as truth-functional operators, consistent with standard practice in formal semantics.

A quantifier utterance (e.g. \( \text{all } A \text{ are } B \)) maps two properties (e.g. \( A \) and \( B \)) to a truth value by consulting the properties of the objects in the situation \( s \) and applying the usual literal meaning. For instance:

\[
\begin{align*}
[u_{\text{no } A \text{ are } B}] &= \{s \in S : \|o_A \cap o_B\| = 0\} \\
[u_{\text{some } A \text{ are } B}] &= \{s \in S : \|o_A \cap o_B\| > 0\} \\
[u_{\text{all } A \text{ are } B}] &= \{s \in S : \|o_A \cap o_B\| = n\} \\
[u_{\text{not all } A \text{ are } B}] &= \{s \in S : \|o_A \cap o_B\| < n\}
\end{align*}
\]

where \( o_A = \{o_i | A_{oi} = 1\} \) and \( o_B = \{o_i | B_{oi} = 1\} \) represent the objects in a situation that have the properties \( A \) and \( B \), respectively. Thus, the quantifier utterances pick out the situations \( s \) where the truth-functional meaning of the utterance is satisfied.

Truth-functional meanings of quantifier expressions are useful here because an expression which assigns a Boolean value to a situation can be used for probabilistic conditioning. That is, these quantifier expressions can be used to update a prior belief distribution over situations into a posterior belief distribution:

\[
P(s | u_1, u_2) \propto P(s) \cdot \delta[u_1][s] \cdot \delta[u_2][s]
\]

where \( u_1, u_2 \) are the two quantifier-utterances corresponding to the premises of a syllogism (e.g. \( u_{\text{all } A \text{ are } B}, u_{\text{some } B \text{ are not } C} \)).

For syllogistic reasoning, we are interested not in the posterior distribution over situations per se, but the distribution on true conclusions that these situations entail: \( P(u_3 | s) \), where \( u_3 \) is a quantifier-utterance corresponding to the conclusion of a syllogism (e.g. \( u_{\text{some } A \text{ are } C} \)). Hence,

\[
P(u_3 | u_1, u_2) \propto P(u_3 | s) \cdot P(s | u_1, u_2)
\]

This model, thus, returns a posterior distribution over conclusions conditioned on the premises of a syllogism being true.
The Bayesian model has a natural way of accounting for the influence of prior beliefs in reasoning. Indeed, beliefs simply specify a prior distribution over situations. In particular, the assumption that properties in a situation are independent and identically distributed (i.i.d.) must be relaxed if we are to consider real-world content. I generalize the model by considering that properties can have correlations; the representation of expectations about the presence or absence of properties will be generalized from one marginal distribution—$P(p) = P(a) = P(b) = P(c)$—to the joint distribution: $P(a,b,c)$.

The model was written in the probabilistic programming language WebPPL\textsuperscript{1} [9]. For background and details on this form of model representation, see http://probmods.org.

2 Experiment 1: Measuring $P(a, b, c)$ for real-world content

Bayesian models of reasoning and language typically measure the relevant prior distribution for a given task in order to generate predictions. For the model of syllogistic reasoning presented by [15], the natural prior to measure is the distribution over the properties mentioned in the syllogism. I constructed content domains to intuitively cover a range of probabilities.

Design I recruited 70 participants on Amazon’s Mechanical Turk to rate the likelihood of various combinations of properties co-occurring. Participants were paid $0.80 for their work.

To assess the reliability of the elicitation task, I ran the experiment using two different dependent measures as a between-subjects variable. Each participant was randomly assigned to either the “frequency” or the “plausibility” dependent measure condition (described below). Within each of these conditions, participants completed the judgment task for 4 content domains\textsuperscript{2}.

Procedure & Materials I selected property domains based on model simulations using qualitatively different priors (elicited from people in my lab). These preliminary simulations suggested that domains with causal structure led to the biggest differences between content domains (possibly due to probability estimates being more reliable for causal domains). Table 1 shows the properties.

The prompts for the “plausibility” condition read: Imagine an X (e.g. a lightbulb). How likely is it that it is ___? The prompts for the “frequency” condition read: Imagine 100 Xs (e.g. lightbulbs). About how many of them are ___? Below these prompts were listed the 8 possible combinations of the presence and

\textsuperscript{1} A fully-specified version of this model can be accessed at: http://forestdb.org/models/syllogisms-esslli2015.html

\textsuperscript{2} The experiment in full can be accessed at: http://stanford.edu/~mtessler/experiments/syllogism-belief-priors/prior-exp.html
absence of 3 properties (e.g. is on, is bright, and is hot). Next to each set of properties, was a slider bar. In the plausibility condition, the slider bar ranged from “Impossible” to “Certain”’. In the frequency condition, the slider bar ranged from “0” to “100”.

### Experiment 1 Domains

<table>
<thead>
<tr>
<th>Noun</th>
<th>Causal relation</th>
<th>Property A</th>
<th>Property B</th>
<th>Property C</th>
</tr>
</thead>
<tbody>
<tr>
<td>crackers</td>
<td>common effect</td>
<td>are soggy</td>
<td>are past expiration date</td>
<td>have lots of flavor</td>
</tr>
<tr>
<td>knives</td>
<td>common effect</td>
<td>are soggy</td>
<td>are rusty</td>
<td>cut well</td>
</tr>
<tr>
<td>light bulbs</td>
<td>common cause</td>
<td>are on</td>
<td>are bright</td>
<td>are hot</td>
</tr>
<tr>
<td>strawberries</td>
<td>common cause</td>
<td>are in the freezer</td>
<td>are soft</td>
<td>are warm</td>
</tr>
</tbody>
</table>

**Table 1.** Content domains used in experiments.

**Data analysis and results** Participants’ responses were normalized within each domain so that the ratings for the 8 property combinations made a well-formed probability distribution (i.e. they added up to 1). I then took the mean rating for each of the 8 property combinations in each of the 4 domains, to arrive at mean empirical priors for all 4 domains. These were used as the empirical \( P(a, b, c) \) for the Bayesian model.

The experiment elicited unique priors for each domain (see Figure 1). The data elicited with different dependent measures were highly correlated (\( r_{\text{pearson}} = 0.78; r_{\text{spearman}} = 0.85 \)). Though the correlation between the prior data elicited by different dependent measures is good, the data set as a whole was substantially more reliable (95% bootstrapped CI for \( r_{\text{split-half}} = [0.95, 0.98] \)), suggesting meaningful differences between the two measurements. At the same time, model predictions based on the different dependent measures were also substantially more reliable (\( r_{\text{pearson}} = 0.95 \)). This suggests that the prior elicitation task captured the relevant variance for the syllogistic reasoning model. For simplicity, I later present the predictions of the reasoning model based on collapsing the prior elicitation ratings across dependent measures, though predictions based on either dependent measure alone are not meaningfully different.

### 3 Experiment 2: Syllogistic reasoning about real world content

In this experiment, I tested if the real world content from Experiment 1 influenced the conclusions drawn from categorical syllogisms.

**Design** I recruited 254 participants from Amazon’s Mechanical Turk. All participants were required to have a 95% approval rating for their previous work on the web service. Participants were paid $0.60 for their work. Each syllogism
Fig. 1. Mean elicited priors collapsed across dependent measure (see text for details). Error bars denote 95% confidence intervals. X-axis shows presence or absence of each property, the ordering of which can be found in Table 1. For example, the tallest bar in the cracker domain (001) is a cracker which isn’t soggy, isn’t past expiration date, and has lots of flavors.

was paired with each domain used in Experiment 1. A total of 8 syllogisms were used, resulting in 32 unique {syllogism, domain} pairs.

Procedures & Materials Each participant completed 4 syllogisms. On each experimental trial, participants were presented with the syllogism (e.g. Some of the lightbulbs that are bright are on. None of the lightbulbs that are hot are bright.) and each possible conclusion (e.g. {All, Some, Not all, None} of the lightbulbs that are hot are on.) and asked Does it follow that: X, for each conclusion. Radio buttons with the options “Doesn’t follow” and “Follows” were provided. Below that was a vertically-oriented slider bar to measure confidence. Participants were required to mark each conclusion before continuing to the next trial3.

Results Shown in Figure 2 (lighter bars) are a subset of the results of the experiment. Content effects can be observed by comparing panels within a row (i.e. comparing across columns). For example, for the some / none syllogism (top row), the proportion of responses endorsing “Some of the lightbulbs that are hot are on” is appreciably higher than the proportion endorsing “Some of the crackers that have lots of flavor are soggy” (2nd row, columns 1 & 3: some conclusion). Effects of reasoning (i.e. of syllogism) can be observed by comparing panels within a column (i.e comparing down rows). For example, “Some of the crackers that have lots of flavor are soggy” is a substantially more endorsed conclusion if the premises are: “All of the crackers that are past expiration date

3 The experiment in full can be viewed at http://stanford.edu/~mtessler/experiments/syllogism-belief/syllbelief-exp2.html
are soggy. Some of the crackers that have lots of flavor are past expiration date.” (4th column, rows 1 & 2; some conclusion).

![Fig. 2. Reasoning patterns and predictions for 3 (of the 8) syllogisms in the experiment. Human reasoning data is in the darkest shade. The lighter shade is the 12-parameter (“independent”) model. The medium shade is the 0-parameter (“empirical prior”) model. All syllogisms shown here were of the form B-A / C-B and the conclusions were of the form C-A (e.g. First row: All B are A, Some C are B). Reasoning effects can be seen by comparing rows within a column. Content effects can be seen by comparing columns within a row.]

4 Bayesian analysis of Bayesian reasoning models

In this section, I explore 4 different models of argument strength that vary in their independence assumptions in the prior\(^4\). The first model—the “abstract” model—uses a single base-rate parameter in its prior\(^5\); this model assumes properties are \emph{i.i.d.} both within domains and across domains. The second model—the “within-i.i.d.” model—is the simplest parametrized model that can predict content effects; it is identical to the first model except in that the base-rate parameter can vary across domains (but not within a domain; hence it is \emph{i.i.d.} within but not across domains). This model has 4 parameters (one base-rate for each domain). The third model—the “fully independent” model—assumes only that

\(^4\) For all of the models I consider, properties are assumed to be independent \emph{across} objects (e.g. \(o_1\) having property A does not influence \(o_2\)’s chance of having property A). It is the assumption of independence \emph{within} objects that is explored in this paper.

\(^5\) This is the model of argument strength used by [15] to model meta-analysis data.
properties of an object are independent (not necessarily identically distributed); it uses a different base-rate parameter for each property within and across domains. This model has 12 parameters (3 properties per domain and 4 domains in total). The final model is a model that uses the empirically elicited priors from Expt. 1; this model has no free variables parametrizing the prior over properties.

One parameter is shared by all models. This is the number of objects in a situation \( n \), which controls the size of the worlds reasoned over. A preliminary analysis revealed that results were highly consistent for \( n \geq 4 \) and so I use \( n = 4 \) for all simulations.

I analyze the models by putting uninformative priors (\( \theta \sim \text{Uniform}(0,1) \)) on the base-rate parameter(s) and conditioning on the observed experimental data. In addition, for all models I include a single data analytic “guessing” parameter \( \phi \sim \text{Uniform}(0,1) \) to account for response noise. This noise parameter is important to accommodate data points that deviate largely from what the reasoning model predicts.

Inference for the combined Bayesian data analysis of the Bayesian reasoning model was done via the Metropolis-Hastings algorithm implemented in the probabilistic programming language WebPPL [9]. Models were analyzed independently, and MCMC chains were run for 10,000 samples.

4.1 Posteriors over model parameters

The posterior value for \( \phi \), the guessing variable, was near 0.25 for all models; this analysis attributes about a quarter of the responses to noise. Note that this estimate of noise is with respect to the reasoning model. In other words, it is an estimate of the proportion of responses better explained by random guessing than by the reasoning model. If the posterior predictive distribution of the model predicts the data well, \( \phi \) would be an accurate measure of the response noise in the syllogistic reasoning task. Alternatively, \( \phi \) could also include aspects of the data set that these particular reasoning models do not predict well.

One hypothesis about the base rate parameters of the parametrized-prior models is that the parameter values that account best for the reasoning data are lower-order representations of the empirically-elicited prior. To test this, I compared the posterior distributions over the base rate parameters of the 12-parameter, “fully independent” model to the marginal distributions of the empirical prior data (Figure 3). For the properties corresponding the conclusion terms of the syllogism (properties A & C), the inferred base-rates based on the

---

6 In some cases, this parameter is actually necessary to analyze the data. This is the case with logically impossible conclusions (e.g. All A are B // All B are C ∴ No A are C); in this case, the reasoning model gives this conclusion probability 0 (i.e. logically impossible means probability 0). If, for whatever reason, the experimental data includes this conclusion as a response, the data analysis model will crash because that particular data point is expected to have probability 0. The guessing parameter lets us accommodate any data point. This is done by postulating that with \( \phi \) probability, the participant selects a conclusion at random. I put a distribution over the probability and infer this value from the experimental data.
12-parameter model are qualitatively consistent with the marginal distributions from Experiment 1. For example, the most likely base rates for property A for the cracker and strawberry domains are definitively smaller than those of the knife and lightbulb domains (Figure 3, column 1, rows 1 & 4 vs. rows 2 & 3). As well, the most likely base rate for property C of the strawberry domain is smaller than the other 3 domains. The inferred base rates from the 12-parameter model have the most uncertainty about property B, possibly because B is only indirectly related to the responses, which are statements about properties A & C. Overall, this is suggestive evidence that the parametrized model of argument strength is using base rates corresponding to those of the marginal distributions over properties elicited in Expt. 1.

Fig. 3. Marginal posterior distributions over the base rates of properties in the 12-parameter, fully independent model. 95% CIs for the marginal distributions derived from the empirical priors elicited in Experiment 1 are shown at the top of each plot. Qualitative consistencies between the two suggest the base rates inferred from the reasoning data resemble lower-order statistics (the marginal distributions) of the richer joint-distributions shown in Figure 1.

4.2 Posterior predictives

The posterior predictive distribution marginalizes over the inferred parameter values to produce predictions about what the data should look like given the posited reasoning model and the observed data. This is akin to fitting the parameters and is an important step in model validation as it shows what data is actually predicted by the model. All of the models did considerably well in accounting for the variance in the syllogistic reasoning data. Table 2 shows the model–data correlations for each of the models. Figure 4 shows the fit for the empirical prior model.
Figure 2 shows predictions for the empirical prior model (darker shade) together with the experimental data (lighter shade) for 3 example syllogisms. The model based on the empirical prior shows effects of content. For example, the knives and lightbulbs domains show lower endorsements for the not all conclusion relative to the crackers and strawberries domains (and the reverse can be observed for the all conclusion), consistent with participants’ reasoning behavior (Figure 2, row 1: columns 2 & 3 vs. 1 & 4, not all conclusion). In addition, the model shows effects of reasoning. For example, the endorsement for the some conclusion is much higher for the all / some premises than for the some / none premises, also consistent with the experimental data (Figure 2, column 3: rows 1 vs. 2, some conclusion). Further, an interaction between content and reasoning can be observed by comparing the lightbulb domain to the knife domain for those same syllogisms (columns 3 vs. 2 X rows 1 vs. 2). In the knife domain, the experimental data shows the effects of the syllogism are much weaker (the light bars are not very different from one another). The model predictions are also very similar for these two syllogisms in the knife domain.

Finally, it’s worth drawing attention to the bottom row of Figure 2, the all / all syllogism. This is a valid syllogism (a maximally strong argument) with two
logically valid conclusions: all and some. Participants show a consistent preference for the all conclusion over the some conclusion, consistent with many other studies of syllogistic reasoning [12]. It’s interesting that this asymmetry is robust across the different content domains. The model predicts consistent responses across the content domains because the syllogistic argument is so strong. However, it has no way to capture the asymmetry between all and some because it is using only a truth-functional semantics (all entails some). [15] predicted this asymmetry by extending to the model to take into account pragmatic reasoning [8]. I leave for future work the incorporation of pragmatic reasoning with syllogistic arguments about real-world content.

5 Discussion

I have demonstrated that a model of syllogistic reasoning with a rich prior distribution can account for the flexibility of interpretation of a syllogistic argument with real world content. The phenomenon of “belief bias” can be viewed in this framework as a natural extension of the notion of “argument strength”. Arguments vary in strength depending on the prior probabilities of the properties in question.

This modeling work reveals that the empirical prior model predicts the data well. Using Bayesian data analytic techniques, I observed that the 4-parameter “within-iid” model and the 12-parameter “fully independent” model can also accommodate the content effects well. I say the models accommodate the data because their predictions are dependent on the particular parameter settings inferred from that data. The empirical prior model, by contrast, predicts the data with no parameter fitting. Additionally, it is likely that performing a formal, Bayesian model comparison between these models would favor the empirical prior model due to Bayes’ Occam’s Razor. However, it is interesting to consider the implications of these modeling results as they stand.

The 4-parameter “within-iid” model is the simplest model that could possibly account for content effects. What this model posits is that there is some difference in the base rates of properties in these four different domains. The posterior over base rate parameters for this model showed that domains with properties that tended to co-occur (e.g. the lightbulbs domain) had relatively high base rates while those for domains with properties that tended not to co-occur (e.g. the crackers domain) had low base rates. The 12-parameter “fully independent” model also accommodates the data well. This model posits that properties in a given domain are independent but differ in their base rates. Correlations between properties need not be tracked explicitly.

An alternative explanation for the ubiquitous good fits is that the experiment itself is confounded. The 8 syllogisms used in my experiment might not be the best syllogisms to distinguish models with subtly different independence assumptions about the priors over properties. I found that the most likely base rate parameter values for the parametrized prior models given the data from Expt. 2 were those that roughly corresponded to the marginal distributions of
properties from Expt. 1. Both models (parametrized priors and empirical priors) modeled the data equally well, suggesting that these experiments were not well suited to disambiguate them. An even more radical proposal is that categorical syllogistic reasoning in general is not the best kind of experiment to distinguish these models. In categorical syllogisms, there are only 3 logical possibilities for conclusions entailed by situations: all, some and not all, or none. It’s possible that these models would make different predictions if we allowed more possible responses, e.g. most and few, exact numbers.

Finally, it is worth noting that classical reasoning experiments such as the one explored here use language to communicate information for participants to reason over. Basic communicative principles, however, are often ignored in the scientist’s analysis of behavior. [15] formalized basic communicative principles in their probabilistic pragmatics model of syllogistic reasoning, as an extension of the Rational Speech-Act theory of language understanding [8]. I leave for future work the interplay between pragmatics and prior beliefs in the syllogistic domain.

References

Automatic alignment using Pair Hidden Markov Models

Johannes Wahle

Seminar für Sprachwissenschaft
Eberhard-Karls Universität Tübingen
johannes.wahle@uni-tuebingen.de

Abstract. This paper shows that Pair Hidden Markov models need to be taken more into consideration for applications in historical linguistics. One important task in historical linguistics is the comparison of two possibly related words. To address this issue historical linguists use alignments. Pair Hidden Markov models are capable of producing such alignments. Their probabilistic nature allows a more careful interpretation of the resulting alignments for phylogenetic applications. This paper compares the alignment performance of Pair Hidden Markov models with the classical Needleman-Wunsch algorithm and the alignment algorithm provided by LingPy. The results show that the Pair Hidden Markov model outperforms both approaches.

1 Introduction

The comparison of sequences of any kind is an important task for several fields of science, computational biology and historical linguistics being only two of those. Historical linguistics uses the comparison of words from different languages, i.e. sequences of sounds, to draw conclusions about the degree of relationship of the respective languages. The comparison of sequences is done on the basis of alignments. An alignment of two sequences tries to match those sounds which are historically related with each other and match inserted sounds with a specific gap symbol. Such an alignment as it is provided by Jäger (2013) [10] is shown in (1).

(1) horn
    kornu

In computational biology, the comparison of DNA or protein sequences is used as a basis for the inference of ancestry relations. Apart from the classical sequence comparison algorithms like the Needleman-Wunsch algorithm [17] computational biologists have brought up Pair Hidden Markov models [5]. While Hidden Markov models operate on one sequence of observation and a set of hidden states, a Pair Hidden Markov model operates on two sequences of observations and a set of

* The author has been supported by the ERC Advanced Grant 324246 EVOLAEMP, which is gratefully acknowledged.
hidden states. Thus, Pair Hidden Markov models model a stochastic process based on two sequences instead of single sequence.

In recent years, the automatic phylogenetic inference about the history and relationship of languages has gained more and more interest. In general, there are two basic approaches which are currently used in this field of study. The character based method [7, 2, among others] classifies a language by a set of features. Such features are, for example, the presence or absence of a particular cognate class or a certain grammatical property. The amount of shared features between two languages is interpreted as a measure of relatedness. The distance based method [9, 10, among others], which draws on the extensive literature in bioinformatics, c.f. the overview in Felsenstein (2004) [6], calculates a pairwise distance between languages. Often on the base of word lists, the distance between two languages is calculated as the distance between the respective word lists. These distances are estimated on the basis of alignments. Accordingly, a short distance should indicate a close relationship. This approach is appropriate for raw data which is not organized in a feature matrix. Pair Hidden Markov models offer an alternative approach to the calculation of distances and alignments. In order to see whether or not Pair Hidden Markov models should be considered for further applications, their performance in an alignment task is compared with other alignment algorithms.

2 Pair Hidden Markov Models

A Pair Hidden Markov model (pHMM) can be defined by the six-tuple \( \lambda = (S, K, \Pi, A, B, B') \), with \( S \) as the set of states, \( K \) as the output alphabet, \( \Pi \) as the initial state probabilities, \( A \) as the state transition probabilities, \( B \) as the emission probabilities for pairs of symbols, and \( B' \) as the probabilities of emitting a gap and a symbol as a pair. A classical pHMM makes use of five states in total. One state (\( M \)) which emits pairs of symbols, one state (\( X \)) which emits a symbol in the first string and a gap in the second string, one state which does the reverse (\( Y \)), one begin and one end state. This general set-up is common to both, computational biology and historical linguistics. An example for a pHMM as it is used in historical linguistics is shown in Fig. 1. In contrast to its relatives in computational biology, the pHMMs in historical linguistics allow a transition between the states \( X \) and \( Y \) (dashed line). This additional transition is necessary since alignments as shown in (2) are a common phenomenon. This example shows the alignment between the Italian (‘due’) and the Spanish (‘dos’) word for ‘two’. One would not hesitate to align the first two sounds of both words, but there is no evidence for a historical sound correspondence between the Italian ‘e’ and the Spanish ‘s’. So it is most suitable to align both sounds to a gap [16].

\[
\begin{align*}
\text{d} & \quad \text{u} & \quad \text{e} & - \\
\text{d} & \quad \text{o} & - \quad \text{s}
\end{align*}
\]

Such an alignment can not be generated by a pHMM which does not allow a transition between the states \( X \) and \( Y \). As well in contrast to the original
Fig. 1. Pair Hidden Markov model as proposed by Mackay & Kondrak [16]. The state $M$ emits aligned pairs of symbols, the state $X$ emits a symbol in the first sequence and a gap in the second sequence, vice versa for state $Y$.

The linguistic pHMM uses two different probabilities $\tau_M$ and $\tau_{XY}$ for the transition into the end state. This split of parameters enables the model to distinguish between the match state ($M$) being the final state or any of the gap states ($X, Y$). This modification preserves the symmetry of the model, while allowing a little bit more freedom. The linguistic version of pHMMs have already been successful in cognate detection and word similarity tasks [16] and dialectological applications [23].

Durbin et. al. [5] show how a pHMM can be derived from a finite state automaton (FSA) for affine gap alignment. It is possible to construct a FSA which mirrors a standard dynamic programming algorithm for affine gap alignment. The dynamic programming algorithm for the affine gap alignment calculates the optimal alignment of two sequences, thus it computes the optimal state path through the FSA. By converting the FSA into a Hidden Markov model, the model can be used as the basis for a probabilistic interpretation of an alignment.

For this conversion to happen, the parts of the dynamic programming have to be replaced by their probabilistic counterpart, i.e. transition probabilities and emission probabilities instead of fixed scores. This process results in a pHMM as described above. These transformations, however, maintain the characteristics of the alignment algorithm, which is to find the optimal state path. In terms of the famous three questions for Hidden Markov models as proposed by Rabiner (1989) [19], this is to find the most likely state path through a Hidden Markov model. This is exactly what the Viterbi-algorithm does for standard Hidden Markov
models, but in this case with an added dimension to account for two sequences. In this fashion, the traditional Forward-, Backward- and Baum-Welch-algorithm are adapted to the needs of pHMMs as well [5].

3 Experiments

The performance of pHMMs in carrying out alignments was tested on the basis of the data developed by Covington [3].\(^1\) Two different sets of parameters were estimated using the Baum-Welch algorithm and tested. All alignments were evaluated against the proposed alignments in the original data. These results were compared to the results of two other alignment algorithms. The test set consists of all word pairs from German, English, French and Latin. Furthermore, all words in the test set were translated from their original IPA encoding into the encoding of “The Automated Similarity Judgment Program” (ASJP) database [22].

For this study the parameters are estimated from a subset of the ASJP database [22]. The ASJP database provides 40 item Swadesh lists for over 6000 languages. From the ASJP database all Indo-European languages are extracted, except those belonging to the Germanic and Romance language families. These language families were excluded to minimize the chance of the same words in the test and training set. For all the remaining languages all words describing the same concept are paired together. If their normalized Levenstein distance was smaller than 0.7, they were treated as probable cognates and thus included into the training set. Actually, a parameter of 0.7 for the normalized Levenstein distance does not require more than one match for two words of length four. A stricter parameter would have resulted in a much smaller training set, which bears the risk of overfitting. This results in 833627 word pairs in the training set. To prevent the influence of the order of the training data, the mirror of each word pair was also included into the training set, i.e. the pair \((\text{word}_1, \text{word}_2)\) as well as \((\text{word}_2, \text{word}_1)\).

Based on this data two different sets of parameters were estimated. Each set of parameters consists of 1600 substitution probabilities, 80 insertion and deletion probabilities and 5 transition probabilities. The first, uninformed set of parameters was estimated on the basis of a uniform distribution of initial weights for emission probabilities.\(^2\) The second, informed set of parameters was estimated on the basis of a non-uniform distribution of initial weights.\(^3\) The weights were selected to mirror the assignment of sounds to the ten sound classes proposed by Dolgopolsky [4] (see Tab. 1). The probability of a match between sounds from the same sound class is higher than the probability of a match between sounds

\(^1\) This study uses the data as provided by List & Prokić [15].

\(^2\) The probability to replace one sound with another was \(\frac{1}{|K|^{2}}\) and the probability of a gap in one string and a symbol in the other was \(\frac{1}{|K|}\).

\(^3\) Such non-uniform distributions are common in computational biology. Those weights are based on empirical observation; c.f. Henikoff and Henikoff (1992) [8] for a prominent example.
Table 1. Dolgopolovsky’s sound classes [4] as provided by List (2012) [12]

<table>
<thead>
<tr>
<th>class description</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>P labial obstruents</td>
<td>p, b, f</td>
</tr>
<tr>
<td>T dental obstruents</td>
<td>d, t, 0, 0, δ</td>
</tr>
<tr>
<td>S sibilants</td>
<td>s, z, j, 3</td>
</tr>
<tr>
<td>K velar obstruents, dental and alveolar fricatives</td>
<td>k, g, ts, tf</td>
</tr>
<tr>
<td>M labial nasal</td>
<td>m</td>
</tr>
<tr>
<td>N remaining nasals</td>
<td>n, ŋ</td>
</tr>
<tr>
<td>R liquids</td>
<td>R, l</td>
</tr>
<tr>
<td>W voiced labial fricative and initial rounded vowels</td>
<td>v, u</td>
</tr>
<tr>
<td>J palatal approximant</td>
<td>j</td>
</tr>
<tr>
<td>∅ laryngeals and initial velar nasal</td>
<td>h, fi</td>
</tr>
</tbody>
</table>

from different sound classes. Based on the work of List [12, 11] the symbols of the alphabet in this study are assigned to one of the sound classes. This prior distribution makes it five-times more probable to match two sounds from the same sound class than to match it with a sound from another sound class. It is also ten times as probable to match a sound with itself then to match it with a sound from a different sound-class. This relaxes the original assumption of Dolgopolovsky where a match of sounds from different sound classes is forbidden. Although these new weights do not ban such a match, they make them very improbable. The probabilities of a gap and symbol remain the same as in the first set. The pHMMs were evaluated against the performance of the Needleman-Wunsch-algorithm [17] and the Sound-Class-Based Phonetic alignment (SCA) method [11].

3.1 The SCA-Method

The SCA-method, which is a part of the LingPy library [13, 14], tries to exploit the basic procedure of manually created alignments. Whilst the SCA-method uses a simple pairwise alignment algorithm, it models the sequences which are going to be aligned in a different way. There are two main aspects which can be considered in this approach, the paradigmatic and the syntagmatic aspect.

The paradigmatic aspect exploits the idea introduced by A. B. Dolgopolovsky [4], which is to organize sounds into different sound classes (see Tab. 1). The SCA approach extends the idea of sound classes to a total of 28 sound classes which are also suitable for an analysis of tone languages. Thus, the sounds of a word are coded as their respective sound class. In the model as proposed by Dolgopolovsky [4] a match between different sound classes was forbidden. The approach taken by the SCA-method relaxes this assumption and introduces a scoring function to code probabilities of divergent matches. These probabilities are estimated based on a weighted directed graph. The direction of the edges mirror the direction of a change and the weights the respective probability. A small weight indicates a
high probability of change. Based on this graph the scoring scheme is calculated [11].

The syntagmatic aspect focuses on the structure of the sequences. As for the SCA-method, prosodic profiles and secondary sequence structures are the focus of attention. The method distinguishes between seven different prosodic positions which are hierarchically ordered. In addition, each segment is assigned a sonority score based on the sonority hierarchy. These two measures are combined into a relative weight. These relative weights influence the probability of insertions or deletions in such a way that it is easier to introduce a gap in a position with a small relative weight. The aspect of secondary structure takes the syllable structure of the word into account by maintaining this structure in the alignment.

In contrast to many other alignment methods, the SCA-method can be used without parameter estimation and is thus suitable for applications with a limited amount of data. Results have shown, that this method outperforms other alignment approaches [11].

4 Results

The trained pHMMs were tested on the data described above [3, 15] and compared with the performance of the SCA-algorithm [11]. As a reference, the performance of the traditional Needleman-Wunsch algorithm is evaluated as well. Using the column score (c score), the sum-of-pairs score (sp score) and the Jaccard score as provided by the LingPy package [13, 14] the alignments were evaluated. Column and sum-of-pairs score are standard measures in evolutionary biology [20].

The sp score is calculated as the number of correctly aligned pairs divided by the pairwise score of the reference alignment, thus a high number of correct alignments results in a high sp score. For each residue $i$ in a test sequence of length $M$ the score $S_i$ is calculated, with $S_i = 1$ if the alignment of residue $i$ is correctly recognized, else $S_i = 0$ [20]. This score is then divided by the score for the reference alignment, where for each residue $i$ in the reference sequence of length $M_r$ the score $S_{ri}$ is calculated in the same manner as described above.

$$\text{sp score} = \frac{\sum_{i=1}^{M} S_i}{\sum_{i=1}^{M_r} S_{ri}} \quad (1)$$

The c score is calculated as the number of correctly aligned columns divided by the total number of columns $M$ in the test alignment, where $C_i = 1$ if all the residues in column $i$ are aligned as in the reference alignment, else $C_i = 0$ [20].

$$\text{c score} = \frac{\sum_{i=1}^{M} C_i}{M} \quad (2)$$

The Jaccard score is calculated as the size of the intersection of residue pairs in test $RP_T$ and reference alignment $RP_G$ divided by their union [1, 12].

$$\text{jaccard score} = \frac{|RP_T \cap RP_G|}{|RP_T \cup RP_G|} \quad (3)$$
Table 2. SCA shows the result of the alignment carried out with LingPy [11], NW show the performance of the Needleman-Wunsch-algorithm, pHMM 1 shows the result of the uniform weights and pHMM 2 shows the result of the non-uniform weights.

<table>
<thead>
<tr>
<th></th>
<th>pHMM 1</th>
<th>pHMM 2</th>
<th>SCA</th>
<th>NW</th>
</tr>
</thead>
<tbody>
<tr>
<td>c score</td>
<td>0.7199</td>
<td>0.7199</td>
<td>0.6</td>
<td>0.3176</td>
</tr>
<tr>
<td>sp score</td>
<td>0.8051</td>
<td>0.8051</td>
<td>0.6974</td>
<td>0.7179</td>
</tr>
<tr>
<td>jaccard-score</td>
<td>0.6382</td>
<td>0.6382</td>
<td>0.4875</td>
<td>0.3310</td>
</tr>
</tbody>
</table>

Using these three measures, Tab. 2 displays a comparison of the different alignment performances. The pHMM outperforms the LingPy alignment as well as the classical Needleman-Wunsch-algorithm on the test set. As the table also shows, the different initial weights did not have any effect on the performance of the pHMM in the alignment task. The signal in the data heavily outweighs the information of the initial weights. The margin by which the probability of matching sounds from different sound classes differs from the probability of matching sounds from the same sound class is nowhere near the difference between the probabilities in the initial weights. In fact, the results show that it is often up to 10-15 times more probable to match two sounds from the same sound class compared to two sounds from different sound classes. This phenomenon can be explained by the very homogeneous training data. During the construction of the training data, potentially troublesome word pairs are already eliminated. Therefore, the extra information at the beginning of the training is not necessary any more.

To address the issue of the same performance of the pHMMs the final parameters of the pHMM are analysed further. Figure 2 shows a hierarchical clustering of the sound substitution probabilities.\(^4\) The clustering shows a main split between the vowels and the consonants. The consonants are further divided into ten smaller classes which largely correspond to observable clusters; e.g. labial/bilabial obstruents p, b, f and labial/bilabial sonorants m, w, v. One exception to this pattern is the position of k within the class of non-labial fricatives (S, s, x). This behaviour can be explained by the centum-satem distinction within the Indo-European languages. The sounds were clustered into eleven classes to mirror the ten sound classes proposed by Dologopolsky [4] and an eleventh class for the vowels. The clustering is indicated by the boxes in Fig. 2. The similarity between the resulting clusters and the Dolgopolsky classes was evaluated using Pearson’s χ-square test [18]. The χ-square test shows that the clusters are highly similar (p-value : 1.824 × 10\(^{-13}\)).

The initial split of vowels and consonants in the hierarchical clustering hints at a further investigation of the substitution probabilities for vowels as well. Following the model of Wieling et. al. (2012) [24] and Jäger (2013) [10] the sub-

\(^4\) To perform the clustering, the substitution probabilities were transformed into distances by subtracting them from the maximal probability. For the hierarchical clustering, Wards method was used [21].
stitution probabilities of the vowels were plotted using multidimensional scaling. The resulting plot looks pretty similar to a rotated version of the articulatory vowel quadrangle (see Fig. 3). The schwa (ASJP symbol 3) is in the center of the quadrangle, i and u form the upper and the a the lower border. This visual inspection of the substitution probabilities shows that pHMMs are able to capture general tendencies of sound similarities.

5 Discussion

This study has shown that Pair Hidden Markov models are able to achieve very good results for the alignment of sequences. Although, this study shows no difference in performance between the introduction of informed initial weights and uniform initial weights, this result can be attributed to the properties of the training data. A further investigation of this aspect has shown a great similarity of proposed and experimentally obtained sound clusters. This suggests that the information of sound classes may be useful for applications with sparser training data. This issue has to be addressed in further research.

In comparison to other alignment algorithms pHMMs make use of a lot of parameters. Although this makes a sufficient training set necessary, it also enables pHMMs to give a more fine grained analysis. A gap in the alignment does not only depend on the probability to open a gap, but also to have a specific sound aligned with a gap. Furthermore, pHMMs do not exclude certain matches they just make them less probable. The SCA-method [11] does allow to match two sounds belonging to the different sound classes as well. However, the probabilities of these matches are estimated class-wise. The pHMM approach does not make such an assumption. All probabilities of matches are estimated sound-wise.
Fig. 3. Vowel quadrangle produced by multidimensional scaling of vowel substitution probabilities. The vowels are represented by their ASJP symbols.

Although the substitution probabilities strongly mirror the behaviour of sound classes, the pHMM approach is less restrictive.

In contrast to the classical Needleman-Wunsch algorithm, the probabilistic nature of pHMMs allows to analyze the reliability of the alignments and also consider suboptimal alignments. This feature makes pHMMs highly attractive for phylogenetic research. The performance of the pHMM in this study also encourages the usage of pHMMs in this line of research.

References

18. Pearson, K. FRS: X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5 50(302), 157–175 (1900)