Proceedings of the ESSLLI & WeSSLLI Student Session 2020

Web Summer School
in Logic, Language, and Information
July 11-17, Brandeis University

Alexandra Pavlova
(Editor)

Preface

These proceedings contain the papers presented at the Student Session of the Web Summer School in Logic, Language, and Information (WeSSLLI), taking place online and organized by the Brandeis University from July 11^{th} to 17^{th}, 2020. The Student Session is a part of the ESSLLI tradition. Due to the circumstances around the coronavirus (COVID-19) pandemic, the 32nd edition of the European Summer School in Logic, Language and Information, that should have taken place in Utrecht, has been postponed to 2-13 August 2021. However, since the ESSLLI Student Session Program Committee had already received several submissions when this decision was made, it was suggested to merge the ESSLLI Student Session with the WeSSLLI 2020 Virtual Student Session (except for the reviewing process). We would like to thank the ESSLLI Organizing Committee for helping us merge the events and especially the WeSSLLI Organizing Committee as well as the WeSSLLI Student Session organizers for this opportunity to run the joint Student Session, for organizing the entire summer school and supporting us in numerous ways. Furthermore, we would like to express our gratitude to the technical program chairs, Sophia Malamud and James Pustejovsky.

The ESSLLI & WeSSLLI Student Session is an excellent venue for students to present their work on a diverse range of topics at the interface of logic, language, and information, and to receive valuable feedback from renowned experts in their respective fields. The ESSLLI Student Session accepts submissions for three different tracks: Language and Computation (LaCo), Logic and Computation (LoCo), and Logic and Language (LoLa). Regarding the 2020 edition, these traditional tracks were laced with two new topics from the WeSSLLI Student Session, namely: second language acquisition (SLA) and phonology. The Student Session attracted submissions from 14 different countries this year from all over the world. As in previous years, the submissions were of high quality, and acceptance decisions were hard to make. However, the coronavirus (COVID-19) pandemic crisis had an impact on the number of submissions which decreased significantly compared to the previous years. Nevertheless, this experimental online format turned out a success regardless of some difficulties that we have encountered while preparing the Student Session. We received 39 submissions in total. At the combined Student Session, 5 of these submissions were presented as talks (30 minutes) and 18 submissions were presented in the form of a poster. As a longer version was not in the original requirements of the WeSSLLI Student Session, not all of its presenters decided to submit an extended version of their work, thus, those 6 papers were not included in the online proceedings.

We would like to thank each of the ESSLLI and WeSSLLI co-chairs for all their invaluable help in the reviewing process and organization of the combined Student Session. Without them, the combined Student Session would not have been able to take place. Additionally, we would like to thank the area experts for their help in the reviewing process and their support of the co-chairs. Thanks go to the chairs of the previous Student Sessions, in particular to Matteo Manighetti and Merijn Beeksma for providing us with the materials from the previous years and for their advice. As in previous years, Springer-Verlag has generously offered prizes for the *Best Paper* and *Best Poster Awards*, and for this we are very grateful. Most importantly, we would like to thank all those who submitted to the combined Student Session, for you are the ones that make it such an exciting event to organize and attend.

July 2020

Alexandra Pavlova
Chair of the ESSLLI 2020 Student Session

Organization

ESSLLI Student Session Program Committee

Chair

Alexandra Pavlova	TU Wien (Technische Universität Wien)

Language & Computation co-chairs

Alexandra Mayn	Saarland University
Eugénio Ribeiro	L^2F – Spoken Language Systems Laboratory, INESC-ID Lisboa Instituto Superior Técnico, Universidade de Lisboa

Logic & Computation co-chairs

Rafael Kiesel	TU Wien (Technische Universität Wien)
Mina Young Pedersen	University of Bergen

Language & Logic co-chairs

Dean McHugh	Institute of Logic, Language and Computation (ILLC), University of Amsterdam
Jonathan Pesetsky	University of Massachusetts, Amherst

WeSSLLI Student Session Organizing Committee

Fox Baudelaire	Brandeis University
Kenneth Lai	Brandeis University
Alex Lưu	Brandeis University
Dom O'Donnell	Brandeis University
Hayley Ross	Brandeis University

Area Experts

Language & Computation

Natasha Korotkova	University of Konstanz
Julian Schlöder	Institute for Logic, Language & Computation (ILLC), University of Amsterdam

Logic & Computation

Ronald de Haan	Institute for Logic, Language & Computation (ILLC), University of Amsterdam

Language & Logic

Alex Göbel	University of Massachusetts, Amherst
Sophia Malamud	Brandeis University
Deniz Ozyldiz	University of Massachusetts, Amherst

Second language Acquisition (SLA)

Tania Ionin	University of Illinois
Keith Plaster	Brandeis University

Phonology

Mary Zampini	Le Moyne College

Table of Contents

ESSLLI

Logic & Computation

Language & Computation

Logic & Language

WeSSLLI

Language & Computation

Logic & Language

Second language Acquisition (SLA)

Phonology

Formalizing Henkin-Style Completeness of an Axiomatic System for Propositional Logic

Asta Halkjær From

Technical University of Denmark

Abstract. We formalize a Henkin-style completeness proof for an axiomatic system for propositional logic in the proof assistant Isabelle/HOL. Our formalization precisely details the structure of this proof method.

Keywords: Propositional logic · Henkin-style completeness · Isabelle/HOL.

1 Introduction

Hilbert proved the completeness of an axiomatic system for propositional logic in 1917-18 [34], Gödel proved the completeness of first-order logic in 1929 [12] and Henkin simplified this proof in 1947 [13], devising what we now know as the Henkin-style method [14]. In this paper we study the structure of a Henkin-style completeness proof for an axiomatic Hilbert system for propositional logic by formalizing it in the proof assistant Isabelle/HOL [21].

Isabelle is a generic proof assistant and Isabelle/HOL is the instance based on higher-order logic. With it, we can state every definition, proposition and proof in the precise language of higher-order logic rather than in natural language. Our proof language is then completely formal which makes it possible for the machine to assist us in our endeavour. By writing our proofs in the Isar language, an acronym of *intelligible semi-automated reasoning* [32], we can have Isabelle check everything that we type. In particular, Isar contains commands such as **assume** to introduce assumptions, **have** to state a partial result and **moreover** to chain several of these together. After these commands, we typically write so-called ‹cartouches› delimited by angle brackets that contain our higher-order logic terms: definitions, statements and so on [33]. Our proofs are checked by the trusted Isabelle/HOL kernel but we do not typically write proofs directly using the kernel's axioms and inference rules. Instead we give the name of a prover that implements a proof search procedure like tableaux or resolution and Isabelle will run the prover to obtain the proof object. By formalizing our proofs like this we know that our conclusions always follow.

Of course, Isabelle cannot verify that our definitions match our intentions, that part is up to us, but formalization still reduces the possibility of mistakes. In particular, it reduces the surface area where mistakes can happen since the proofs themselves are checked by the machine. Not only does a formalization like the one we present increase the trust in our result, it can also serve as a reference to understand the proof since every detail is given: no case can be omitted as "trivial" or left as an "exercise for the reader." Our work can also act as starting point for formalizing other results based on the same techniques.

The full formalization, just below 400 lines, is available online:

https://github.com/logic-tools/axiom

We reproduce the essential pieces of it here and introduce parts of the syntax as we go along but forgo any thorough explanation.

1.1 Structure of the paper

After giving a brief history of formalized completeness proofs we start off by formalizing the syntax and semantics of our propositional logic (§ 2) and defining a sound proof system (§ 3). The idea of the completeness proof is as follows: given a formula ϕ valid under assumptions $\psi_1, \ldots, \psi_k$, assume for the sake of contradiction that there is no corresponding derivation:

$$\nvdash \psi_1 \longrightarrow \ldots \longrightarrow \psi_k \longrightarrow \phi$$

This means we cannot derive falsity, $\bot$, when also assuming $\neg\phi$, i.e.:

$$\nvdash \neg\phi \longrightarrow \psi_1 \longrightarrow \ldots \longrightarrow \psi_k \longrightarrow \bot$$

The set $\{\neg\phi, \psi_1, \ldots, \psi_k\}$ is therefore *consistent* and can be turned into a *maximal consistent set* (§ 4) through an *extension* (§ 5). Such sets are *Hintikka* sets (§ 6) and their elements have a model. This contradicts the validity assumption, proving that a derivation must exist. The proof system is therefore complete (§ 7) and we conclude with possible extensions (§ 8).

1.2 A history of formalized completeness proofs

Our formalization is only one in a long line of formalized completeness proofs.

Completeness proofs can generally be split into two categories based on their approach: semantic proofs in the style of Gödel [12] and Henkin [13] on the one hand and syntactic proofs in the style of Beth and Hintikka [17] and Gallier [11] on the other. Fitting and Mendelsohn call the semantic proofs "synthetic" because they start from a formula and *synthesize* new ones, building up larger and larger sets of formulas that are consistent with the starting point [9]. Formulas in such sets are then shown to have a model and this is the approach we take in this paper. Fitting and Mendelsohn contrast this with the syntactic proofs that they dub "analytic" because they work by *analyzing* the given formula, breaking it into smaller and smaller subformulas and reasoning from those. In these proofs we typically construct a counterexample from the open leaves or an infinite path of a failed derivation attempt. The synthetic approach is remarked to have a *mathematical*, abstract feeling whereas the analytic approach is more *computational* and often resembles an actual prover for the logic [5].

The Henkin-style completeness method has been applied to modal logic from the beginning, notably to system S5 as early as 1959 by Bayart (in French) [1]. Bentley recently formalized such a proof in the proof assistant Lean [2]. Jørgensen et al. adapted the synthetic approach to a tableau system for hybrid logic [16] with a formalization in Isabelle/HOL due to the present author [10]

In 1985, Shankar formalizes Shoenfield's first-order logic and axiomatic proof system in the Boyer-Moore theorem prover [28]. They show propositional completeness of the system analytically by defining a tautology checker for a fragment of the syntax based on negation and disjunction.

In 1996, Persson shows constructive completeness for intuitionistic first-order logic in Martin-Löf type theory using the proof assistant ALF [24]. Their proof has a synthetic flavor and their result is constructive: they obtain a program that transforms a proof of validity into a derivation in either natural deduction or sequent calculus. Persson also formalizes an axiomatic system but without proving its completeness.

By early 2000, Margetson formalizes the completeness of first-order logic and the cut elimination theorem for sequent calculus in Isabelle/HOL and Ridge later updates the formalization to the Isar language [18]. Their completeness proof is in the Beth-Hintikka style and based on analyzing failing branches in proof trees.

In 2005, Braselmann and Koepke follow in the Mizar system but using a Henkin-style argument for their sequent calculus [6].

In 2007, Berghofer formalizes Fitting's synthetic work on natural deduction [8] in Isabelle/HOL [3]. The formalized model existence theorem is based on Smullyan's abstract consistency properties [30] and Berghofer follows Fitting in reusing the result to show the Löwenheim-Skolem theorem. The present author has extended the completeness result in that formalization to also cover open formulas [3]. In 2016, Schlichtkrull extended Berghofer's work in another direction, namely to prove the completeness of first-order resolution [26,27].

In 2010, Ilik investigates Henkin-style arguments for both classical and intuitionistic first-order logic in the proof assistant Coq [15].

In 2017, Michaelis and Nipkow formalize a number of proof systems for propositional logic in Isabelle/HOL: natural deduction, sequent calculus, an axiomatic Hilbert system similar to ours and resolution [19,20]. They give a syntactic completeness proof for the sequent calculus and show that sequent calculus derivations can be translated into natural deduction and further into their Hilbert system, obtaining completeness for the three proof systems. Independently of this approach, they formalize the propositional model existence theorem by Fitting [8] and use this result to reprove completeness of the sequent calculus and Hilbert system, respectively. Their formalization is more ambitious than ours and therefore more involved. We start from a smaller syntax and focus on only one proof system and one approach. This leads to a simpler formalization and helps us understand the essential pieces of the approach.

Blanchette, Popescu and Traytel have recently advanced the state of completeness proofs for sequent calculus and tableau systems in Isabelle/HOL [5]. They explicitly shy away from Henkin in favor of the Beth-Hintikka style and use codatatypes to model possibly infinite derivation trees. Their result can be instantiated for different variations of sequent calculus or tableau and various flavors of first-order logic.

Blanchette gives an overview of the formalized metatheory of various other logical calculi and automatic provers in Isabelle [4].

If we move to Gödel's incompleteness theorems, the first one has been formalized in the Boyer-Moore theorem prover by Shankar in 1986 [29] and in Coq by O'Connor in 2003 [22]. Both incompleteness theorems have been formalized in Isabelle/HOL by Paulson in 2013 [23] and by Popescu and Traytel in 2019 [25].

In summation, the Henkin style is ubiquitous and we have seen it applied to examples such as sequent calculus and natural deduction for first-order logic, system S5 for modal logic and a tableau system for hybrid logic. Most work either extends the technique to cover more advanced logics or abstracts it so that it applies to several at once. Our contribution is to boil this proof style down to its essence, motivating each step as we present it and using a proof assistant to ensure precision, correctness and comprehensiveness. Our work may also serve as a fast-paced introduction to proof assistants.

2 Syntax and Semantics

We pick a minimal syntax consisting of a logical constant representing falsity, natural numbers as propositional symbols, and implication. We model the syntax as a datatype, *form*, with a constructor for each case separated by "|":

```
datatype form = Falsity (‹⊥›) | Pro nat | Imp form form (infixr ‹⟶› 25)
```

The annotations in parentheses allow us to construct formulas using standard notation in bold. Our definition of negation as an abbreviation makes use of this:

```
abbreviation Neg (‹¬ -› [40] 40) where ‹¬ p ≡ p ⟶ ⊥›
```

We define the semantics as a primitive recursive predicate on formulas given an interpretation of propositional symbols:

```
primrec semantics :: ‹(nat ⇒ bool) ⇒ form ⇒ bool› (‹- ⊨ -› [50, 50] 50) where
  ‹(I ⊨ ⊥) = False›
| ‹(I ⊨ Pro n) = I n›
| ‹(I ⊨ (p ⟶ q)) = ((I ⊨ p) ⟶ (I ⊨ q))›
```

The first line gives the type and infix notation while the remaining lines define the predicate by each case of the syntax. The first case states that no interpretation models ⊥, the second case that the semantics of a propositional symbol is given by the interpretation and finally we delegate to the meta-logic implication, ⟶, to interpret the object logic implication **⟶**.

3 Proof System

We pick a simple axiomatic proof system for our purposes, consisting of modus ponens and three axiom schemas. This is Church's axiom system P_1 [7]:

```
inductive Axiomatics :: ‹form ⇒ bool› (‹⊢ -› [50] 50) where
  MP: ‹⊢ p ⟹ ⊢ (p ⟶ q) ⟹ ⊢ q›
| Imp1: ‹⊢ (p ⟶ q ⟶ p)›
| Imp2: ‹⊢ ((p ⟶ q ⟶ r) ⟶ (p ⟶ q) ⟶ p ⟶ r)›
| Neg: ‹⊢ (((p ⟶ ⊥) ⟶ ⊥) ⟶ p)›
```

The proof system is sound with respect to the semantics, which means that every derivable formula is true under any interpretation:

```
theorem soundness: ‹⊢ p ⟹ I ⊨ p›
  by (induct rule: Axiomatics.induct) simp-all
```

The second line shows that the simplifier can easily verify the theorem once we state that the proof should be performed by induction over the rules.

4 Consistency and Maximality

We want to work with sets of formulas where no finite subset S' syntactically entails falsity, i.e. we cannot derive $\bot$ given S'. Our provability predicate, $\vdash$, has no notion of entailment but we can use implication, $\longrightarrow$, to serve the same purpose. As such, we use the following function, *imply*, to build a chain of implications from a given list of assumptions to a conclusion. A list is a finite sequence and is either empty, [], or built from an element, the separator #, and a smaller list. We say that q can be *derived from ps* when we can derive ⊢ *imply ps q*.

```
primrec imply :: ‹form list ⇒ form ⇒ form› where
  ‹imply [] q = q›
| ‹imply (p # ps) q = (p ⟶ imply ps q)›
```

The set S is consistent exactly when there is no list S' that, when treated as a *set*, is a subset of S and that entails $\bot$ in the sense of *imply*:

```
definition consistent :: ‹form set ⇒ bool› where
  ‹consistent S ≡ ∄S'. set S' ⊆ S ∧ ⊢ imply S' ⊥›
```

A set is maximal when any proper extension makes it inconsistent:

```
definition maximal :: ‹form set ⇒ bool› where
  ‹maximal S ≡ ∀p. p ∉ S ⟶ ¬ consistent ({p} ∪ S)›
```

Note that we allow for inconsistent maximal sets to separate concerns.

5 Extension

We need to grow a consistent set into a *maximal* one while preserving consistency. According to Lindenbaum's lemma, attributed to him by Tarski [31], we can always do this. Given an enumeration of formulas, (ϕ_n), we construct a corresponding sequence of consistent sets (S_n).

Assuming S_n has been constructed, its immediate extension is given by:

$$S_{n+1} = \begin{cases} \{\phi_n\} \cup S_n & \text{if } \{\phi_n\} \cup S_n \text{ is consistent,} \\ S_n & \text{otherwise.} \end{cases}$$

That is, we only add the corresponding formula to the previous set if consistency is preserved. In the Isabelle code, we use the function *extend S f n* to construct S_n from $S = S_0$ given an enumeration of formulas represented by f:

```
primrec extend :: ⟨form set ⇒ (nat ⇒ form) ⇒ nat ⇒ form set⟩ where
  ⟨extend S f 0 = S⟩
| ⟨extend S f (Suc n) =
    (if consistent ({f n} ∪ extend S f n)
     then {f n} ∪ extend S f n
     else extend S f n)⟩
```

To construct our *maximal consistent set* we take the infinite union $\bigcup S_n$:

```
definition Extend :: ⟨form set ⇒ (nat ⇒ form) ⇒ form set⟩ where
  ⟨Extend S f ≡ ⋃n. extend S f n⟩
```

It is easy to see that the starting set is a subset of the union:

```
lemma Extend-subset: ⟨S ⊆ Extend S f⟩
  unfolding Extend-def by (metis Union-upper extend.simps(1) range-eqI)
```

And that any element S_m is a superset of previous elements:

```
lemma extend-bound: ⟨(⋃n ≤ m. extend S f n) = extend S f m⟩
  by (induct m) (simp-all add: atMost-Suc)
```

5.1 Consistency

When the initial S is consistent, so is any S_n by construction:

```
lemma consistent-extend: ⟨consistent S ⟹ consistent (extend S f n)⟩
  by (induct n) simp-all
```

Finally, we show that the limit, $\bigcup S_n$, is also consistent:

```
lemma consistent-Extend:
  assumes ⟨consistent S⟩
  shows ⟨consistent (Extend S f)⟩
```

We prove this by classical contradiction using the *ccontr* rule:

```
  unfolding Extend-def
proof (rule ccontr)
```

Assuming the union is inconsistent, we can derive $\bot$ from some subset S':

```
  assume ⟨¬ consistent (⋃n. extend S f n)⟩
  then obtain S' where ⟨⊢ imply S' ⊥⟩ ⟨set S' ⊆ (⋃n. extend S f n)⟩
    unfolding consistent-def by blast
```

This subset is finite so it must be a subset of a finite segment of the union, say $S_0 \cup \ldots \cup S_m$ for some m:

```
  then obtain m where ⟨set S' ⊆ (⋃n ≤ m. extend S f n)⟩
    using UN-finite-bound by (metis List.finite-set)
```

But every element in (S_n) is a subset of the next, so S' is a subset of S_m:

```
  then have ⟨set S' ⊆ extend S f m⟩
    using extend-bound by blast
```

And we already established that any such element is consistent:

```
  moreover have ⟨consistent (extend S f m)⟩
    using assms consistent-extend by blast
```

So there cannot be an inconsistent subset S' and we have our contradiction:

```
  ultimately show False
    unfolding consistent-def using ⟨⊢ imply S' ⊥⟩ by blast
qed
```

In conclusion, $\bigcup S_n$ is consistent when S_0 is.

5.2 Maximality

Importantly, the union $\bigcup S_n$ is also maximal (regardless of the choice of S_0):

```
lemma maximal-Extend:
  assumes ⟨surj f⟩
  shows ⟨maximal (Extend S f)⟩
  (proof omitted)
```

The proof is similar to the one for consistency. If the union is not maximal then there is some $\phi_k \notin \bigcup S_n$ such that $\{\phi_k\} \cup \bigcup S_n$ is consistent. Since $\phi_k \notin \bigcup S_n$, it was not added to the sequence, i.e $\phi_k \notin S_{k+1}$, and by construction this must be because $\{\phi_k\} \cup S_k$ is inconsistent. But $\{\phi_k\} \cup \bigcup S_n$ is a superset of $\{\phi_k\} \cup S_k$, so $\{\phi_k\} \cup \bigcup S_n$ must be inconsistent too, contradicting our assumption.

6 Hintikka Sets

The completeness proof works by showing that every maximal consistent set is a Hintikka set, where Hintikka sets are defined as follows:

```
locale Hintikka =
  fixes H :: ‹form set›
  assumes
    NoFalsity: ‹⊥ ∉ H› and
    Pro: ‹Pro n ∈ H ⟹ (¬ Pro n) ∉ H› and
    ImpP: ‹(p ⟶ q) ∈ H ⟹ (¬ p) ∈ H ∨ q ∈ H› and
    ImpN: ‹(¬ (p ⟶ q)) ∈ H ⟹ p ∈ H ∧ (¬ q) ∈ H›
```

The idea is to ensure that every formula in a set is satisfiable by ensuring through syntactic criteria that the set is *downwards saturated* [30], i.e. that the satisfiability of any complex formula is guaranteed by conditions on its subformulas. Since ⊥ is unsatisfiable it should never occur (*NoFalsity*), and if a propositional symbol occurs then its negation should not (*Pro*). An implication is satisfied if either the antecedent is false or the consequent is true, so if an implication occurs in a Hintikka set, then either the negated antecedent or the consequent should too (*ImpP*). If a negated implication occurs in a Hintikka set then so should both the antecedent and negated consequent (*ImpN*).

6.1 Model existence

The downwards saturation ensures that if we interpret every proposition in a Hintikka set as true then every larger formula in the set will be modelled by this interpretation. We therefore base the interpretation on set membership:

```
abbreviation (input) ‹model H n ≡ Pro n ∈ H›
```

This models any formula in a Hintikka set:

```
lemma Hintikka-model:
  ‹Hintikka H ⟹ (p ∈ H ⟶ model H ⊨ p) ∧ ((¬ p) ∈ H ⟶ ¬ model H ⊨ p)›
  by (induct p) (simp; unfold Hintikka-def, blast)+
```

6.2 Maximal consistency

Our final task is to show that a maximal consistent set is a Hintikka set:

```
lemma Hintikka-Extend:
  assumes ‹maximal S› ‹consistent S›
  shows ‹Hintikka S›
```

The proof has four cases based on the cases of the Hintikka definition and we show two of them here. Consider first propositional symbols:

```
fix n
assume ‹Pro n ∈ S›
moreover have ‹⊢ imply [Pro n, ¬ Pro n] ⊥›
  by (simp add: FalsityE)
ultimately show ‹(¬ Pro n) ∉ S›
  using assms(2) unfolding consistent-def
  by (metis bot.extremum empty-set insert-subset list.set(2))
```

We have assumed a fixed but arbitrary propositional symbol n that occurs positively in S. We can derive $\bot$ from this in combination with a negative occurrence. Thus, both cannot appear in the consistent S and this case of the Hintikka definition is satisfied.

Next, assume that a negated implication occurs in S. We show half of the Hintikka condition by contradiction, namely that so does the antecedent:

```
assume *: ‹(¬ (p ⟶ q)) ∈ S›
show ‹p ∈ S ∧ (¬ q) ∈ S›
proof (rule conjI; rule ccontr)
```

The set S is maximal, so if it does not contain p there must be some finite subset S' of S that we can derive falsity from when adding p:

```
  assume ‹p ∉ S›
  then obtain S' where S': ‹⊢ imply (p # S') ⊥› ‹set S' ⊆ S›
    using assms inconsistent-head by blast
```

We can *cut* out p and derive $\bot$ directly from the negated implication:

```
  moreover have ‹⊢ imply ((¬ (p ⟶ q)) # S') p›
    using add-imply ImpE1 deduct by blast
  ultimately have ‹⊢ imply ((¬ (p ⟶ q)) # S') ⊥›
    using cut' by blast
```

These assumptions, however, are a subset of S, contradicting its consistency:

```
  moreover have ‹set ((¬ (p ⟶ q)) # S') ⊆ S›
    using *(1) S'(2) by fastforce
  ultimately show False
    using assms unfolding consistent-def by blast
```

7 Completeness

Isabelle can automatically prove the countability of formulas, providing a surjective function *from-nat* for obtaining specific elements of the enumeration (ϕ_n):

```
instance form :: countable by countable-datatype
```

Finally we reach the completeness lemma itself. We assume that p is valid under the assumptions *ps* and show that we can derive p from *ps*:

```
lemma imply-completeness:
  assumes valid: ‹∀ I s. list-all (λq. I ⊨ q) ps ⟶ I ⊨ p›
  shows ‹⊢ imply ps p›
```

We proceed by contradiction and the application of a similar derivation rule:

```
proof (rule ccontr)
  assume ‹¬ ⊢ imply ps p›
  then have *: ‹¬ ⊢ imply ((¬ p) # ps) ⊥›
    using Boole by blast
```

We abbreviate the starting consistent set *?S* and its maximal extension *?H*:

```
  let ?S = ‹set ((¬ p) # ps)›
  let ?H = ‹Extend ?S from-nat›
```

And use the previous results to show that *?H* is a Hintikka set:

```
  have ‹consistent ?S›
    unfolding consistent-def using * imply-weaken by blast
  then have ‹consistent ?H› ‹maximal ?H›
    using consistent-Extend maximal-Extend surj-from-nat by blast+
  then have ‹Hintikka ?H›
    using Hintikka-Extend by blast
```

We have seen that we have a model for any formula in such an *?H*:

```
  have ‹model ?H ⊨ p› if ‹p ∈ ?S› for p
    using that Extend-subset Hintikka-model ‹Hintikka ?H› by blast
```

So in particular for $\neg p$ and all of *ps*:

```
  then have ‹model ?H ⊨ (¬ p)› ‹list-all (λp. model ?H ⊨ p) ps›
    unfolding list-all-def by fastforce+
```

Our validity assumption then gives us that *model ?H* also models p:

```
  then have ‹model ?H ⊨ p›
    using valid by blast
```

But this is a contradiction:

```
  then show False
    using ‹model ?H ⊨ (¬ p)› by simp
qed
```

As such, we must be able to derive any valid formula:

```
theorem completeness: ‹∀ I. I ⊨ p ⟹ ⊢ p›
  using imply-completeness[where ps=‹[]›] by simp
```

8 Conclusion

We have shown how to formalize the soundness and completeness of a simple axiomatic proof system for propositional logic in Isabelle/HOL. The proof assistant is sophisticated enough that we can do the soundness proof almost automatically and use constructions like infinite sets in the proof of completeness.

Our choice of propositional logic means that we miss out on an aspect of Henkin's original proof: the use of special constants to witness existential statements. In return, our formalization is more manageable.

The formalization is simple to extend. The supplementary material contains a file where we have added binary disjunction and conjunction operators to the syntax and updated the proof system and so on accordingly. The result is only around 130 lines longer and we did not have to modify any existing line, only to add new ones. The biggest changes are in the Hintikka definition and maximal consistency lemma while model existence is still completely automatic.

Acknowledgements We thank Jørgen Villadsen, Alexander Birch Jensen, Frederik Jacobsen and the anonymous reviewers for valuable comments.

References

1. Bayart, A.: Quasi-adéquation de la logique modale du second ordre S5 et adéquation de la logique modale du premier ordre S5. Logique et Analyse **2**(6/7), 99–121 (1959)
2. Bentzen, B.: A Henkin-style completeness proof for the modal logic S5 (2019), http://arxiv.org/abs/1910.01697, CoRR,
3. Berghofer, S.: First-Order Logic According to Fitting. Archive of Formal Proofs (Aug 2007), http://isa-afp.org/entries/FOL-Fitting.html
4. Blanchette, J.C.: Formalizing the metatheory of logical calculi and automatic provers in Isabelle/HOL (invited talk). In: Mahboubi, A., Myreen, M.O. (eds.) Proceedings of the 8th ACM SIGPLAN International Conference on Certified Programs and Proofs, CPP 2019. pp. 1–13. ACM (2019)
5. Blanchette, J.C., Popescu, A., Traytel, D.: Soundness and completeness proofs by coinductive methods. Journal of Automated Reasoning **58**(1), 149–179 (2017)
6. Braselmann, P., Koepke, P.: Gödel's completeness theorem. Formalized Mathematics **13**(1), 49–53 (2005)
7. Church, A.: Introduction to Mathematical Logic. Princeton Mathematical Series, Princeton University Press (1956)
8. Fitting, M.: First-Order Logic and Automated Theorem Proving, Second Edition. Graduate Texts in Computer Science, Springer (1996)
9. Fitting, M., Mendelsohn, R.L.: First-Order Modal Logic. Springer (2012)
10. From, A.H.: Formalizing a Seligman-style tableau system for hybrid logic. Archive of Formal Proofs (Dec 2019), http://isa-afp.org/entries/Hybrid_Logic.html, Formal proof development
11. Gallier, J.H.: Logic for computer science: foundations of automatic theorem proving. Courier Dover Publications (2015)
12. Gödel, K.: Über die Vollständigkeit des Logikkalküls. Ph.D. thesis, University of Vienna (1929)

13. Henkin, L.: The Completeness of Formal Systems. Ph.D. thesis, Princeton University (1947)
14. Henkin, L.: The Discovery of My Completeness Proofs. Bulletin of Symbolic Logic **2**(2), 127–158 (1996)
15. Ilik, D.: Constructive completeness proofs and delimited control. Ph.D. thesis, École polytechnique (2010)
16. Jørgensen, K.F., Blackburn, P., Bolander, T., Braüner, T.: Synthetic completeness proofs for Seligman-style tableau systems. In: Proceedings of the 11th conference on Advances in Modal Logic. pp. 302–321 (2016)
17. Kleene, S.C.: Mathematical Logic. Wiley, London (1967)
18. Margetson, J., Ridge, T.: Completeness theorem. Archive of Formal Proofs (Sep 2004), http://isa-afp.org/entries/Completeness.html, Formal proof development
19. Michaelis, J., Nipkow, T.: Propositional proof systems. Archive of Formal Proofs (Jun 2017), http://isa-afp.org/entries/Propositional_Proof_Systems.html, Formal proof development
20. Michaelis, J., Nipkow, T.: Formalized proof systems for propositional logic. In: Abel, A., Forsberg, F.N., Kaposi, A. (eds.) 23rd Int. Conf. Types for Proofs and Programs (TYPES 2017). LIPIcs, vol. 104, pp. 6:1–6:16. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2018)
21. Nipkow, T., Paulson, L.C., Wenzel, M.: Isabelle/HOL - A Proof Assistant for Higher-Order Logic, Lecture Notes in Computer Science, vol. 2283. Springer (2002)
22. O'Connor, R.: Essential incompleteness of arithmetic verified by Coq. In: International Conference on Theorem Proving in Higher Order Logics. pp. 245–260. Springer (2005)
23. Paulson, L.C.: A machine-assisted proof of Gödel's incompleteness theorems for the theory of hereditarily finite sets. The Review of Symbolic Logic **7**(3), 484–498 (2014)
24. Persson, H.: Constructive completeness of intuitionistic predicate logic. Licenciate thesis, Chalmers University of Technology (1996)
25. Popescu, A., Traytel, D.: A formally verified abstract account of Gödel's incompleteness theorems. In: Fontaine, P. (ed.) Automated Deduction – CADE 27. pp. 442–461. Springer International Publishing, Cham (2019)
26. Schlichtkrull, A.: The resolution calculus for first-order logic. Archive of Formal Proofs (Jun 2016), http://isa-afp.org/entries/Resolution_FOL.html, Formal proof development
27. Schlichtkrull, A.: Formalization of the resolution calculus for first-order logic. Journal of Automated Reasoning **61**(1-4), 455–484 (2018)
28. Shankar, N.: Towards mechanical metamathematics. Journal of Automated Reasoning **1**(4), 407–434 (1985)
29. Shankar, N.: Metamathematics, machines and Gödel's proof. No. 38, Cambridge University Press (1997)
30. Smullyan, R.M.: First-Order Logic. Springer-Verlag (1968)
31. Tarski, A.: Logic, Semantics, Metamathematics: Papers from 1923 to 1938. Hackett Publishing (1983)
32. Wenzel, M.: Isabelle/Isar—a generic framework for human-readable proof documents. From Insight to Proof—Festschrift in Honour of Andrzej Trybulec **10**(23), 277–298 (2007)
33. Wenzel, M.: The Isabelle/Isar Reference Manual (2020), part of the Isabelle distribution.
34. Zach, R.: Completeness before Post: Bernays, Hilbert, and the development of propositional logic. Bulletin of Symbolic Logic **5**(3), 331–366 (1999)

A Logical Framework for Understanding Why[⋆]

Yu Wei

Department of Philosophy, Peking University, China
wei_yu@pku.edu.cn

Abstract. Epistemic logic pays barely any attention to the notion of understanding, and stands in total contrast to the current situation in epistemology and in philosophy of science. This paper studies understanding why in an epistemic-logic-style. It is generally acknowledged that understanding why moves beyond knowing why. Inspired by philosophical ideas, we consider whereas knowing why requires knowing horizontal explanations, understanding why additionally requires vertical explanations. Based on justification logic and existing logical work for knowing why, we build up a framework by introducing vertical explanations, and show it could accommodate different philosophical viewpoints via adding conditions in the models. A sound and complete axiomatization for the most general case is given.

Keywords: Understanding why · Knowing why · Justification logic.

1 Introduction

There has been a resurgence of interest among epistemologists and philosophers of science in the nature of understanding recently. Different uses of 'understanding' seem to mean so many different things. Literature tends to suppose three main types of understanding (cf. [5]):

- Propositional understanding or understanding-that: "I understand that X."
- Atomistic understanding or understanding-wh: "I understand why/how X."
- Objectual understanding or holistic understanding: "I understand X."

Among all the types above, plenty of recent work focus on understanding-why, which is also called a narrow conception of understanding.

While a lot of discussions have been taking place among philosophers, there is barely any attention to characterizing understanding in literatures on epistemic logic (an exception being [4], which is about "understanding a proposition"). As we know, apart from "knowing that", there has been a growing interest in epistemic logic in various knowledge expressions in terms of "knowing what", "knowing how", "knowing why" and so on (see the survey in [22]). It would become interesting to introduce the notion of understanding into current framework of epistemic logic, to see what would happen between understanding and

[⋆] The author wishes to thank Fernando Velazquez Quesada, Sonja Smets and Yanjing Wang for their patient guidance and helpful suggestions.

knowing and what the distinct logical principles for understanding are. This paper focuses on understanding why, and the main motivation is to contribute to the explication of "understanding why" from the perspective of epistemic logic.

As noted in [18], the relation between understanding and knowing has been a prominent theme in the search for a satisfactory account of understanding. Hence we start by considering the logic of knowing why. Xu, Wang and Studer recently take the ideas similar to justification logic together with the standard notions of epistemic logic to capture knowing why in [24]. There is a very general connection between knowledge and wh-questions discovered by Hinttika in the framework of quantified epistemic logic (cf. [10]). The authors thus view knowing why φ as knowing an answer to the question "Why φ?", which intuitively amounts to knowing an explanation of φ. The explanatory relation between explanations and propositions is characterized by the format $t : \varphi$, which is a formula from justification logic originally stating that "t is a justification of φ". The analysis of "knowing why φ" is $\exists t \mathsf{K}_i(t : \varphi)$.

According to the philosophical ideas, it is widely assumed that understanding why φ moves beyond knowing why φ, in which knowing why φ is commonly analyzed as identifying the dependencies, say knowing that "φ because ψ". This view, called non-reductionist, is understanding why cannot be reduced to knowing why. Pritchard [16] introduces a scenario where a child knows via testimony that a house burned down because of faulty wiring. The child then could answer a corresponding why-question since she accepts the information, and say, ready to repeat it to her friends. However, while the parent understands why the house burned down because the parent also knows how the faulty wiring caused the fire, the child has no conception of that and thus has no understanding why.

It seems plausible that understanding why requires more than merely knowing an answer to the why-question. This "more" is usually illustrated by more questions in literatures like [17]: if the child were asked the question of why the introduction of faulty wiring caused the fire, she would be unable to respond. Non-reductionists argue that one having understanding-why could in addition answer a kind of "vertical" follow-up why question (see [13]) or a "what-if-things-had-been-different" question (see [7]). We will present both notions in Sect.2.1. Since providing an explanation amounts to answering a why-question, these philosophical insights inspire us to introduce more (sorts of) explanations into the notion of understanding why, so as to respond to asking for further information. We will combine the idea that understanding why φ requires answers to more questions with the apparatus in [24], and analyze understanding why φ as $\exists t_1 \exists t_2(\mathsf{K}(t_2 : (t_1 : \varphi))$, where t_1 is an answer to "Why φ?" and t_2 is an answer to the vertical follow-up question "Why t_1 is the answer to 'Why φ?'?", or to the question "What if things in t_1 had been different?".

The paper is organized as follows. Sect. 2 looks at the philosophical discussions and logical work relevant to our topic in a little more details. Sect. 3 provides a logical framework for making such analysis of understanding why precise. Sect. 4 gives an axiomatization of the general version of understanding why. We conclude in Sect. 5 with discussions on the potentially future work.

2 Preliminary

2.1 Philosophical views

According to the authors of [12,15,20] and [21] etc, views on the nature of understanding why fall into two broad camps: reductionists and non-reductionists, in which the former hold that: one understands why φ iff she knows why φ. Knowing why φ is analyzed as knowledge of causes of φ, or more generally, knowledge of dependencies (cf. [6], [8]). By contrast, non-reductionists mainly argue that knowing why is not sufficient for understanding why, and their view can be illustrated with Pritchard's case of the house fire above.

In response to the counterexamples, on one side, Grimm [8] holds that the counterexamples contain an inadequate idea of what it means to have knowledge of causes. Knowing why amounts to having a sufficient conception of how cause and effect might be related, which is called "modal relationship" in [8], rather than just assenting to the proposition that describe this relationship. The notion of knowing why such understood is a kind of (limited) understanding why.

On the other side, Pritchard [17] famously proposes that while knowing-why requires identifying the cause, understanding-why requires having a sound explanatory story regarding how cause and effect are related, which is a kind of cognitive achievement.When trying to clarify the "sound explanatory story" mentioned by Prichard, Lawer [13] borrows the idea in [19]: whenever you answer a why-question, you create an opportunity for your questioner to immediately ask "why?" about your answer. Recall the experiences with children. There are two importantly different ways to ask "why?" about the answer to a why-question:

1. "horizontal" follow-up why-question: someone says "φ because r" and you ask "why is it the case that r?" Traces this chain of reasons "backward".
2. "vertical" follow-up why-question: we step outside the chain of reasons, and ask what the facts in the chain have done to belong in the chain.

Generally speaking, while the "horizontal" follow-up why-questions seek for lower-level explanations, the "vertical" follow-up why-questions seek higher-order explanations, that is, the explanations why those explanations are explanations. As an example of the different levels of reasons why from [19], Suzy throws a rock at a window but Billy sticks his mitt out, thereby catching the rock before it hits. The fact that Billy stuck his mitt out is a reason why the window didn't break. And the fact that Suzy threw the rock is a reason why "the act that Billy stuck his mitt out is a reason why the window didn't break". Lawler [13] suggests that the essence of a sound explanatory story regarding how cause and effect are related is an answer to the "vertical" follow-up why-question.

Besides, Hills [9] suggests that the distinction between knowing why and understanding why lies in "grasping" an explanation, which means answering questions of "What if . . . ?" sort, like "what-if-things-had-been-different" proposed by Woodward [23].

Although these philosophical views are varied, we can find a common thread: understanding why requires at least two explanations of different levels. Bermúdez

[1] acknowledges a distinction between *horizontal explanation* and *vertical explanation*, which could be made use of to refer to these different levels. Think of horizontal explanation as the explanation required in ordinary knowing-why, vertical explanations can broadly be characterized as explaining the grounds of horizontal explanations, which is able to accommodate "modal relationship" in the knowledge of causes and answers to vertical follow-up why questions and "what if" questions.

2.2 Logic of Knowing Why and Fitting Model

For lack of space, we only look at the logic introduced in [24], which inspires our techniques. Initially, the analysis of "knowing why φ" is $\exists t \mathsf{K}_i(t : \varphi)$. Xu et al. pack the quantifier and modality together, and introduce a new operator Ky_i to denote $\exists t \mathsf{K}_i(t : \varphi)$ into the language of standard multi-agent epistemic logic.

The semantics is defined in a classical epistemic model with some apparatus similar to Fitting model of justification logic. A knowing why model $\mathfrak{M}$ is defined as a tuple $(W, E, \{R_i \mid i \in I\}, \mathcal{E}, V)$ where $(W, \{R_i \mid i \in I\}, V)$ is an epistemic model, E is a non-empty set of explanations, and $\mathcal{E}$ is an admissible explanation function specifying the set of worlds where $t \in E$ is an explanation of φ.

The truth conditions for the classical operators from epistemic logic are routine, and with: $\mathsf{Ky}_i\varphi$ holds at $\mathfrak{M}, w$ iff (1) there exists $t \in E$ such that for all v with wR_iv, $v \in \mathcal{E}(t, \varphi)$; and (2) for all v with wR_iv, φ holds at v.

A Fitting model $\mathfrak{M}^J$ for justification logic is a tuple $(W^J, R^J, \mathcal{E}^J, V^J)$ based on a single-agent Kripke model (W^J, R^J, V^J), in which $\mathcal{E}^J$ is an evidence function assigning justification terms to formulas on each world. The evaluation of the format $t : \varphi$ follows that: $t : \varphi$ holds at a pointed model $\mathfrak{M}^J, w$ iff (1) $w \in \mathcal{E}^J(t, \varphi)$; and (2) for all v with wR^Jv, φ holds at v.

As noted in [24], Fitting models typically have monotonicity condition, i.e. $w \in \mathcal{E}^J(t, \varphi)$ and wR^Jv imply $v \in \mathcal{E}^J(t, \varphi)$. When R^J is an equivalence relation, it follows that all indistinguishable worlds have the same justification for the same formula, that is, $w \in \mathcal{E}^J(t, \varphi)$ iff $v \in \mathcal{E}^J(t, \varphi)$ whenever wR^Jv. Compared with knowing why models, Fitting models only store known explanations (justifications) but all other possible explanations (justifications) are dropped. Therefore Fitting models cannot tell the difference between $\exists t\mathsf{K}(t : \varphi)$ and $\mathsf{K}\exists t(t : \varphi)$, which is thought of by the authors as essential for the analysis of knowing why. Justification formula $t : \varphi$ accommodates a strict 'justificationist' reading in which it means t is accepted by the agent as a justification of φ. However, in [24] the format $t : \varphi$ actually is assigned an externalist and nonjustificationist reading, which could be used to formalize understanding.

Given the vertical explanation idea, one may be tempted to consider that the semantics for understanding why φ as $\mathsf{KyKy}\varphi$. Unfortunately it is infelicitous. $\mathsf{KyKy}\varphi$ states some explanation (say t_2) is known as an explanation of that some explanation (say t_1) is known as an explanation of φ, which is indeed a matter of introspection of one's knowing why. As a simple example from [24], the window is broken since someone threw a rock at it, and an agent knows that because she saw it, or someone told her about it. This kind of explanations is certainly not what we have in mind of the vertical explanations, and cannot be an answer to what-if questions as well. Thus we need a new logical framework.

3 A Framework for Understanding Why

In this section we introduce formally the language and the semantics. We will be interested in the issue of what it means to ascribe understanding to individual agents, so for the time being we set multi-agent aside for simplicity.

Definition 1 (Epistemic language of understanding-why). *Fix nonempty set P of propositional letters, the language* **ELUy** *is defined as (where $p \in P$):*

$$\varphi ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathsf{K}\varphi \mid \mathsf{Ky}\varphi \mid \mathsf{Uy}\varphi$$

The explication of understanding why is intended to be done by studying its relations with knowing why, so a new "packed" modality Uy for understanding why is introduced into the language in [24]. Besides, K is included because we intend to connect the notion of explanation with justification in our logic. That is to say, the explanation packed in $\mathsf{KyK}\varphi$ is considered as a justification for $\mathsf{K}\varphi$.

We accept the view in [24] that although something is a tautology, you may not know why it is a tautology. A special set of "self-evident" tautologies Λ is introduced, which the agent is assumed to know why. For example, we can let all the instances of $\varphi \wedge \psi \to \varphi$ and $\varphi \wedge \psi \to \psi$ be Λ. At present, we do not suppose any necessitation rule for Uy. L^P is used to denotes all formulas w.r.t. P below.

Definition 2. *An* **ELUy** *model $\mathcal{M}$ is a tuple $(W, E, R, \mathcal{E}, V)$ where:*

- *W is a non-empty set of possible worlds.*
- *E is a non-empty set of explanations satisfying:*
  1. *If $t, s \in E$, then $t \cdot s \in E$,*
  2. *If $t \in E$, then $!t \in E$,*
  3. *A special symbol c is in E.*
- *$R \subseteq W \times W$ is an equivalence relation over W.*
- *$\mathcal{E} : E \times (L^P \cup \langle E \times L^P \rangle) \to 2^W$ is an explanation function satisfying:*
  1. *Horizontal Application: $\mathcal{E}(t, \varphi \to \psi) \cap \mathcal{E}(s, \varphi) \subseteq \mathcal{E}(t \cdot s, \psi)$,*
  2. *Constant Specification: If $\varphi \in \Lambda$, then $\mathcal{E}(c, \varphi) = W$,*
  3. *Vertical Application I: $\mathcal{E}(t_2, \langle t_1, \varphi \to \psi \rangle) \cap \mathcal{E}(s, \varphi) \subseteq \mathcal{E}(t_2, \langle t_1 \cdot s, \psi \rangle)$,*
  4. *Vertical Application II: $\mathcal{E}(\langle t, \varphi \to \psi \rangle) \cap \mathcal{E}(s_2, \langle s_1, \varphi \rangle) \subseteq \mathcal{E}(s_2, \langle t \cdot s_1, \psi \rangle)$,*
  5. *Vertical Explanation Factivity: $\mathcal{E}(t_2, \langle t_1, \varphi \rangle) \subseteq \mathcal{E}(t_1, \varphi)$.*
  6. *Epistemic Introspection: $\mathcal{E}(t, \bigcirc\varphi) \subseteq \mathcal{E}(!t, \langle t, \bigcirc\varphi \rangle)$ for $\bigcirc = \mathsf{K}, \mathsf{Ky}, \mathsf{Uy}$.*
- *$V : P \to 2^W$ is a valuation function.*

The set E is closed under the application operator $\cdot$, which combines two explanations into one, and the (positive) introspection operator !. The sum operator + is excluded since otherwise, in situations different worlds have different explanations ($t_1, \ldots, t_n$ respectively) for the same formula φ, $\mathsf{Ky}\varphi$ will possibly hold by virtue of a uniform explanation $t_1 + \ldots + t_n$. Moreover, the special element c in E is the self-evident explanation for all formulas in the designated set Λ.

The admissible explanation function $\mathcal{E}$ specifies the set of worlds for both horizontal explanations ($\mathcal{E}(t, \varphi)$) and vertical explanations ($\mathcal{E}(t_2, \langle t_1, \varphi \rangle)$). One may wonder why stop at two levels of explanations in $\mathcal{E}$. This is because understanding why is investigated by studying what sets it apart from knowing-why, and from the common thread found in philosophical viewpoints, the distinguishing feature is: whereas understanding why additionally requires a vertical explanation, knowing why does not. Therefore two levels suffice, just as only one level of explanation is considered in lots of work about knowing why other than knowing that. More levels than two can certainly be explored technically for finding whether there are interesting results, and we leave it to a future occasion. Before further discussion on the conditions of $\mathcal{E}$, we give the truth clauses:

Definition 3.

$\mathcal{M}, w \vDash p$	$\Leftrightarrow$	$w \in V(p)$
$\mathcal{M}, w \vDash \neg\varphi$	$\Leftrightarrow$	$\mathcal{M}, w \nvDash \varphi$
$\mathcal{M}, w \vDash \varphi \wedge \psi$	$\Leftrightarrow$	$\mathcal{M}, w \vDash \varphi$ *and* $\mathcal{M}, w \vDash \psi$
$\mathcal{M}, w \vDash \mathsf{K}\varphi$	$\Leftrightarrow$	$\mathcal{M}, v \vDash \varphi$ *for all* v *such that* wRv
$\mathcal{M}, w \vDash \mathsf{Ky}\varphi$	$\Leftrightarrow$	*(1) there exists* $t \in E$ *such that for all* $v \in W$ *with* wRv,$v \in \mathcal{E}(t, \varphi)$ *(2) for all* $v \in W$ *with* wRv, $\mathcal{M}, v \vDash \varphi$
$\mathcal{M}, w \vDash \mathsf{Uy}\varphi$	$\Leftrightarrow$	*(1) there exist* $t_1, t_2 \in E$ *such that for all* $v \in W$ *with* wRv, $v \in \mathcal{E}(t_2, \langle t_1, \varphi \rangle)$*;* *(2) for all* $v \in W$ *with* wRv, $\mathcal{M}, v \vDash \varphi$

Returning to the explanation function $\mathcal{E}$ in $\mathcal{M}$. The first two conditions are for horizontal explanations as in [24]. The third and the fourth are for vertical applications. The fifth condition says vertical explanations yield horizontal explanations, and do not specify any more conditions on vertical explanations. That is why we call it *a general framework*. Further conditions on $\mathcal{E}$ corresponding to philosophical views discussed in Sect. 2.1 will be referred to later. Note that $\mathcal{E}(\langle t_2, \langle t_1, \varphi \to \psi \rangle\rangle) \cap \mathcal{E}(s_2, \langle s_1, \varphi \rangle) \subseteq \mathcal{E}(t_2, \langle t_1 \cdot s_1, \psi \rangle)$ and $\mathcal{E}(\langle t_2, \langle t_1, \varphi \to \psi \rangle\rangle) \cap \mathcal{E}(s_2, \langle s_1, \varphi \rangle) \subseteq \mathcal{E}(s_2, \langle t_1 \cdot s_1, \psi \rangle)$ hold by condition $3, 4, 5$.

This is the reason the last condition of $\mathcal{E}$ is introduced. As mentioned before, $\mathsf{KyK}\varphi$ corresponds to a why-question: why one knows φ. Typically, the inquirer does not expect the agent to give reasons for why she is not being Gettiered in her belief that φ; rather, she should simply provide her reasons for believing that φ, that is, her justification for φ (some relevant arguments could be found in [14]). Justification logics on the market have such a logical principle: $t : \varphi \to !t : (t : \varphi)$. Following the arguments in [3], we are generally able to substantiate the reasons we have for our knowledge in everyday life. This principle says that $!t$ is always a justification for $t : \varphi$, or $!t$ is an introspective act confirming that $t : \varphi$. It is interesting to note that understanding why so conceived goes with a connection with justification principles: if there is an explanation t for $\mathsf{K}\varphi$, then there always exists a vertical explanation $!t$ of t so as to bring about $\mathsf{UyK}\varphi$. In other situations,

t being an explanation for the non-epistemic fact φ does not entail t can be transformed into a vertical explanation for that t explains φ.

Based on this general framework, many shades of assumptions mentioned in Sect. 2.1 could be reflected in distinct conditions for $\mathcal{E}$ in some sense:

Grimm's Limited Understanding: $w \in \mathcal{E}(t_1, \varphi) \Longrightarrow \exists t_2, w \in \mathcal{E}(t_2, \langle t_1, \varphi \rangle)$. According to Grimm [8], if, in Pritchard's case, we credit the child with knowing-why, then she truly has some conception of "the why". Properly understood knowing why or limited understanding why equals to $\exists t_1 \mathsf{K} \exists t_2 (t_2 : (t_1 : \varphi)) \wedge \mathsf{K}\varphi$.

Answering Vertical Follow Up Question: $\mathcal{E}(t_2, \langle t_1, \varphi \rangle) \subseteq \mathcal{E}(t_1, \varphi) \cap \mathcal{E}(t_2, \mathsf{Ky}\varphi)$. We might think an explanation for why φ could also constitute a propositional justification for why φ is known. Literature on epistemology beginning with [2] makes a distinction between propositional and doxastic justification. The distinction is that: one can have propositional justification without actually believing it. Hence we could assume $\mathsf{KyKy}\varphi$ does not say that the agent knows why she "knows" why φ, but rather she knows why she knows "why" φ, that is, she can seek a convincing propositional justification for her knowing why, so as to answer a relevant vertical why-question.

Answering What If Question: $\mathcal{E}(t_2, \langle t_1, \varphi \rangle) \subseteq \mathcal{E}(t_1, \varphi) \cap \mathcal{E}(t_2, \neg\varphi)$. An example from [11] is adapted: a firm hires Jones because he had extensive prior experience. Moreover, his other credentials were fairly nondescript such that he would not have been hired had he lacked this experience. Agent a knowing that Jones is hired because of his superior experience knows why Jone is hired. Agent b understanding why knows that other hiring criteria (e.g. education) as the deciding factor could explains why it was not the case that Jone is hired. Note that $w \in \mathcal{E}(t_2, \neg\varphi)$ does not entail $w \vDash \neg\varphi$. The understanding models do not assume the factivity of horizontal explanations.

4 An Axiomatization

More conditions on $\mathcal{E}$ may invoke more debates. In this section we provide a sound and complete axiomatization for the general case, that is the system SUY:

	Axiom Schemes		
(TAU)	Propositional Tautologies	(KYK)	$\mathsf{Ky}(\varphi \to \psi) \to (\mathsf{Ky}\varphi \to \mathsf{Ky}\psi)$
(K)	$\mathsf{K}(\varphi \to \psi) \to (\mathsf{K}\varphi \to \mathsf{K}\psi)$	(UYK1)	$\mathsf{Uy}(\varphi \to \psi) \to (\mathsf{Ky}\varphi \to \mathsf{Uy}\psi)$
(T)	$\mathsf{K}\varphi \to \varphi$	(UYK2)	$\mathsf{Ky}(\varphi \to \psi) \to (\mathsf{Uy}\varphi \to \mathsf{Uy}\psi)$
(4)	$\mathsf{K}\varphi \to \mathsf{KK}\varphi$	(UK)	$\mathsf{Uy}\varphi \to \mathsf{Ky}\varphi$
(5)	$\neg\mathsf{K}\varphi \to \mathsf{K}\neg\mathsf{K}\varphi$	(4*)	$\bigcirc\varphi \to \mathsf{K}\bigcirc\varphi$ (for $\bigcirc = \mathsf{Ky}, \mathsf{Uy}$)
(IMP)	$\mathsf{Ky}\varphi \to \mathsf{K}\varphi$	(KYU)	$\mathsf{Ky}\bigcirc\varphi \to \mathsf{Uy}\bigcirc\varphi$(for $\bigcirc = \mathsf{K}, \mathsf{Ky}, \mathsf{Uy}$)
	Rules		
(MP)	Modus Ponens	(NE)	If $\varphi \in \Lambda$, then $\vdash \mathsf{Ky}\varphi$
(N)	$\vdash \varphi \Rightarrow \vdash \mathsf{K}\varphi$		

It is worth noting that the axiom (KYU) expresses that "understanding why" is necessary for "knowing why" in epistemic situations, which corresponds to *epistemic introspection* condition in the model. $\mathsf{KyUy}\varphi \to \mathsf{UyUy}\varphi$, as an instantiation of (KYU), suggests that the introspection of Uy (i.e. $\mathsf{Uy}\varphi \to \mathsf{UyUy}\varphi$) will be obtained once we accept $\mathsf{Uy}\varphi \to \mathsf{KyUy}\varphi$ as a reasonable new axiom. However both are not valid without further conditions on **ELUy** models.

Theorem 1. *SUY is sound over **ELUy** models.*

Definition 4. *Let Ω be the set of all maximal SUY-consistent sets of formulas. The **canonical model** $\mathcal{M}^c$ for SUY is a tuple $(W^c, E^c, \mathcal{F}^c, R^c, \mathcal{E}^c, V^c)$ where:*

- *E^c is defined in BNF: $t ::= c \mid \varphi \mid (t \cdot t) \mid !t$ where $\varphi \in L^P$.*
- *$W^c := \{\langle \Gamma, F, G, f, g, h\rangle \mid \langle \Gamma, F, G\rangle \in \Omega \times \mathcal{P}(E^c \times L^P) \times \mathcal{P}(E^c \times (E^c \times L^P)), f : \{\varphi \mid \mathsf{Ky}\varphi \in \Gamma\} \to E^c, g : \{\varphi \mid \mathsf{Uy}\varphi \in \Gamma\} \to E^c, h : \{(g(\varphi), \varphi) \mid \mathsf{Uy}\varphi \in \Gamma\} \to E^c$ such that f and g satisfy the following conditions$\}$*

   1. *If $\langle t, \varphi \to \psi\rangle, \langle s, \psi\rangle \in F$, then $\langle t \cdot s, \psi\rangle \in F$.*
   2. *If $\varphi \in \Lambda$, then $\langle c, \varphi\rangle \in F$.*
   3. *If $\langle t_2, \langle t_1, \varphi \to \psi\rangle\rangle \in G, \langle s, \psi\rangle \in F$, then $\langle t_2, \langle t_1 \cdot s, \psi\rangle\rangle \in G$.*
   4. *If $\langle t, \varphi \to \psi\rangle \in F, \langle s_2, \langle s_1, \psi\rangle\rangle \in G$, then $\langle s_2, \langle t \cdot s_1, \psi\rangle\rangle \in G$.*
   5. *$\langle t_2, \langle t_1, \varphi\rangle\rangle \in G$ implies $\langle t_1, \varphi\rangle \in F$.*
   6. *$\langle t, \bigcirc\varphi\rangle \in F$ implies $\langle !t, \langle t, \bigcirc\varphi\rangle\rangle \in G$ for $\bigcirc = \mathsf{K}, \mathsf{Ky}, \mathsf{Uy}$.*
   7. *$\mathsf{Ky}\varphi \in \Gamma$ implies $\langle f(\varphi), \varphi\rangle \in F$.*
   8. *$\mathsf{Uy}\varphi \in \Gamma$ implies $\langle h(g(\varphi), \varphi), \langle g(\varphi), \varphi\rangle\rangle \in G$.*

- *$\langle \Gamma, F, G, f, g, h\rangle R^c \langle \Delta, F', G', f', g', h'\rangle$ iff (1) $\{\varphi \mid \mathsf{K}\varphi \in \Gamma\} \subseteq \Delta$, and (2) $f = f', g = g', h = h'$.*
- *$\mathcal{E}^c : E^c \times (L^P \cup \langle E^c \times L^P\rangle) \to 2^{W^c}$ is defined by*

$$\begin{cases} \mathcal{E}^c(t, \varphi) = \{\langle \Gamma, F, G, f, g, h\rangle \mid \langle t, \varphi\rangle \in F\} \\ \mathcal{E}^c(t_2, \langle t_1, \varphi\rangle) = \{\langle \Gamma, F, G, f, g, h\rangle \mid \langle t_2, \langle t_1, \varphi\rangle\rangle \in G\} \end{cases}$$

- *$V^c(p) = \{\langle \Gamma, F, G, f, g, h\rangle \mid p \in \Gamma\}$.*

In the construction, the definition of E^c and W^c are based on those in [24]. Since the nested explanations are needed here, we introduce $!t$ in E^c. For each world in W^c, it contains information about the horizontal and/or vertical explanations for all Ky and Uy formulas belonging to it. More specifically, f is a witness function picking one horizontal t for each formula in $\{\varphi \mid \mathsf{Ky}\varphi \in \Gamma\}$, while g picking one horizontal t_1 for every $\varphi \in \{\varphi \mid \mathsf{Uy}\varphi \in \Gamma\}$. h is a witness function picking one vertical explanation t_2 for each pair $\langle t_1, \varphi\rangle$ in $\{\langle t, \varphi\rangle \mid \mathsf{Uy}\varphi \in \Gamma \ \& \ g(\varphi) = t\}$. Note that the cases of $\mathsf{Ky}\varphi$ and $\mathsf{Uy}\varphi$ can have different horizontal explanations for φ, i.e. we can have $\langle f(\varphi), \varphi\rangle \in F \ \& \ \langle g(\varphi), \varphi\rangle \in F \ \& \ f(\varphi) \neq g(\varphi)$. The following shows such W^c is indeed nonempty.

Definition 5. *Given any $\Gamma \in \Omega$, construct $F^\Gamma, G^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma$ as follows:*

- *$F_0^\Gamma = \{\langle \varphi, \varphi\rangle \mid \mathsf{Ky}\varphi \in \Gamma\} \cup \{\langle c, \varphi\rangle \mid \varphi \in \Lambda\}$, $G_0^\Gamma = \{\langle \varphi \cdot \varphi, \langle !\varphi, \varphi\rangle\rangle \mid \mathsf{Uy}\varphi \in \Gamma\}$*
- *$F_{n+1}^\Gamma = F_n^\Gamma \cup \{\langle t \cdot s, \psi\rangle \mid \langle t, \varphi \to \psi\rangle, \langle s, \varphi\rangle \in F_n^\Gamma$ for some $\varphi\} \cup \{\langle t_1, \varphi\rangle \mid \langle t_2, \langle t_1, \varphi\rangle\rangle \in G_n^\Gamma\}$*
- *$G_{n+1}^\Gamma = G_n^\Gamma \cup \{\langle t_2, \langle t_1 \cdot s, \psi\rangle\rangle \mid \langle t_2, \langle t_1, \varphi \to \psi\rangle\rangle \in G_n^\Gamma, \langle s, \varphi\rangle \in F_n^\Gamma$ for some $\varphi\} \cup \{\langle s_2, \langle t \cdot s_1, \psi\rangle\rangle \mid \langle t, \varphi \to \psi\rangle \in F_n^\Gamma, \langle s_2, \langle s_1, \varphi\rangle\rangle \in G_n^\Gamma$ for some $\varphi\} \cup \{\langle !t, \langle t, \bigcirc\varphi\rangle\rangle \mid \langle t, \bigcirc\varphi\rangle \in F_n^\Gamma$ for $\bigcirc = \mathsf{K}, \mathsf{Ky}, \mathsf{Uy}\}$*

- $F^{\Gamma} = \bigcup_{n \in \mathbb{N}} F_n^{\Gamma}$, $G^{\Gamma} = \bigcup_{n \in \mathbb{N}} G_n^{\Gamma}$
- $f^{\Gamma} : \{\varphi \mid \mathsf{Ky}\varphi \in \Gamma\} \to E^c, f^{\Gamma}(\varphi) = \varphi$.
- $g^{\Gamma} : \{\varphi \mid \mathsf{Uy}\varphi \in \Gamma\} \to E^c, g^{\Gamma}(\varphi) = !\varphi$.
- $h^{\Gamma} : \{(g^{\Gamma}(\varphi), \varphi) \mid \mathsf{Uy}\varphi \in \Gamma\} \to E^c, h^{\Gamma}(!\varphi, \varphi) = \varphi \cdot \varphi$.

Proposition 1. *For any $\Gamma \in \Omega$, $\langle \Gamma, F^{\Gamma}, G^{\Gamma}, f^{\Gamma}, g^{\Gamma}, h^{\Gamma} \rangle \in W^c$.*

Proof. We show that the conditions $1-8$ in the definition of W^c are all satisfied. However, for lack of space, merely selected conditions are discussed below:

- For the condition 3, suppose $\langle t_2, \langle t_1, \varphi \to \psi \rangle \rangle \in G^{\Gamma}$, $\langle s, \psi \rangle \in F^{\Gamma}$. Then there exist $k, l \in \mathbb{N}$ such that $\langle t_2, \langle t_1, \varphi \to \psi \rangle \rangle \in G_k^{\Gamma}$, $\langle s, \psi \rangle \in F_l^{\Gamma}$. Assume without loss of generality that $k > l$. Then we get $\langle t_2, \langle t_1 \cdot s, \psi \rangle \rangle \in G_{k+1}^{\Gamma}$ by the construction. Therefore $\langle t_2, \langle t_1 \cdot s, \psi \rangle \rangle \in G^{\Gamma}$.
- For the condition 8, suppose $\mathsf{Uy}\varphi \in \Gamma$. Then we get $\langle \varphi \cdot \varphi, \langle !\varphi, \varphi \rangle \rangle \in G^{\Gamma}$ by the constructions of G_0^{Γ} and G^{Γ}. Moreover, we have $\langle h^{\Gamma}(g^{\Gamma}(\varphi), \varphi), \langle g(\varphi), \varphi \rangle \rangle \in G^{\Gamma}$ by the construction of g^{Γ} and h^{Γ}. □

For the construction of R^c in the canonical model in Definition 4, we claim:

Proposition 2. *R^c is an equivalence relation.*

Proof. It is trivial by the construction of R^c and axioms (4) and (5). □

As for the construction of $\mathcal{E}^c$, we can check the following without special tricks:

Proposition 3. *$\mathcal{E}^c$ satisfies all the conditions in* **ELUy** *model definition.*

Hence the canonical model is well-defined, based on Proposition 1, 2 and 3.

Proposition 4. *The canonical model $\mathcal{M}^c$ is well-defined.*

Now we prove the existence lemmas for K, Ky and Uy respectively.

Lemma 1 (K Existence Lemma). *For any $\langle \Gamma, F, G, f, g, h \rangle \in W^c$, if $\widehat{\mathsf{K}}\varphi \in \Gamma$, then there exists a $\langle \Delta, F', G', f', g', h' \rangle \in W^c$ such that $\langle \Gamma, F, G, f, g, h \rangle R^c \langle \Delta, F', G', f', g', h' \rangle$, and $\varphi \in \Delta$.*

Proof. (Sketch) Suppose $\widehat{\mathsf{K}}\varphi \in \Gamma$. Let $\Delta^- = \{\psi \mid \mathsf{K}\psi \in \Gamma\} \cup \{\varphi\}$. First, Δ^- is consistent. The proof is routine by (K) and (N). Next we extend Δ^- into an MCS Δ. Finally, we construct F', G', f', g' and h' to form a world in W^c. We can simply let $F' = F, G' = G$, and $f' = f, g' = g, h' = h$. □

In order to refute $\mathsf{Ky}\psi$ while keeping $\mathsf{K}\psi$ semantically, we could construct an accessible world where the horizontal explanation for ψ is not identical to that at the current world.

Lemma 2 (Ky Existence Lemma). *For any $\langle \Gamma, F, G, f, g, h \rangle \in W^c$ where $\mathsf{K}\psi \in \Gamma$, if $\mathsf{Ky}\psi \notin \Gamma$, then for any $\langle t, \psi \rangle \in F$, there exists a $\langle \Delta, F', G', f', g', h' \rangle \in W^c$ such that $\langle t, \psi \rangle \notin F'$ and $\langle \Gamma, F, G, f, g, h \rangle R^c \langle \Delta, F', G', f', g', h' \rangle$.*

Proof. Suppose $\mathsf{Ky}\psi \notin \varGamma$, we construct $\langle \varDelta, F', G', f', g', h'\rangle$ as follows.

- $\varDelta = \varGamma$
- $F' = \{\langle s, \varphi\rangle \mid \langle s, \varphi\rangle \in F \text{ and } \mathsf{Ky}\varphi \in \varGamma\}$
- $G' = \{\langle s', \langle s, \varphi\rangle\rangle \mid \langle s', \langle s, \varphi\rangle\rangle \in G \text{ and } \mathsf{Ky}\varphi \in \varGamma\}$
- $f' : \{\varphi \mid \mathsf{Ky}\varphi \in \varDelta\} \to E^c$ is defined as: $f'(\varphi) = f(\varphi)$
- $g' : \{\varphi \mid \mathsf{Uy}\varphi \in \varDelta\} \to E^c$ is defined as: $g'(\varphi) = g(\varphi)$
- $h' : \{(g'(\varphi), \varphi) \mid \mathsf{Uy}\varphi \in \varDelta\} \to E^c$ is defined as: $h'(g'(\varphi), \varphi) = h(g(\varphi), \varphi)$

The main idea behind the constructions of F' and G' is to "carefully" delete all horizontal explanations for $\{\psi \mid \mathsf{Ky}\psi \notin \varGamma\}$. Clearly $\langle t, \psi\rangle \notin F'$ for any $\langle t, \psi\rangle \in F$ by the construction. In order to complete this proof, firstly, we show that $\langle \varDelta, F', G', f', g', h'\rangle \in W^c$ by checking the conditions $1-8$ in the definition of W^c. Only the last case is written below:

- For condition 8, suppose $\mathsf{Uy}\varphi \in \varDelta$. Then we get $\mathsf{Uy}\varphi \in \varGamma$ by $\varGamma = \varDelta$, thus $\langle h(g(\varphi), \varphi), \langle g(\varphi), \varphi\rangle\rangle \in G$. By (UK) and the property of MCS, we have $\mathsf{Ky}\varphi \in \varDelta$, so $\langle h'(g'(\varphi), \varphi), \langle g'(\varphi), \varphi\rangle\rangle = \langle h(g(\varphi), \varphi), \langle g(\varphi), \varphi\rangle\rangle \in G'$.

Secondly, $\langle \varGamma, F, G, f, g, h\rangle R^c \langle \varDelta, F', G', f', g', h'\rangle$ holds:

- Since $\varDelta = \varGamma$, obviously we have $\{\varphi \mid \mathsf{K}\varphi \in \varGamma\} \subseteq \varDelta$.
- Since $\varDelta = \varGamma$, it is clear that $dom(f) = dom(f')$, and $dom(g) = dom(g')$. Then for any $\varphi \in \{\varphi \mid \mathsf{Ky}\varphi \in \varDelta\}$, by definition of f', we have $f(\varphi) = f'(\varphi)$. Similarly, for any $\varphi \in \{\varphi \mid \mathsf{Uy}\varphi \in \varDelta\}$, we have $g(\varphi) = g'(\varphi)$, and so $dom(h) = dom(h')$. Then by definition of h', we have $h'(g'(\varphi), \varphi) = h(g(\varphi), \varphi)$. Hence $f = f'$, $g = g'$ and $h = h'$. □

Similarly, to refute $\mathsf{Uy}\chi$ while keeping $\mathsf{Ky}\chi$, we construct an accessible world where the vertical explanation for χ is not identical to that at the current world.

Lemma 3 (Uy Existence Lemma). *For any $\langle \varGamma, F, G, f, g, h\rangle \in W^c$ where $\mathsf{Ky}\chi \in \varGamma$, if $\mathsf{Uy}\chi \notin \varGamma$, then for any $\langle s, \langle t, \chi\rangle\rangle \in G$, there exists a $\langle \varDelta, F', G', f', g', h'\rangle \in W^c$ such that $\langle s, \langle t, \chi\rangle\rangle \notin G'$ and $\langle \varGamma, F, G, f, g, h\rangle R^c \langle \varDelta, F', G', f', g', h'\rangle$.*

Proof. Suppose $\mathsf{Uy}\chi \notin \varGamma$, note that $\chi \neq \bigcirc\varphi$ for any φ since $\mathsf{Ky}\chi \in \varGamma$. We construct $\langle \varDelta, F', G', f', g', h'\rangle$ by deleting all current vertical explanations for χ:

- $\varDelta = \varGamma$
- $F' = \{\langle s, \varphi\rangle \mid \langle s, \varphi\rangle \in F \text{ and } \mathsf{Ky}\varphi \in \varGamma\}$
- $X = \{\langle s', \langle s, \varphi\rangle\rangle \mid \langle s', \langle s, \varphi\rangle\rangle \in G \text{ and } \mathsf{Uy}\varphi \notin \varGamma \text{ and } \varphi \neq \bigcirc\psi \text{ for any } \psi\}$
- $G' = G \setminus X$
- $f' : \{\varphi \mid \mathsf{Ky}\varphi \in \varDelta\} \to E^c$ is defined as: $f'(\varphi) = f(\varphi)$
- $g' : \{\varphi \mid \mathsf{Uy}\varphi \in \varDelta\} \to E^c$ is defined as: $g'(\varphi) = g(\varphi)$
- $h' : \{(g'(\varphi), \varphi) \mid \mathsf{Uy}\varphi \in \varDelta\} \to E^c$ is defined as: $h'(g'(\varphi), \varphi) = h(g(\varphi), \varphi)$

Clearly, for each $\langle s, \langle t, \chi\rangle\rangle \in G$, we have $\langle s, \langle t, \chi\rangle\rangle \notin G'$ by the construction. In order to complete remaining proof, firstly, we show that $\langle \varDelta, F', G', f', g', h'\rangle \in W^c$, i.e. this tuple satisfies the conditions $1-8$ in the definition of W^c. We omit some cases due to limited space:

- For the condition 4, suppose $\langle t, \varphi \to \psi\rangle \in F' \subseteq F, \langle s_2, \langle s_1, \varphi\rangle\rangle \in G' \subseteq G$, then $\mathsf{Ky}(\varphi \to \psi)$ and $\langle s_2, \langle t \cdot s_1, \psi\rangle\rangle \in G$. Moreover either $\mathsf{Uy}\varphi \in \Gamma$ or $\varphi = \bigcirc\psi$ for some ψ. If $\mathsf{Uy}\varphi \in \Gamma$, we have $\mathsf{Uy}\psi \in \Gamma$ due to the axiom (`UYK2`) and the property of MCS, hence $\langle s_2, \langle t \cdot s_1, \psi\rangle\rangle \in G \setminus X = G'$. If $\varphi = \bigcirc\psi$ for some ψ, $\langle s_2, \langle t \cdot s_1, \psi\rangle\rangle \in G'$ holds clearly.
- For the condition 5, suppose $\langle t_2, \langle t_1, \varphi\rangle\rangle \in G' \subseteq G$. Then we have $\langle t_1, \varphi\rangle \in F' \subseteq F$ and $\mathsf{Uy}\varphi \in \Gamma$. So $\mathsf{Ky}\varphi \in \Gamma$ by (`UK`), which means $\langle t_1, \varphi\rangle \in F'$.

Secondly, we can show $\langle \Gamma, F, G, f, g, h\rangle R^c \langle \Delta, F', G', f', g', h'\rangle$ as above. □

Lemma 4 (Truth Lemma). *For all φ, $\langle \Gamma, F, G, f, g, h\rangle \vDash \varphi$ iff $\varphi \in \Gamma$.*

Proof. The proof is by induction on the structure of φ. The atomic case and boolean cases are routine. For the case of $\varphi = \mathsf{K}\psi$, it is clear by Lemma 1. For the case of $\varphi = \mathsf{Ky}\psi$, the proof is not hard with the help of Lemma 1 and 2.

For the case of $\mathsf{Uy}\psi$,

- $\Longleftarrow$: Suppose $\mathsf{Uy}\psi \in \Gamma$. Then for any $\langle \Delta, F', G', f', g', h'\rangle$ such that $\langle \Gamma, F, G, f, g, h\rangle R^c \langle \Delta, F', G', f', g', h'\rangle$, we get $\mathsf{Uy}\psi \in \Delta$, which implies $\varphi \in \Delta$ by (`4*`), (`UK`), (`IMP`), (`T`), and the property of MCS. Thus $\langle \Gamma, F, G, f, g, h\rangle \vDash \psi$ by IH. Furthermore, we have $\langle\langle h(g(\psi), \psi), \langle g(\psi), \psi\rangle\rangle \in G$, $\langle\langle h'(g'(\psi), \psi), \langle g'(\psi), \psi\rangle\rangle \in G'$ and $g = g', h = h'$, which means there exists $g(\psi) = g'(\psi) \in E^c$ such that $h(g(\psi), \psi) = h'(g'(\psi), \psi) \in E^c$, and $\langle \Delta, F', G', f', g', h'\rangle \in \mathcal{E}^c(h(g(\psi), \psi), \langle g(\psi), \psi\rangle)$. Hence $\langle \Gamma, F, G, f, g, h\rangle \vDash \mathsf{Uy}\psi$.
- $\Longrightarrow$: Suppose $\mathsf{Uy}\psi \notin \Gamma$. Then we have the following three cases:
  - $\mathsf{K}\psi \notin \Gamma$. By Lemma 1, we have $\langle \Gamma, F, G, f, g, h\rangle \nvDash \mathsf{K}\psi$, thus $\langle \Gamma, F, G, f, g, h\rangle \nvDash \mathsf{Uy}\psi$.
  - $\mathsf{K}\psi \in \Gamma$ and $\mathsf{Ky}\psi \notin \Gamma$. By Lemma 2, we have $\langle \Gamma, F, G, f, g, h\rangle \nvDash \mathsf{Ky}\psi$, thus $\langle \Gamma, F, G, f, g, h\rangle \nvDash \mathsf{Uy}\psi$.
  - $\mathsf{K}\psi \in \Gamma$, and $\mathsf{Ky}\psi \in \Gamma$. If $\langle t_2, \langle t_1, \psi\rangle\rangle \notin G$ for any $t_1, t_2 \in E^c$, then by the semantics, $\langle \Gamma, F, G, f, g, h\rangle \nvDash \mathsf{Uy}\psi$. If there exist t_1 and t_2 with $\langle t_2, \langle t_1, \psi\rangle\rangle \in G$, then we complete the proof by Lemma 3.

□

Theorem 2. *The system* SUY *is complete over* **ELUy** *models.*

5 Conclusions and Future Work

Understanding why is considered requiring more than knowing why. But philosophers differ on the nature of this "more". Inspired by non-reductionists, we think of this "more" as providing more explanations to more questions. We build up a general framework by introducing vertical explanations, and show that it could accommodate different points on the nature of understanding why via adding different conditions in the models. Only one axiomatization for models without these conditions is provided. Hence one of the future directions is to develop axiomatizations with more reasonable conditions on $\mathcal{E}$ in the models. Besides, understanding why can be studied on basis of knowing how as well. It should not be overlooked that philosophical literatures are glutted with expressions such as "understanding why requires *knowing how* cause and effect are related".

References

1. Bermúdez, J.L.: Philosophy of psychology: A contemporary introduction. Routledge (2004)
2. Firth, R.: Are epistemic concepts reducible to ethical concepts? In: Values and morals, pp. 215–229. Springer (1978)
3. Fitting, M.: A logic of explicit knowledge. Logica Yearbook pp. 11–22 (2004)
4. Gattinger, M., Wang, Y.: How to agree without understanding each other: Public announcement logic with boolean definitions. Electronic Proceedings in Theoretical Computer Science **297**, 206–220 (2019)
5. Gordon, E.C.: Is there propositional understanding? Logos & Episteme **3**(2), 181–192 (2012)
6. Greco, J.: Episteme: Knowledge and understanding. Virtues and their vices pp. 285–302 (2014)
7. Grimm, S.R.: Is understanding a species of knowledge? The British Journal for the Philosophy of Science **57**(3), 515–535 (2006)
8. Grimm, S.R.: Understanding as knowledge of causes. In: Virtue epistemology naturalized, pp. 329–345. Springer (2014)
9. Hills, A.: Understanding why. Noûs **49**(2), 661–688 (2015)
10. Hintikka, J.: New foundations for a theory of questions and answers. In: Questions and answers, pp. 159–190. Springer (1983)
11. Khalifa, K.: Understanding, explanation, and scientific knowledge. Cambridge University Press (2017)
12. Lawler, I.: Reductionism about understanding why. In: Proceedings of the Aristotelian Society. vol. 116, pp. 229–236. Oxford University Press (2016)
13. Lawler, I.: Understanding why, knowing why, and cognitive achievements. Synthese **196**(11), 4583–4603 (2019)
14. McKinnon, R.: How do you know that 'how do you know?'challenges a speaker's knowledge? Pacific Philosophical Quarterly **93**(1), 65–83 (2012)
15. Palmira, M.: Defending nonreductionism about understanding. Thought: A Journal of Philosophy **8**(3), 222–231 (2019)
16. Pritchard, D.: Knowing the answer, understanding and epistemic value. Grazer Philosophische Studien **77**(1), 325–339 (2008)
17. Pritchard, D.: Knowledge and understanding. In: Virtue epistemology naturalized, pp. 315–327. Springer (2014)
18. Ross, L.D.: Is understanding reducible? Inquiry **63**(2), 117–135 (2020)
19. Skow, B.: Reasons why. Oxford University Press (2016)
20. Sliwa, P.: Iv—understanding and knowing. In: Proceedings of the Aristotelian Society. vol. 115, pp. 57–74. Oxford University Press Oxford, UK (2015)
21. Sullivan, E.: Understanding: not know-how. Philosophical Studies **175**(1), 221–240 (2018)
22. Wang, Y.: Beyond knowing that: a new generation of epistemic logics. In: Jaakko Hintikka on Knowledge and Game-Theoretical Semantics, pp. 499–533. Springer (2018)
23. Woodward, J.: Making things happen: A theory of causal explanation. Oxford university press (2005)
24. Xu, C., Wang, Y., Studer, T.: A logic of knowing why. Synthese pp. 1–27 (2019)

Assessing the Effect of Text Type on the Choice of Linguistic Mechanisms in Scientific Publications

Iverina Ivanova[0000−0003−2026−9448]

Goethe University Frankfurt,
Norbert-Wollheim-Platz 1, 60323
Frankfurt am Main, Germany
{I.Ivanova}@em.uni-frankfurt.de

Abstract. In this paper, we report a qualitative and quantitative evaluation of a hand-crafted set of discourse features and their interaction with different text types. To be more specific, we compared two distinct text types—scientific abstracts and their accompanying full texts—in terms of linguistic properties, which include, among others, sentence length, coreference information, noun density, self-mentions, noun phrase count, and noun phrase complexity. Our findings suggest that abstracts and full texts differ in three mechanisms which are size and purpose bound. In abstracts, nouns tend to be more densely distributed, which indicates that there is a smaller distance between noun occurrences to be observed because of the compact size of abstracts. Furthermore, in abstracts we find a higher frequency of personal and possessive pronouns which authors use to make references to themselves. In contrast, in full texts we observe a higher frequency of noun phrases. These findings are our first attempt to identify text type motivated linguistic features that can help us draw clearer text type boundaries. These features could be used as parameters during the construction of systems for writing evaluation that could assist writing tutors in text analysis, or as guides in linguistically-controllable neural text generation systems.

Keywords: Linguistic Mechanisms · Discourse Coherence · Text Types · Linguistic Features for Text Generation · Noun Density · Self-mentions

1 Introduction

Writing is a creative process which involves not only the generation of a sequence of sentences, but also a mechanism of how these sentences relate to each other. In fact, how to produce a coherent text has always been a challenge for all those who are actively involved in the creative writing process, for example, instructors, researchers working on scientific papers, as well as students who make their first endeavours in academic writing [4].

The recent development in neural transformer-based language modeling [3] has made tremendous progress in automatic generation of coherent texts. Radford et

al.[14], for instance, have successfully demonstrated that neural text generation can produce syntactically valid and meaningful texts. Justified concerns about the large-scale generation of disinformation have been raised already[1] and it is quite certain that sequence prediction models will soon find their way also into the fields of computer-assisted writing. First interactive editors, for example, the one by Wolf et al.[19] propose automatic completion of text fragments and incorporate next sentence prediction objectives in their underlying models [3].[2] Although these text predictions incorporate some notion of "discourse understanding", they still suffer from being controllable as text productions are greedily chosen and typically represent only random predictions. Keskar et al. [10] make one of the first attempts to encode a way of control mechanism in their language model objectives, however, to-date it is still an unresolved problem how these powerful models can be used to conform to text-level coherence and what exactly the linguistic factors are that determine text coherence for neural language models. In this paper, we try to fill the gap between the formal theoretical approaches to modeling discourse coherence on the one hand, for example, the one by Grosz et al. [5], and the latest neural advancements on the other, which do not incorporate any linguistic signals other than plain n-grams. We set the scope of our work into the context of Benz and Jasinskaja [2], who argue that a text is produced as an answer to a question and that the text structure, as well as the choice of language expression in terms of information packaging and the use of cohesive devices, is constrained by the communicative goal of the **type of text**, which has also been pointed out by von Stutterheim and Klein [17].
The goal and contribution of our present study is to verify this claim and to **identify distinctive linguistic features** for two different text types. These features can be applied to other text genres, for example, academic essays, or even to discourse segments in scientific publications such as introductions, methods, discussions, conclusions, and used as parameters during the development of tools for automated writing evaluation [12] or automated essay scoring [8]. To be more specific, such distinctive features could provide a better understanding of the typical underlying linguistic characteristics that set one text type or discourse segment apart from another one. This could facilitate the development of more informative tools that can provide hints about the features that are expected to be found in a concrete text type or segment. The presence or the absence of the target features could assist tutors in the analysis and evaluation of the text quality. Furthermore, these features can be employed as an interpretable guiding signal that controls the output of neural text generation systems across the sentence boundary. In this paper, we focus on their identification; their integration into downstream applications is left for future work.

1.1 Related Work

Various attempts have been made to extract distinctive features from different text types, both in purely linguistic contexts but also in text classification

[1] https://openai.com/blog/better-language-models/

[2] https://transformer.huggingface.co/

settings. Previous studies have focused, for example, on the linguistic characteristics that distinguish scientific English from literary English. Ahmad [1], for instance, found that scientific language differs from non-scientific language in the use of impersonal constructions marked by the passive voice, which makes the authors' expression objective; the use of nominalizations, which adds to the technicality in scientific discourse, and the use of hedging as a means of achieving a consensus among scientists on the subject matter under discussion [1]. Other researchers, by contrast, have found out that academic expression is not entirely devoid of authors' presence. In fact, Hyland [7] and Yazılarda et.al. [20] analyse the frequency of self-mentions in research articles and emphasize that authors make use of self-referring words to achieve various rhetorical purposes such as to present the aim of the study, to explain the research procedure, to elaborate on an argument, or to make claims. Others investigate the internal organization of information in academic abstracts by analysing the grammatical and lexical patterns that indicate the problem–solution–evaluation–conclusion moves. Such patterns can be implemented in models that measure the overall text coherence in abstracts by means of automated detection of the moves [13]. von Stutterheim and Klein [17] analyse how the nature of the question constrains the text structure and the choice of referential movements, i.e. what type of information is transferred from one utterance to another and what linguistic devices are adopted to signal these movements by comparing narrative with descriptive texts.

Unlike previous studies which contrast the linguistic characteristics of texts representing different genres and disciplines, our study examines two different text types - pairs of an abstract and its accompanying full text - which both appear in the same corpus of a scientific article but because of the differences in their size and purpose are considered distinct text types. Therefore, the current research seeks, on the one hand, to elaborate on the linguistic mechanisms present in scientific discourse and, on the other hand, to verify if the choice of these mechanisms can be constrained by the text type. To achieve this, we compare the text types in terms of a set of features, which are automatically extracted. Thus, we will find out which of the analysed features are text-specific and will try to explain what justifies their dominance.

2 Experimental Setup

We analyze the two text types on the basis of a predefined, linguistically-motivated set of features reflecting the size and purpose constraints imposed by the text type. In order to extract these features from the target texts, we use automated annotations as a proof-of-concept for the feasibility of experiments involving more data and more features, which is currently beyond the scope of this present study. The purpose of this experiment is to obtain distinctive features that can help us draw clearer text type boundaries that can facilitate both text generation and text evaluation/analysis.

For the purposes of our research we analyze the abstracts and the accompanying full texts of 1,761 scientific papers in the field of computational linguistics available from the ACL Anthology Reference Corpus [3]. The analysis involves an automatic extraction of a set of linguistic features using the `StanfordCoreNLP` module.[4]

Table 1. Our linguistic features involved in this study and how they are measured.

	Feature	Description
1	**Sentence length**	The total number of tokens normalized by the total number of sentences.
2	**Coreference**	The total number of coreference chains normalized by the total number of sentences.
3	**Noun density**	The sum of all token-based differences between noun occurrences normalized by the number of noun occurrences.
4	**Self-mentions**	The total number of self-mention occurrences normalized by the total number of noun phrases.
5	**Noun phrase count**	The total number of noun phrases per document normalized by the total number of tokens.
6	**Noun phrase complexity**	– The total number of embedded that-clauses in the noun phrases normalized by the total number of noun phrases (excluding pronouns). – The total number of embedded past participle clauses normalized by the total number of noun phrases (excluding pronouns). – The total number of embedded to-infinitive clauses normalized by the total number of noun phrases (excluding pronouns).

Table 2. Overview of analysed text types, mean size in tokens, and mean number of sentences.

Text type	Avg length	Avg # of sentences
Abstract	104 tokens	4.97
Main Part	3,262 tokens	151.8

[3] `https://acl-arc.comp.nus.edu.sg/`

[4] `https://stanfordnlp.github.io/CoreNLP/index.html`

A paper's abstract and its main part are considered two distinctive text types as they clearly differ in size (see Table 2) and purpose. Abstracts introduce the target reader to the aim, the methods, the results, and the possible applications of a study in a concise fashion and their aim is to inform and involve the reader in reading on the paper's main part [13]. The main part, by contrast, is intended to inform the reader about the findings of the study in an extended form by providing a comprehensive description of the background, the methods, the results of the study and the possible conclusions that can be drawn from them.

2.1 Linguistic Features under Consideration

Our features include sentence length, coreference, noun density, self-mentions, noun phrase (NP) count, and noun phrase complexity; see Table 1 for a full overview of the features. Such lexical, syntactic, and discourse signals can inform us about the lexical and syntactic sophistication of the texts and could therefore be employed as predictors of text quality and writing proficiency in automated writing tools [11]. In this study, sentence length is measured by the mean number of tokens per sentence and the feature is used for normalization purposes. Coreference is a type of grammatical cohesive device [6] which provides insights into the topic persistence in the texts indicated by the presence of coreferential relations that hold between threads of meaning. The `StanfordCoreNLP` module displays these relations between entities in the form of coreference chains. A coreference chain stands for the relation between an anaphora and its antecedent and a chain can contain two or more mentions of the same entity. Our expectations are that full texts will contain a higher frequency of coreference chains, which could be size motivated, i.e. the longer the text, the higher the frequency of coreference chains. Noun density refers to the mean distance between noun occurrences in the text. Following Witte and Faigley [18], we measure density by calculating the mean number of tokens that occur between nouns–thus, the smaller the distance, the greater the density of nouns. We expect that in abstracts nouns will be more densely distributed, which could be both size and purpose bound. Self-mentions are occurrences of personal pronouns such as *I, me, we, us* and possessive pronouns such as *my, mine, our, ours* that authors use to make self references. Taking Hyland's[7] and Yazilarda et.al's [20] findings into account, we expect that in abstracts there will be a higher occurrence of self-mentions since authors use them to mark rhetorical moves, for example, to introduce their research topic, to explain the research procedure, to emphasize the significance of their research, and to make conclusions based on the research results. NP count refers to the number of NP occurrences in texts. By an NP occurrence we understand the noun head along with its dependents. Since nominalization is a distinctive feature of scientific language [1], we believe that there are NP occurrences in both text types but their frequency will be higher in full texts due to their length. The NP complexity measures the types of modification that are present in the internal structure of NPs. We extract the frequency of embedded finite and non-finite clauses from the NP structures. Considering the nominalization feature, we expect that noun heads in NPs are heavily modified by that-clauses or non-finite clauses introduced by to-infinitive or a past participle in both abstracts and full texts. We would like to check which type of NP modification is predominant in the two target texts.

3 Evaluation

3.1 Quantitative Assessment

The distinctive features are shown in Figure 1. The current results indicate that the two text types differ significantly in three of the analysed features: noun density, NP count, and self-mentions. The abstract and full text samples for each feature were compared pairwise. Abstracts and full texts differ in terms of noun density. The mean number of tokens that appear between nouns in abstracts is lower than that in the accompanying full text. In abstracts, a noun occurs once every 2.91 tokens, whereas in full texts - once every 3.28 tokens. Since the data in both samples was normally distributed, a paired t-test [9] with 0.95 percent confidence interval was conducted. The test result confirmed that the means of the two samples differ significantly with a p-value < 0.05, t = -48.402 and df = 1712. Another distinctive feature is the frequency of NP occurrences. This frequency in full texts is higher (0.46) than that in abstracts (0.41). The data in both samples was normally distributed and the paired t-test confirmed that the difference is again statistically significant with a p-value < 0.05, t = -30.722 and df = 2912.8. Finally, a third distinctive feature is the frequency of self-mentions, which is higher in abstracts (0.07) than in full texts (0.03). The Wilcoxon test [15] was used since the data in the abstract sample was non-normally distributed and its result showed that the difference in means is statistically significant (p < 0.05, V = 1170812). For all the rest of the features, the tests did not reveal any statistically significant differences (cf. Table 3).

Table 3. All linguistic features and their computed mean values per text type.

Feature	Abstract	Main Part
Sentence length	21.14	21.56
Coreference	0.33	0.35
Noun density	2.91	3.28
Self-mentions	0.07	0.03
NP count	0.41	0.46
Embedded finite clauses	0.05	0.06
Embedded past part clauses	0.08	0.07
Embedded to-inf clauses	0.03	0.03

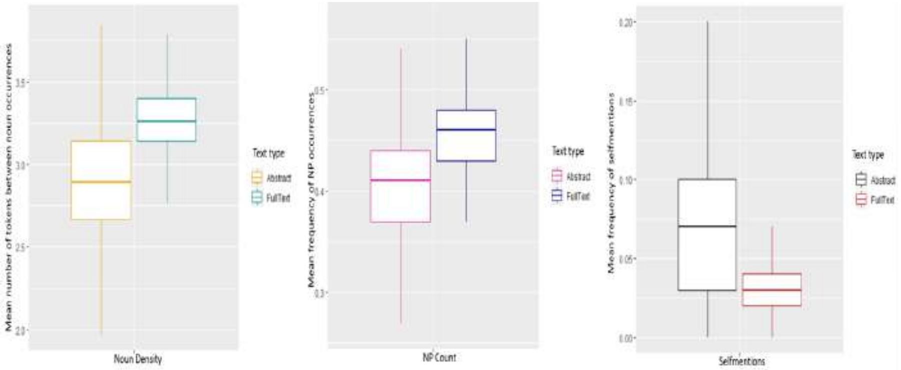


Fig. 1. Distinctive linguistic features between two text types: Abstracts exhibit a lower number of tokens between noun occurrences and a higher frequency of self-mentions, whereas full texts have a higher frequency of noun phrases.

3.2 Discussion

Our findings support the observations of previous work by von Stutterheim and Klein [17] that the communicative goal of the text places constraints on the text structure and the selection of linguistic devices that add to the overall text coherence. It also confirms our preliminary expectations that there are genre-bound differences in terms of linguistic mechanisms. The results from our quantitative study suggest that nouns in abstracts are more densely distributed than those in full texts. This could be explained, on the one hand, by the compact size of abstracts in which authors tend to present the essential points of their research by using nominal forms that are information burdened. On the other hand, it could also be motivated by the text purpose, i.e. abstracts have to be informative per se. Moreover, abstracts and their accompanying full texts also differ in the frequency of NP occurrences. The frequency of NPs could correlate with the text size. The longer the text, the higher the frequency. In abstracts there might be fewer NPs but these NPs might contain a sequence of nouns that modify the noun heads, which could also be a possible explanation for their high noun density. Finally, the third feature on the basis of which the two text types differ is the high frequency of self-mentions in abstracts, which means that authors tend to refer to themselves more often in the abstract than in the main part of the article. This could be motivated by the rhetorical purposes that the authors want to achieve, namely, to present in a succinct and engaging form the goal of the conducted research, to introduce us to the experimental setup, the methods, and the results, and to provide us with their interpretation of the results. The use of self-mentions also makes the expression more personal and thus improves the writer-reader interaction.

Contrary to our initial assumptions, the two text types did not display any significant differences in terms of sentence length, the frequency of coreferential relations, or the complexity of NPs marked by embedded finite and non-finite clauses. This could be due to the data sparsity. Nevertheless, we believe that these findings provide more insights into the text-specific linguistic mechanisms and help us understand better how the aim of the text imposes constraints on the language expression. They also make us more confident in our claims that automated tools for writing evaluation or text generation can greatly benefit from such text-specific characteristics. The integration of such pre-defined distinctive features in natural language processing tools could improve their functionality by enabling a more fine-grained analysis of the most common linguistic mechanisms present in a particular text type. The automated detection of these features in an input text could give us an informative feedback on whether the author of the text has achieved the communicative goal of the text or the discourse segment. What is more, these linguistic properties could also find a good application as predictors of the underlying syntactic and discourse mechanisms when integrated in natural language text generation systems.

4 Conclusion and Future Work

We have investigated the effect of text type on the choice of linguistic mechanisms. By comparing scientific abstracts and their accompanying full texts, we found that the size and the communicative goal of the text type could influence the linguistic devices that are employed in the text. The results showed that the features of noun density and self-mentions are predominant mechanisms in abstracts and are both size and purpose bound. The high frequency of noun phrases is a predominant feature of the full text, which turned out to be size bound.

The current study requires further investigation in various directions. First, we would like to investigate other types of features which are related to the internal structure of the NP such as the mean NP length measured by the mean number of tokens, as well as the frequency of nominal and adjectival modification. Second, another feature that is associated with scientific language is hedging. Hedging expresses the degree of confidence with which authors present the information in their studies. Our expectation is that there will be a higher occurrence of hedges in full texts, especially in the Discussion part where authors present their interpretation of the research results. Third, since relations between utterances can be signalled not only by means of reference, but also by lexical markers such as repetitions, synonyms, and antonyms, we would like to see which are the predominant lexical cohesive devices and what is their distribution in abstracts and in the different sections of the accompanying full texts. Finally, the mechanisms and the proposed research methodology of acquiring linguistic features described in this study will be further extended and applied to larger text corpora and other academic genres.

Along with the paper submission, we publicly release the annotations including the source code for use to the linguistic community.

Acknowledgements

I would like to thank Dr. Niko Schenk for his constant support, inspiration, and supervision of the extracted and analysed data. Sincere gratitude to Prof. Gert Webelhuth, Dr. Janina Radó, and Prof. Manfred Sailer for their guidance and assistance in the interpretation of the research results.

References

1. Ahmad, J.(2012). Stylistic Features of Scientific English: A Study of Scientific Research Articles. English Language and Literature Studies, 2(1). https://doi.org/10.5539/ells.v2n1p47
2. Benz, A. and Jasinskaja, K. (2017). Questions under discussion: From sentence to discourse, Discourse Processes, 54:3, 177–186. https://doi.org/10.1080/0163853X.2017.1316038
3. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.(2018). Bert: Pre-training of deep bidirectional transformers for language understanding. `https://arxiv.org/abs/1810.04805`. Last accessed 4 July 2020
4. Flower, L., Hayes, J. (1981). A cognitive process theory of writing. College Composition and Communication, 32(4), 365–387.
5. Grosz, B.J., Joshi, A.K., and Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. Computational Linguistics, 21(2):203–225.
6. Halliday, M. and Hasan, R. (1976). Cohesion in English. Longman Group Ltd London.
7. Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. English for Specific Purposes Volume 20, Issue 3, 2001, pp. 207–226.https://doi.org/10.1016/S0889-4906(00)00012-0
8. Jin, C., He, B., Hui, K., and Sun, L. (2018). TDNN: A two stage deep neural network for prompt-independent automated essay scoring. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1088–1097, Melbourne, Australia, July. Association for Computational Linguistics.
9. Kalpic, D., Hlupic, N., and Lovric, M. (2011). Student's tTests, pages 1559–1563. Springer Berlin Heidelberg, Berlin, Heidelberg.
10. Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., and Socher, R. (2019). CTRL: A conditional transformer language model for controllable generation. CoRR,abs/1909.05858.
11. McNamara, Danielle Mccarthy, Philip. (2010). Linguistic Features of Writing Quality. Written Communication - WRIT COMMUN. 27. pp. 57–86. https://doi.org/10.1177/0741088309351547
12. McNamara, D. S., Graesser, A. C. (2012). Coh-Metrix: An Automated Tool for Theoretical and Applied Natural Language Processing. In P. McCarthy, C. Boonthum-Denecke (Eds.), Applied Natural Language Processing: Identification, Investigation and Resolution, pp. 188–205), Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-60960-741-8.ch011
13. Orasan, Constantin. (2001). Patterns in Scientific Abstracts. Proceedings Corpus Linguistics, 433–445.

14. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners.
15. Rey, D. and Neuhauser, M. (2011). Wilcoxon-Signed-Rank Test, pages 1658–1659. Springer Berlin Heidelberg, Berlin, Heidelberg.
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In I. Guyon, et al.,editors, Advances in Neural Information Processing Systems 30, pages 5998–6008. Curran Associates, Inc.
17. von Stutterheim, C. and Klein, W. (1989). Referential Movement in Descriptive and Narrative Discourse. North-Holland Linguistic Series: Linguistic Variations, Elsevier, Volume 54, 1989, pages 39–76.https://doi.org/10.1016/B978-0-444-87144-2.50005-7
18. Witte, S. P. and Faigley, L. (1981). Coherence, cohesion, and writing quality. College Composition and Communication, 32(2):189–204.
19. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T.,Louf, R., Funtowicz, M., and Brew, J. (2019). Huggingfaces transformers: State-of-the-art natural language processing. https://arxiv.org/abs/1910.03771v4. Last accessed 4 July 2020
20. Yazılarda, A., İşaret, Y., Kullanımı, E.S., Kafes, H. (2017). The use of authorial self-mention words in academic writing. International Journal of Language Academy Volume 5/3, Summer 2017, pp. 165–180. https://doi.org/10.18033/ijla.3532

Cancelling the Maxim of Quantity & Reasoning under Uncertainty

Cathy Agyemang

Carleton University, Ottawa, K1S 5B6, Canada

Abstract. Fox (2014) points out that cancelling the Maxim of Quantity can minimally dissociate between the pragmatic (Gricean and neo-Gricean) and grammatical approaches to scalar implicature. The pragmatic views make no predictions about the availability of scalar implicatures when Quantity is unavailable. The grammatical view, on the other hand, states that scalar implicatures arise from mechanisms independent of conversational maxims and should thus be available. This study adopts Fox's game-show scenario, where participants are tasked with deducing which items have money associated with them. The host, who is reticent about their knowledge, provides partial information (disjunctions and numerals) as hints to help the contestants. Results demonstrate that participants exhaustify the meaning of partially informative statements in order to help them make judgments about the relevant alternatives. The current work provides experimental evidence for the availability of scalar implicatures in contexts where Quantity need not be satisfied.

Keywords: Maxim of Quantity · Scalar Implicature · Exhaustification.

1 Introduction

Scalar implicatures strengthen the basic meaning of a sentence by maximizing the quantity of information communicated to a listener. For example, the logical disjunction *Susie ate cake or ice-cream* can be strengthened to an exclusive disjunction, *Susie ate cake or ice-cream, but not both.* Scalar implicatures are also relevant to reasoning under uncertainty. A Question Under Discussion (QUD) is a choice between relevant alternatives, further specified by the number of alternatives that can answer the question. Thus, an answer will fully or partially satisfy a question by choosing at least one of the specified alternatives [20, 11]. Scalar implicatures reduce the uncertainty in the set of alternatives that could sufficiently answer the QUD.

The derivation of scalar implicatures has been described by two oppositional views: the pragmatic view, specifically, Gricean, neo-Gricean and other pragmatic approaches [10, 12] and the grammatical view [5, 7]. Gricean pragmatics establishes that scalar implicatures are the result of the maxim of Quantity [10]. The maxim of Quantity stipulates that a conversational contribution ought to be maximally informative to the current purposes of the exchange and is based on the assumption of speaker rationality. Under Quantity, the contribution is presumed to represent all of the speaker's knowledge, which strengthens the asserted statement by denying an even stronger claim. That said, the maxim of Quantity alone cannot account for scalar implicatures, as it yields the symmetry problem: the scalar implicature and the scalar alternative are both equally likely inferences based on the basic meaning of the sentence. The neo-Gricean pragmatic approach rectifies this by specifying that the maxim of Quantity operates on a set of formally restricted alternatives [12, 21]. The grammatical approach establishes that scalar implicatures are generated as a result of exhaustification [5, 4]. Grammatical exhaustification is the application of a covert operator *exh*, analogous to the term *only*, which parses the sentence to a strengthened meaning. Here, scalar implicatures should arise independently of the Quantity maxim.

2 Implicature Cancellation

In Grice's formalization of the Cooperative Principle, participants can opt out of the principle and its subsequent maxims [10]. This can be achieved by simply stating that they will not be cooperative towards the conversational goals (e.g., one refusing to disclose information because they were sworn to secrecy). When the speaker is not expected to be fully informative, for example, in the context of a treasure hunt, the grammatical and pragmatic approaches[1] [10, 9, 22, 14] predict that ignorance inferences would not be generated. Using the example of a treasure hunt, the listener would not necessarily know if the speaker was ignorant about where the treasure is buried, it could reasonably be that the speaker does know but will not disclose this information based on the rules of the treasure hunt. Said differently, the listener cannot justifiably generate the inference that the speaker is ignorant about where the treasure is hidden. An ignorance inference can even be explicitly cancelled, as Grice points out, if the speaker were to say that they do know where the treasure is but for the purpose of the game, they won't explicitly tell their conversational partner this information [10, 9]. To this effect, Fox theorizes on the effect that cancelling Quantity has on the types of inferences that a listener can generate from the conversation [8].

The objective of Fox's thought experiment on cancelling Quantity is to demonstrate the difference in how grammatical and pragmatic approaches account for circumstances when a speaker does not need to communicate all the information relevant to the conversational goals. Namely, pragmatic approaches cannot make any predictions as to why scalar implicatures are conceivably still available when Quantity is deactivated.

Fox presents a hypothetical game show scenario, where there are five boxes out of a hundred that have a million dollars inside. The game show host knows which boxes have money inside of them but mentions that they will not explicitly tell the contestants this information, cancelling the ignorance inference. The host

[1] As an anonymous reviewer points out, this statement primarily applies to pragmatic approaches that adopt a less granular view of semantics (see [25] and section 4 for more context.

will provide the contestant with some hints/partial information to help them make a choice. Fox illustrates one potential round of the game show (p. 12):

(1) There is money in box 20 or 25.

The scalar implicature in (1) would be that there is money either in box 20 or there is money in box 25, but not both. An inclusive interpretation is also available and could possibly arise given the parameters of the game (Quantity is cancelled), such as in (2):

(2) There is money in box 20 or 25 or both.

The grammatical approach predicts that scalar implicature should be available despite Quantity being deactivated as scalar implicatures are a feature of the grammar rather than pragmatic strengthening, through grammatical exhaustification *exh*:

(3) $p \vee q$
$exh(p \ \vee \ q)$
$(p \ \vee \ q) \wedge \neg(p \ \wedge \ q)$

Under the pragmatic approaches implicatures are derived from conversational maxims. Since the maxim of Quantity is deactivated in the context of a game show, this view does not make any predictions as to why the listener is still able to conclude that there is money only in box 20 or only box 25 from (1). Based on the assertions made from the pragmatic views, Fox derives the prediction that interpretations under the pragmatic approach should not find any difference in meaning between (1) and (2). Under these views, both (1) and (2) represent an inclusive interpretation of the disjunction under the pragmatic constraints.

Again, these predictions oppose what can be predicted by the grammatical approach. Fox further motivates this by demonstrating that it is acceptable for a contestant to refute that there is money in both box 20 and 25 when given (1) as a hint, where exclusivity is available and it is odd to refute this when given (2) as a hint where it is not [8].

Fox's original characterization only includes disjunctive statements, however, in this thesis, I aim to extend this line of reasoning to numerals, especially since the scalar implicatures that arise from numerals are demonstrably salient and preferred in some contexts [19, 15]. Consider (4)

(4) a. There is money in one box.
b. There is money in *exactly* one box.
c. There is money in *at least* one box.

A strengthened reading as in (4.b) would be a precise answer to the QUD of how many boxes contain a million dollars. This should be greatly preferred to a basic reading as in (4.c). Going further, it is likely that a strengthened meaning derived from a scalar numeral as in (4) compared a disjunctive sentence (1) is more useful to a contestant in this scenario. It would be more informative to know that there is *exactly* one box that contains a million dollars than to known that either box 20 or box 25 *but not both* have a million dollars.

All this being said, Meyer argues that it is unnecessary to state that Quantity is cancelled in a game show scenario [17]. Including Quantity-1 here for convenience, Grice states "Make your contribution as informative as is required for **for the purposes of the conversation**". Providing all of the relevant information to the contestant about where the money is would defeat the purpose of the game. Therefore, Meyer argues Quantity-1 accounts for circumstances outside of cooperative conversation. She also demonstrates out that Fox's game show scenario can be used to defend against the claim that Quantity-2 (Do not make your contribution more informative than is required) and the maxim of Relevance are redundant, as mentioned by Horn [13] and others. If the host is aware of which boxes have money, telling a contestant this as a stronger alternative is certainly relevant to the conversation, yet for the purposes of the game, is more informative than necessary.

Nevertheless, Meyer still found that the pragmatic approach has difficulty to account for the scalar implicature. The non-assertion of either disjunct (e.g., box 20, box 25) is attributed to Quantity-2, where asserting either disjunct (e.g., the money is in box 20) would be a stronger statement than the current conversational goals. This cancels the ignorance inference, as it is uncertain that the speaker is ignorant about which box has money or that the speaker will not disclose this information. If neither disjunct can be asserted, then their conjunction (the money is in box 20 and box 25) also cannot be asserted. Difficulty arises because the non-assertion of the conjunction results from the constraints of the game show scenario, rather than belief that the truth of the stronger conjunctive statement does not hold, which would typically be used to generate the scalar implicature (*not both* in this case). In this case, the pragmatic approach still cannot account for why the scalar implicature should be available in the game show scenario. To rectify this and problems that she additionally finds with the grammatical approach, the author proposes a third approach based of Matrix K Theory (see [17] for a full description).

3 Experiment

I adapted Fox's [8] game show scenario in order to dissociate between the competing views on the nature of scalar implicature. The debate concerning scalar implicatures largely considers contexts where conversational participants are assumed to be maximally informative. The current study aims to determine the results of cancelling the

3.1 Methods

Participants. 210 participants were recruited either as volunteers from a community sample or as undergraduate students from the Carleton University undergraduate research pool. Volunteers received an invitation to the study shared via social media and did not receive any compensation for their participation. Undergraduate participants received partial course credit (0.25%) in an introductory Cognitive Science course. The study was approved by the Carleton University Research Ethics Board.

3.2 Design and Materials

I adapted Fox's game scenario paradigm, using a 2(Implicature Availability) X 2(Previous Outcome) X 2(Scalar Item) within-subjects design, where the participant saw a disjunction or a numeral that either licenses or cancels a scalar implicature. Participants saw four disjunctive sentences and four numerals[2]. Additionally, the participant was informed of the approximate likelihood associated with a particular answer based on the host's hint and the outcome (winning or losing) from a previous contestant. Response choices and response time data (in seconds) were collected. See Tables 1 and 2 for item design. Response time data was comprised of the time to fully read the scenario and make a decision about which alternative is more likely.

Table 1: Disjunction experimental item design

Your task is to choose a numbered box. There are 100 numbered boxes in total and five of them contain a million dollar prize. The host tells the first contestant that there is money in **box 20 or box 25/box 20 or box 25 both**. This contestant picks box 20 and **finds a million dollars there/ discovers that the box is empty**. Imagine you are the next contestant in this game.
The host does not give you any new hints. Which action are you most likely to take?
a.) Choose box 25.
b.) Choose another box.

Table 2: Numerals experimental item design

Your task is to choose a numbered door. There are eight numbered doors and four of them are associated with a million dollar prize. The host tells the first contestant that there is money associated with **one/at least one** door with a number less than 3. The contestant before you picks Door 1 and **wins a million dollars /does not win any money**. Imagine you are the next contestant in this game.
The host does not give you any new hints. Which action are you most likely to take?
a.) Choose Door 2.
b.) Choose another door.

[2] The numerical expressions were constrained to a choice between only two possibly relevant alternatives to be comparable to the disjunctive statements and to make the other alternative particularly salient.

Predictions. For the response times, the results may be consistent with the respective literature on numerals and disjunctions. Namely, the basic interpretation in disjunctions will be faster to process compared to its strengthened counterpart [18, 23]. For numerals, the opposite finding could arise, where the strengthened meaning would be more easily accessed to the basic counterpart [19, 15]. That said, empirical data is lacking on the time course of scalar implicature when conversational maxims do not apply. Similarly, they may be emergent patterns as a result of one of the alternatives being asserted or negated (e.g., faster responses times when the previous contestant won after picking a given box), although what they might be is unclear.

Under the pragmatic approach, there should be no difference between inclusive and exclusive disjunctions as scalar implicatures are wholly unavailable. Under the grammatical approach, the prediction is that there should be an interaction between the implicature (whether it is available or cancelled) and the previous outcome (winning or losing). Specifically, when the previous contestant did not win any money, participants ought to choose the specified alternative (i.e., box 25) regardless of whether the implicature is licensed or cancelled, based on the host's hint. However, when the previous contestant did win money, participants under the condition where the implicature is licensed (box 20 or box 25 ⇒ but not both) should not choose the alternative and instead choose another box. When the implicature is cancelled (box 20 or box 25 **or both**), participants should be more likely to choose the alternative given as a hint, as in this case having a million dollars in box 20 does not negate that there is also money in box 25. Table 3 outlines the experimental conditions and their predicted outcomes under the competing approaches.

Table 3: Predicted outcomes from pragmatic and grammatical approaches for disjunctions

Approach	Disjunction	Previous Outcome	Predicted Choices
Pragmatic	**box 20 or 25**	**won**	**another box $\approx$ box 25**
	box 20 or 25 or both	won	another box $\approx$ box 25
	box 20 or 25 or both	lost	box 25 $>$ another box
	box 20 or 25	lost	box 25 $>$ another box
Grammatical	**box or 25**	**won**	**another box $>$ box 25**
	box 20 or 25 or both	won	another box $\approx$ box 25
	box 20 or 25 or both	lost	box 25 $>$ another box
	box 20 or 25	lost	box 25 $>$ another box

For the condition in bold, the pragmatic and grammatical approaches differ in their predicted choices, due to the debate on the availability of the implicature. While only the predictions for the disjunctions are presented here, the same logic applies for numerals. Specifically, when there is money in one of two options (e.g., door 1 or door 2) and the previous contestant won, the pragmatic approach predicts that a specified alternative (e.g., door 2) and the choice for "another option" should be equally likely. The grammatical approach instead predicts a preference for another option over the specified alternative, due to the availability of the implicature *exactly* one.

3.3 Results

Descriptive Statistics. Extreme outliers were identified using the boxplot method and further corroborated using the interquartile range. Outliers were trimmed from the dataset before subsequent analysis. Table 4 denotes the mean proportions of responses for the alternative given in the experimental item (e.g., box 25) and the mean response times per experimental condition. Higher responses corresponded to more choices for the specified alternative and lower responses indicate more choices for "another option".

Table 4: Response proportions and response times (s) and standard errors for each experimental condition.

Item	Implicature	Prev. Outcome	Response Prop'n (SE)	RT (SE)
Disjunctions	Available	Won	0.22 (0.05)	32.1 (1.6)
	Available	Lost	0.82 (0.05)	27.4 (1.5)
	Unavailable	Lost	0.86 (0.04)	29.2 (1.5)
	Unavailable	Won	0.39 (0.06)	33.6 (2.0)
Numerals	Available	Won	0.26 (0.06)	31.0 (1.8)
	Available	Lost	0.86 (0.04)	28.5 (1.8)
	Unavailable	Lost	0.81 (0.05)	27.8 (1.4)
	Unavailable	Won	0.44 (0.07)	33.1 (1.7)

Response times. Response times were collected as the total time for the participant to read the context and make a response to the question of interest. Response times were transformed to a logarithmic scale to approximate normality and were subsequently analyzed using a linear mixed model with the "lmer" function from the "lme4" package in R [3]. Implicature availability and prior outcome were coded as fixed effects. Based on recommendations by [2], a maximal random effects structure was used, namely one that would did not fail to converge. The random structure adopted a per-participant random adjustment to the fixed outcome intercept (SD = 0.13, r = -0.21). Additionally, the fixed effect of outcome was included to the random slope term. Item as a random effect was not included in the structure as it had a low variance and would fail to converge or overfit the data when added to the formula. A likelihood ratio test determined that there was significant main effect of prior outcome ($\chi^2(1)$ = 16.06, $p < 0.0001$). If previous contestant won, participants were on average four seconds slower to respond (32.1s) than if the previous contestant lost (28.1s; β = -0.13, t = -4.04, $p < 0.0001$). The p-value was adjusted using the Tukey method to correct for multiple comparisons. There was no influence of the type of scalar item (disjunction or numeral) on the response time ($\chi^2(1)$ = 2.62, p = 0.11). Likewise, there was no main effect of implicature ($\chi^2(1)$ = 0.76, p = 0.38). Looking at the relationship between response times and response choices , while not significant, a linear regression suggests a general trend ($\beta = 0.22$, $t = 0.82$ $p = 0.41$)[3]. Participants were slightly faster to choose the given alternative (29.1s) compared to choosing another option (31.4s). This effect arose independently of the effects prior outcome, type of scalar item or the availability of the implicature.

Response Choices. Response choices for both disjunctions and numerals were analyzed using a logistic mixed model with the 'glmer" function from the "lme4" package in R [3]. Implicature availability, type of scalar implicature and previous outcome were coded as fixed effects. Similarly for the random effects structure used for the response times, the model used a per-participant random adjustment to the fixed outcome intercept and outcome was added as a fixed effect to the random slopes (SD = 1.73, r = -0.98). Overall, there was a significant two-way interaction between previous outcome and implicature availability, as determined by a likelihood ratio test ($\chi^2(1) = 5.93$, $p < 0.015$). When the previous contestant won and the implicature was unavailable (e.g., box 20 or 25 or both) participants were more likely to choose the specified alternative (box 25) than when the implicature was available (box 20 or 25, but not both; $\beta = 0.84$, z = 2.88, p = 0.021; Tukey adjusted). When the previous contestant did not win money, participants strongly preferred the specified alternative regardless of whether the implicature was available or blocked (β = -0.38, z = -0.93, p = 0.79; Tukey adjusted). Again, there was no influence of the type of scalar item (disjunction or numeral) on the response choices ($\chi^2(1) = 1.53$, $p = 0.22$). This is the pattern of results that was predicted by the grammatical approach. The pragmatic approach would have predicted a main effect of outcome with no interactions (two parallel lines with greater values when the previous contestant lost and smaller values when the previous contestant won).

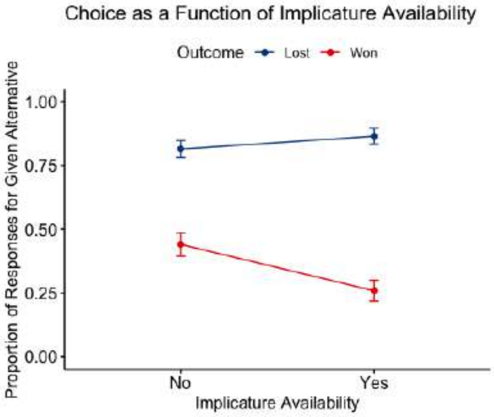


Fig. 1: Proportion of response choices as function of implicature availability and outcome. Error bars represent the standard errors of the mean.

[3] Random effects were not included in this analysis to avoid overfitting the data.

4 Conclusion

The results of this study are consistent with the predictions of [8] under the grammatical approach to scalar implicature, where, scalar implicatures remain available when Quantity is not active. This current study shows that when a scalar implicature is in principle available to a listener (e.g., box 20 or box 25) reveal that that individuals generate scalar implicatures to strengthen the meaning of a sentence. As a manipulation check, when the previous contestant lost, the predictions from either theoretical approach state that participants ought to choose the specified alternative, regardless of whether the implicature is available or not. Generally, participants chose responses consistent with this prediction, with higher proportions of choices for the specified alternative compared to the choice of another option.

Response times also support this, such that, participants were quicker to respond to an item when they are told that the participant previously lost than when they won. These results for response choice were not influenced by the type of scalar implicature (disjunction/numeral). There were no difference in response times for the specified alternative compared to another option. This result does not support the prediction that response times would be faster for those who did not generate the scalar implicature in disjunctions and faster for those who did generate them in numerals. This effect is partially attributable to the experimental design, as the response time included the time to read the scenario and then respond to the question. This metric is likely to not be precise enough to capture the respective processing mechanisms of disjunctions and numerals.

One limitation of the study is that it was administered as a within-subjects design, where participants saw conditions in which the implicature was available and in which the implicature was not. A potential consequence of this could be that participants were comparing the difference between the plain disjunction or numeral (e.g., box 20 or 25) with such items that cancelled the implicature (e.g., box 20 or 25 or both). This comparison may have prompted individuals to generate the scalar implicature for the plain scalar implicature independently of Quantity being cancelled. Thus a between-subjects design should be conducted to provide additional support for the current findings. Additionally, the use of Hurford's disjunctions to cancel the implicature presents a challenge to what can be expected as a prediction, particularly from the neo-Gricean view. Such that, the disjunction "box 20 or 25 or both" is presumed to be equivalent to "box 20 or 25", irrespective of cancelling the maxim of Quantity ([1] and see [16] for further discussion). While no such caveat exist for the numerals, this observation presents added considerations for the feasibility of the neo-Gricean views of scalar implicature outside of ideal contexts.

Another consideration is that, while this study primarily considers specific Gricean and neo-Gricean views as representative of the pragmatic approach, there are other pragmatic approaches that make use of a semantics where "A or B" is not equivalent to "A or B or both", for example, attentional pragmatics [24]. Predictions under these accounts should also be investigated. Additionally, the results do not explicitly confirm that participants are deriving the scalar implicature under grammatical exhaustification. As *exh* here is presumed to be optionally applied [5, 4], there is no specification as to how it is generated when Quantity is cancelled and other non-cooperative contexts.

More broadly, this investigation points to further required insights into the pragmatic conventions and expectations that may persist and arise in non-cooperative contexts. The present study and others (e.g., Dulcinati et al., [6]) demonstrate one such phenomenon, where scalar implicatures can arise in contexts were the speaker is non-cooperative or deliberately un-cooperative (in competition with a listener).

5 Acknowledgments

I would like to express my immense gratitude to my supervisor, Dr. Raj Singh, for his fruitful conversations and astute comments that greatly contributed in developing this work. I would also like thank the ESSLLI reviewers whose detailed feedback helped to improve the quality of the paper. All remaining errors are my own.

References

1. Alonso-Ovalle, L.: Disjunction in Alternative Semantics. Ph.D. thesis, University of Massachusetst, Amherst (2006)
2. Barr, D.J., Levy, R., Scheepers, C., Tily, H.J.: Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language **68**(3), 255–278 (2013)
3. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. Journal of Statistical Software **67**(1), 1–48 (2015). https://doi.org/10.18637/jss.v067.i01
4. Chierchia, G., Fox, D., Spector, B.: The grammatical view of scalar implicatures and the relationship between semantics and pragmatics (2008)
5. Chierchia, G., et al.: Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. Structures and beyond **3**, 39–103 (2004)
6. Dulcinati, G., Pouscoulous, N.: Cooperation and exhaustification. Pre-proceedings of Trends in Experimental Pragmatics pp. 39–45 (2016)
7. Fox, D.: Free choice and the theory of scalar implicatures. In: Presupposition and implicature in compositional semantics, pp. 71–120. Springer (2007)
8. Fox, D.: Cancelling the Maxim of Quantity: Another challenge for a Gricean theory of Scalar Implicatures. Semantics and Pragmatics **7**, 1–20 (2014)
9. Grice, H.P.: Studies in the Way of Words. Harvard University Press (1989)
10. Grice, H.P.: Logic and conversation. In: Speech acts, pp. 41–58. Brill (1975)
11. Groenendijk, J.A.G., Stokhof, M.J.B.: Studies on the Semantics of Questions and the Pragmatics of Answers. Ph.D. thesis, Univ. Amsterdam (1984)
12. Horn, L.R.: On The Semantic Properties Of Logical Operators in English. Ph.D. thesis, University of California (1972)
13. Horn, L.R., Ward, G.L.: The Handbook of Pragmatics. Wiley Online Library (2004)

14. Levinson, S.C.: Presumptive meanings: The theory of generalized conversational implicature. MIT press (2000)
15. Marty, P., Chemla, E., Spector, B.: Interpreting numerals and scalar items under memory load. Lingua **133**, 152–163 (2013)
16. Meyer, M.C.: Deriving hurford's constraint. In: Semantics and linguistic theory. vol. 24, pp. 577–596 (2014)
17. Meyer, M.C., et al.: Ignorance and grammar. Ph.D. thesis, Massachusetts Institute of Technology (2013)
18. Noveck, I.A.: When children are more logical than adults: Experimental investigations of scalar implicature. Cognition **78**(2), 165–188 (2001)
19. Papafragou, A., Musolino, J.: Scalar implicatures: experiments at the semantics–pragmatics interface. Cognition **86**(3), 253–282 (2003)
20. van Rooy, R.: Conversational implicatures and communication theory. In: Current and new directions in discourse and dialogue, pp. 283–303. Springer (2003)
21. Sauerland, U.: Scalar implicatures in complex sentences. Linguistics and philosophy **27**(3), 367–391 (2004)
22. Sperber, D., Wilson, D.: Relevance: Communication and cognition, vol. 142. Harvard University Press Cambridge, MA (1986)
23. Tieu, L., Romoli, J., Zhou, P., Crain, S.: Children's knowledge of free choice inferences and scalar implicatures. Journal of Semantics **33**(2), 269–298 (2015)
24. Westera, M.: An attention-based explanation for some exhaustivity operators. In: Proceedings of Sinn und Bedeutung. vol. 21, pp. 1307–1324 (2018)
25. Westera, M.: Hurford disjunctions: an in-depth comparison of the grammatical and the pragmatic approach. Under review (2020)

More Truthmakers for Vagueness

Shimpei Endo[1]

Hitotsubashi University, Kunitachi, Tokyo, Japan
endoshimpeiendo@gmail.com
endoshimpeiendo.github.io

Abstract. Towards the sorites paradox, many theorests have suggested their own solutions. Among these many, Sorensen suggests his version of epistecism solution, *truthmaker gap epistecism*, according to which the sorties paradox arises because there are unground truths – true with no truthmaker. Nevertheless, Sorensen's version has been criticized for abandoning *higher-order* vagueness. This paper suggests a novel version of truthmaker theory for vague predicates, which not only solves the sorites paradox but also encompasses higher-order vagueness.

Keywords: truthmaker · vagueness · the Sorites paradox · higher-order vagueness

1 Introduction

When vagueness matters, truth matters a lot. A well known paradox of the sorites has cast a question on the concept of truth. For example, a solution by degree theorists [1] to the sorites paradox questions our widely-believed presupposition on truth by suggesting that truth values do not have to be just false (truth value 0) nor just true (truth value 1) but somewhere between them such as 0.2 or 0.9876. Given that truth matters when we talk about vagueness, *truthmakers* may matter as well. At least, Sorensen [5] employs the idea of truthmakers to present a variant of the epistemicist solution (e.g. [7]) to the sorites paradox. According to his account labeled *truthmaker gap epistemicism*, there are sentences which are true but *ungrounded* i.e. has *no* truthmaker. In the sorites case, an epistemicist of this type insists on the existence of some natural number k such that a sentence (K) "a person with k hairs is bald" is true without any truthmaker. To solve the sorites paradox, we need to explain how someone turns into non-bald from bald. In terms of truthmaking, we need to explain what makes (K) true and (K+1) "a person with $k+1$ hairs is bald" false. Sorensen's answer is *nothing*. Such a natural number k exists but (K) has no truthmaker at all. Sorensen holds a version of epistemicism, claiming that we do not know the exact value of k because (K) has nothing that makes it true.

However, Sorensen's solution calls for an unwanted byproduct. [2] As Jago [2] has already pointed out, Sorensen's approach is incompatible with higher-order

[1] For example, see [3].

[2] Another important issue, which is rather metaphysical, is about *truthmaker maximalism*. Most truthmaker theorists assume truthmaker maximalism, that is, every

vagueness: vagueness about whether something is a borderline case, vagueness on whether something is a borderline case of being a borderline case and so on. Sorensen's framework cannot explain this higher-order vagueness because his account of truthmaking is understood as clear-cut: either truth is *wholly* made true or truth is not made *at all*. Sorensen allows borderline cases in the first order but no higher-order borderline cases. This is problematic because we have many instances of higher-order vagueness. We – even Sorensen himself [6] [4] – often say that "vague is vague". We can reasonably discuss a borderline case of a borderline case, and a borderline case of a borderline case of a borderline case, and so on. Any reasonable theory of vagueness should explain this phenomenon.

This paper suggests a more fine-grained account to the sorites paradox, adopting several theoretical resources from truthmaker theories (*cf.* [1]). My approach is inspired by Sorenson's idea: truth-makers play a significant role in sorites scenarios. My proposition is richer and more fine-grained which better captures higher-order vagueness. To reconcile higher-order vagueness, I posit considerations that are neither subversive nor challenging, rather, I wish to utilize vocabulary with which we have already been familiar with in the ongoing truthmaking community. Instead of the black-and-white notion of "lack" and "existence" of truthmakers as Sorensen employed, we can utilize the richer toolkits of truthmakers such as distinction between *partial* and *full* truthmakers.

The rest of this paper is constructed as follows. The next sections explains an updated version of truthmaker theory for vague predicates (§2) and demonstrates this solves the sorites paradox (§3). The following section (§4) examines how the new one can deal with higher-order vagueness, which Sorensne fails.

2 More truthmakers and truthmakings

This section offers my version of truthmaker theory for the sorites paradox. Following my theory, it turns out that the sorites paradox arises only when you believe two unreasonable assumptions on certain facts (about the number of hairs). Unless we adopt these conditions, we avoid the counterintuitive conclusion such as "a person with 2,000,000 hairs is bald".

Notation. Before diving into the discussion, let me note several notations. Let B be a 1-ary vague predicate "is bald". Write a natural number n indicating a person with n hairs. Hence, read $B(n)$ as "a person with n hairs is bald".

Let $f, g, h, ...$ be *facts*. If a single fact f fully or partially makes a proposition ϕ true, then f is a (full or partial) truthmaker for ϕ. For example, the fact that I am being a student is the full truthmaker for a proposition "I am a student". Similarly, the fact that I am being a student is a partial truthmaker for a proposition "I am a student in Tokyo" since we need other facts to make it true (like the fact that I am living in Tokyo).

truth has its truthmaker, as a default setting. Sorense does not seem to provide enough reasons to reject this default principle. I will leave this issue for another project.

Let Γ be a set of facts $\Gamma = \{f, g, h, ...\}$. A fact can be about anything. We introduce *types* of facts. In particular, let f^n be a fact that the person in question has n hairs. Facts other than this type such as facts about allocations of hairs, facts that a person puts wig are written in different alphabets. Call a collection of facts (truthmakers) a *truthmaking*, highliting the difference between each fact and a collection of facts. [3] Read $\Gamma \vdash B(n)$ "a truthmaking Γ makes $B(n)$ true". We often drop brackets of a truthmaking for the sake of simplicity. For example, write $f, g \vdash \phi$ instead of $\{f, g\} \vdash \phi$.

The sorites paradox reformalized. Now, let us write the sorites paradox in the truthmaking terminology. (Seemingly) reasonable assumptions (1,2) lead to an unintuitive conclusion (3).

1. Base case. $f^0 \vdash B(0)$. *A person with 0 hair (no hair at all) is bald.*

2*. Tolerance principle. There is no $n \in N$ such that $f^n \vdash B(n)$ and $f^{n+1} \vdash \neg B(n+1)$. *A small change such as pulling a single hair does not make a person bald from non-bald.*

3. Conclusion. $f^{2,000,000} \vdash B(2,000,000)$. *A person with 2,000,000 hair is bald.*

The spirit of tolerance principle, a key assumption of the sorites paradox, is "small change does not matter". Intuitively, pulling one single hair seems to change nothing about whether somebody is bald or not bald. Its equivalent claim also sounds reasonable as well: We do not have the clear threshold between baldness and non-baldness. If we do, we have the exact number k – which is the largest number of hairs for being bald and $k+1$ is the smallest number for being non-bald. Most non-truthmaker theories formulate the tolerrance in a normal if-then clause such as: if $B(n)$, then $B(n+1)$ for an arbitrary n. Or, equivalently, non-existence claim such as there is no n satisfying both $B(n)$ and $\neg B(n+1)$.

From the truthmaker perspective, we can understand the sorites scenario in greater detail. First, the base case is now understood in the terms of truthmaking as "the fact of her having no hair at all is sufficient to make the person bald". With no help of other facts, it is enough to determine that the person is bald just by the fact about her number of hair, which is, zero. Second, this truthmaking talk gives a closer view to the tolerance principle –the key clause of this paradox. We should understand the principle as follows: there is no $n \in N$ such that $f^n \vdash B(n)$ and $f^{n+1} \vdash \neg B(n+1)$. This principle translated into the truthmaker terminology is read as: you cannot find a pair of facts f^n "having n hairs" which makes someone bald and f^{n+1} "having $n+1$ hairs" which makes someone not bald. [4]

[3] You can reasonably claim that truthmaking – a collection of truthmakers – is also a (kind of) truthmaker. But the metaphysical debate on this does not affect our discussion on the sorites in this paper.

[4] Note this is weaker than the following version.

2!. The tolerance principle (stronger). If $f^n \vdash B(n)$, then $f^{n+1} \vdash B(n+1)$.

This stronger version claims that if the fact f^n of having n hairs makes someone bald, then another fact f^{n+1} of having $n+1$ hairs also makes someone bald.

I am not claiming that such $B(k+1)$ is indeterminate. f^{K+1} is not enough but it may have its truthmaking. With extra facts, say Δ, $B(k+1)$ (or its negation) would

The paradox avoided. Following this formalization, the sorites paradox does not necessarily happen. Let us assume the base case **1**. From the truthmaking perspective, this says more than a person with no hair at all is bald; it claims that the single fact f^0 about the number of hair is enough to make the person bald. Let us also assume the tolerance **2***. Notice that the tolerance here only claims that there is no pair of a single fact f^n about the number of hair n which makes $B(n)$ true and another single fact f^{n+1} about $n+1$ hairs which make $\neg B(n)$ true. In other words, the truthmaking of a single fact f^k about the number of hair k for $B(k)$ does *not* promise another fact f^{k+1} about $k+1$ hairs does the same job for $B(k+1)$. Note that I do not mean that such k is a threshold between baldness and non-baldness. $B(k+1)$ may be still true with other truthmakers (like $\{f^k, g, h\}$) or by other truthmakings (like $\{g, h, i\}$). In such cases, the tolerance does not block another pair of two facts, say, g^m which makes $B(m)$ true and g^{m+1} which makes $\neg B(m+1)$ true because the tolerance is only about facts about the number of hair.

Conditions needed. Now, we avoid the sorites paradox by offering my trutuhmaker picture. But this is not the end of the story. We also need to explain why this is a paradox – why people find the sorites paradox paradoxical at all. For this purpose, we name the two extra conditions over truthmakers behind the paradox. Namely, the facts about the number of hairs should be (i) *full* and (ii) *obligatory* trutuhmakers. In other words, the sorites paradox arises because people (mistakingly) adopt these questionable and unjustified assumptions saying the number of hair is (i) sufficient and (ii) necessary for someone to be bald.

Condition 1: full. Firstly, to reach the problmatic conclusion, the truthmakers f^n – facts about the number of hairs – need to be *full* with respect to the baldness predicate B. By f^n being full with respect to $B(n)$, I mean that either f^n suffices to make $B(n)$ true or f^n suffices to make $\neg B(n)$ by itself and no other fact $g \neq f$ is needed. That is, the number of hairs sufficiently determines the truth value of $B(n)$.

What happens if we drop this fullness character? The sorites paradox does not pop up! Let us see in detail. Assume the base case **1**. The fact of having no hair f^0 is enough to make a person with no hair at all bald. Formally written, $f^0 \vdash B(0)$. However, our tolerance **2*** does not block us to move forward to the following case: $f^n \vdash B(n)$ and $f^{n+1} \nvdash B(n+1)$ and $f^{n+1} \nvdash \neg B(n+1)$. This says that when the number of hair is n, the fact about the number of hair is enough to make true $B(n)$. But as for $n+1$ and $B(n+1)$, f^{n+1} is not enough. We need extra facts about the person with $n+1$ hairs, say, g, to make $B(n+1)$ true: $f^{n+1} \nvdash B(n+1)$ but $f^{n+1}, g \vdash B(n+1)$. It is easy to see that the problematic consequence ("a person with 2,000,000 hairs is bald") never arises after such n. The fact that a person has $2,000,000$ hairs is not sufficient to make $B(2,000,000)$ true.

be determinated as $\Delta, f^{k+1} \vdash B(n)$ (or $\Delta, f^{k+1} \vdash \neg B(n)$). There may not be such supporting Δ. This would be Sorensen's picture of absolute borderline case. I will discuss this later.

Condition 2: obligatoriness. We have just observed that the sorites paradox presupposes that the number of hair should be a *full* trutuhmaker for baldness. The other required condition is *obligatoriness*: the number of hair should be *obligatory* in the sense that the number of hair appears in any truthmaking Γ for $B(n)$. In other words, the number of hairs *always* matters. More formally speaking, f^n is obligatory for a predicate B if $\Gamma \vdash B(n)$ implies $f^n \in \Gamma$.

Let us check that the sorites paradox needs to presuppose this condition. For the sake of the purpose of this current argument (to highlight the necessity to reach the sorites paradox), we assume that the number of hairs is full for baldness. Let us start with the base case $f^0 \vdash B(0)$. And suppose $f^n \vdash B(n)$ for some n. Here let us drop the obligatoriness. Then, we can have another fact $g \neq f$ such that $g \vdash B(n)$. Now we have two independent truthmakings for the same proposition $B(n)$: $f^n \vdash B(n)$ and $g \vdash B(n)$. The tolerance **2*** states only about the facts about the number of hairs. So it has nothing to do with facts with another type g, h, i or anything other than f, which is not about the number of hairs. Hence, some non-f fact may make $\neg B(2,000,000)$ true — $g_{\neq f} \vdash \neg B(2,000,000)$. Or, at least, may not make $B(2,000,000)$ true — $g_{\neq f} \not\vdash B(2,000,000)$

3 Arguments against idealized properties

In the previous section, we specified two conditions over truthmakers required to make the sorites scenario paradoxical. This section presents less-formal arguments claiming that these two conditions are too ideal in most cases. Our starting point (a person with no hair at all is bald because she has no hair at all) and ending point (a person with 2,000,000 hairs is not bald because she has 2,000,000 hairs) are just exceptional cases where the numbers of hairs solely matter.

3.1 Fullness?

Let us start with the condition to be full i.e. the fact about the number of hair is sufficient to determine truth value of the bald claim. This principle is not always the case for baldness and other vague predicates [5] because we often rely on facts other than the number of hairs when we evaluate the predicate. In fact, the number of hair alone does not tell us enough information about the

[5] For example, consider "is tall". In some cases, the number of centimeters (or any units of heights) is enough to tell whether the person is tall or not. 230 cm is tall. We can say so without any fact other than the height in centimeters. However, in many cases, the number in centimeters does not determine whether the person is tall or not. Say, 173 cm. S/he may be tall for the first grader. But it is not sure when she is in the middle age. We need other facts to consider. A similar arguement goes for "is a heap" and the number of grains. In some apparent cases like zero, the number of grains is enough for us to tell a heap from a non-heap. But for many numbers, we may need information on other facts such as arrangement or form of sand grains.

baldness. Suppose you are given the exact number of someone's hair, say, $5,302$. Do you have a clear idea whether s/he is bald or not? Maybe, in some exceptional apparent cases (e.g. where the person has 0 hair or 10^{80} hairs), the exact number of hairs by itself provides enough information to determine the baldness. But it is highly questionable to think that the number of hairs is always sufficient to know whether someone is bald.

3.2 Obligatoriness?

Canceling the obligatory condition means that things independent of the number of hairs would make the baldness claim true. Someone is bald due to things which have little to do with the number of hairs.

This is possible and plausible in many cases. First of all, we seem to make a clear and unquestionable evaluation on baldness predicate without knowing the number of hairs. My father is certainly not bald. But I do not know the exact or even approximate number of his hair. My grandfather was certainly bald. But I do not know the exact or even approximate number of his hair. You have a clear evaluation of whether someone is bald or not. But when you have such a clear vision, do you know the exact value of the number of hairs? The number of hairs may play an important role in many cases. Not only ordinary people but professional philosophers have assumed, mistakenly, the number of hair is a *necessary* component of what makes baldness predicates true or false. Facts which make the baldness true include allocation, length, compositions, and so on. And sometimes, such facts, which are independent of the number of hairs, make a person bald or non-bald. I am *not* claiming here that the number of hair *never* matters when we evaluate someone is bald. Rather, I claim that it is compatible that the fact about the number of hairs f^n makes $B(n)$ true and another fact g independent of f^n makes $B(n)$ true at the same time.

4 Higher-order vagueness

Now, the sorites has been solved in the truthmaking terms: the sorites paradox arises only when we assume the two unreasonable assumptions over truthmakers. My remedy to the paradox is simple. To overcome the paradox, just cancel one of these unjustified assumptions. However, any solution has its competitors. In fact, Sorensen also solves the paradox in his more simpler way. So we need to show what points my solution is better than others (at least Sorensen's).

To this end, this section argues that importing different types of truthmakers and truthmaking gives enough space for *higher-order vagueness*, which Sorensen's system fails to capture. Instead of the number of hairs, I employ the *number of truthmakers (facts) which ground*. Some sentences just need a single truthmaker to determine its truth value. Others may need two. Still others need more. The main idea here is that the number of truthmakers corresponds to how higher-order its vagueness is. The more facts needed to support, the more vague it is. Here are some quick examples for this idea. "is unemployed" seems less vague than “is poor”. Why? According to my truthmaker oriented explanation, it is because while the former needs only single type of facts (just check whether she is unemployed or not) the latter needs more types of facts such as posessions of properties, the amount of spending, prices of commodities, and so on. “is good” is another instance. Being good is, without question, not only ambiguous but vague. My view tells why so – because it needs to consider many different types of facts about the person to determine whether she is good.

This line of thought offers us some space for higher-order vagueness in the sorites scenario — higher-order borderline cases. Even if we consider only the single type of facts (in the sorites scenario, the number of hairs), as we have already seen, we face many instances of the first-order vague i.e. first-order borderline cases like 72 hairs or 5789 hairs. Taking another type of facts (like allocation of hairs) into consideration dissolves some borderline cases into non-borderline cases. Others still remain borderline cases. These remaining ones are now called the higher(second)-order borderline cases. Taking consideration to another type of facts, we eliminate second-order borderline cases and face third-order borderline cases.

Borderlines at first glance – first-order borderlines. As we have already observed, in some instances of the sorites series, the number of hair f^n is insufficient to determine $B(n)$ or $\neg B(n)$ for some n. When we consider only these factors about the numbers of hairs, we will face many indeterminate cases. For example, consider a person with j hairs such that $f^j \not\vdash B(j)$ and $f^j \not\vdash \neg B(j)$. Considering another fact with another type g which is about other than the number of hairs (such as the area covered by hair?) may desolve the indeterminacy. For some k, it may be indeterminate whether $B(k)$ or $\neg B(k)$ by considering only facts f^k about the number of hair. But considering another type of facts, say g such as the allocation of hair, may fix the truth value. $f^k, g^k \vdash B(k)$. Then, this case k is vague cases in a *higher-order* than n.

Further borderlines. We expand this idea above – a *single* truthmaker is not suffice to make true but two truthmakers, if rightly selected, are suffcie – for defining the order of borderline cases. Let us see the first-order cases. When you cannot determine $B(n)$ or $\neg B(n)$ by considering *any* single fact, $B(n)$ is said to be a first-order vague instances (with respect to baldness predicate B). More formally written, n is a first-order borderline case with respect to B if $f \not\vdash B(n)$ and $f \not\vdash \neg B(n)$ for any fact f.

The idea of higher-order vagueness is naturally given by putting more facts (truthmakers) under consideration. For the second-order borderline cases, define as follows: n is a second-order borderline case if $\Gamma \not\vdash B(n)$ and $\Gamma \not\vdash \neg B(n)$ for any Γ such that $|\Gamma| = 2$.

More generally, the m-th order borderline case is defined case as follows: n is a m-th order borderline case if $\Gamma \not\vdash B(n)$ and $\Gamma \not\vdash \neg B(n)$ for any Γ such that $|\Gamma| = m$.

We can capture Sorensen’s version of truthmaker gap epistemicism as a tiny subset of my grand theory. My truthmaking formulation can capture what

Sorensen mean by *absolute borderline cases*. Absolute vagueness is written as the limit of this scheme of adding further facts to determine $B(n)$ or $\neg B(n)$. Formally written, n is an *absolute* borderline case (i.e. $B(n)$ is absolutely vague) if $\Gamma \nvdash B(n)$ and $\Gamma \nvdash \neg B(n)$ for *any* Γ (hence any $|\Gamma|$). In other words, an absolute borderline case is where no collection of truthmakers makes true $B(n)$ nor $\neg B(n)$. My formulation and this paper leaves it open the existence of such a limit point. But my system is capable of describing Sorensen's opinion comparing it with others.

Higher-order apparent case! My framework can do even more than the higher-order of vague cases. We can account for higher-order *apparent* cases, which enable us to depict the more fine-grained graduation between apparent cases (0 hair and 2,000,000 hairs). When you think of borderline cases, you may also think of apparent cases. Suppose $B(567)$ is true, but not for completely sure. $B(15)$ is true more certainly than $B(567)$. But $B(2)$ is true further more seemingly.

To give grades of being apparent, we can use the formal idea of truthmakers. Our formal setup allows us to take two (or more) different collections of truthmakers to make the same statement true. How apparent the claim is corresponds to the number of the collections of truthmakers which make the claim true. Put more formally, n is a first-order apparent case for B if there is only one Γ such that $\Gamma \vdash B(n)$. Naturally, higher-order apparent case is defined: n is a mth-order apparent case for B if there are exact m $\Gamma_1, ..., \Gamma_m$ such that $\Gamma_1 \vdash B(n) ... \Gamma_m \vdash B(n)$.

For example, let $B(2034)$ have 5 different collections of truthmakers $\Gamma_1, ...\Gamma_5$. Let, also $B(2)$ have more truthmakers, say, 23 of them – $\Delta_1, ..., \Delta_{23}$. Given that, we can describe how apparent $B(2034)$ and $B(2)$ and we can say the latter is more apparent than the former in terms of the number of collections of truthmakers.

Together with the higher-order borderline cases, we can describe the sorites paradox in the following reasonable way. The paradox begins with assuming that the fact of having no hair at all f^0 is enough for determining a person is bald $B(0)$. In this sense, this case with zero hair is not vague at all (given you consider the number of hairs). But also, many other facts such as allocation, how it looks and so on, make them true as well. $B(0)$ is a many-order apparent case; So no one is against this assumption of the paradox. The following case $B(1)$ may be similar to $B(0)$ with respect to not only its truth but also its apparency. But adding further number, it becomes less and less apparent – less collections of truthmakers are available.

A key observation is that being apparent and being vague are independent from each other. Apparent cases and borderline cases can overlap. For example, consider $B(k)$ such that $f, g \vdash B(k)$ and $h, i \vdash B(k)$ and no other Γ make it true. $B(k)$ is an apparent case because $B(k)$ has two different collections of truthmakers. But also $B(k)$ is (the first-order) vague in a sense that it requires more than one facts to support. This feature fits well the epistemicists' project because this overlaps explain why we often overlook thresholds in any order. We fail to find thresholds because such cases can be not only vague but also apparent.

5 Conclusion

Sorensen's importing truthmakers in the long-disputed topic of vagueness and the sorites paradox is a small but great step still in advancing our understanding of the nature of truth. However, unfortunately, his move sacrifices higher-order vagueness, which epistemicists have been very good at dealing with. I offer a more fine-grained interpretation of the sorites scenario using several common notions of truthmakers. I point out that the fact of the number of hairs does not necessarily play the sole and decisive role in truthmaking vague predicates. The fact about the number of hairs may need further extra facts to make the baldness claim true. Moreover, facts independent of the number of hairs may make the baldness claim true. The sorites paradox appears paradoxical because we often take for granted two conditions over truthmakers, which are questionable in many scenarios.

As a bonus, this richer framework of mine covers the higher-order vagueness. The last section demonstrates how my truthmaker formalization can talk about higher-order vagueness in an unified manner. This paper should not be perceived as a negative critique of Sorensen's work. His approach – and his goal to capture absolute vagueness – was also explained in a reasonable way under my scheme.

References

1. Fine, K.: Truthmaker semantics. In: Hale, B., Wright, C., Miller, A. (eds.) A Companion to the Philosophy of Language, pp. 556 – 577. John Wiley & Sons Ltd., 2 edn. https://doi.org/10.1002/9781118972090.ch22
2. Jago, M.: The problem with truthmaker-gap epistemicism **1**(4), 320–329. https://doi.org/10.1002/tht3.49
3. Smith, N.J.: Vagueness and Degrees of Truth. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199233007.001.0001, ISSN: 0004-8402
4. Sorensen, R.: Borderline hermaphrodites: Higher-order vagueness by example **119**(474), 393–408. https://doi.org/10.1093/mind/fzq030, `https://academic.oup.com/mind/article-lookup/doi/10.1093/mind/fzq030`
5. Sorensen, R.: Vagueness and Contradiction. Oxford University Press
6. Sorensen, R.A.: An argument for the vagueness of 'vague' **45**(3), 134. https://doi.org/10.2307/3327138, `https://www.jstor.org/stable/3327138?origin=crossref`
7. Williamson, T.: Vagueness. Routledge

Towards a cross-linguistic description of morphological causatives: issues in syntax-semantics linking

Valeria Generalova

Heinrich Heine University of Dusseldorf

1 Introduction

Causatives have been in the focus of linguistic attention for decades. It seems that during this time, many papers written in the perspective of formal semantics have been using English lexical causatives as their object (e. g. the classical work Talmy 1976). Causative constructions of other types (morphological, periphrastic) seem to have been studied mostly in languages with a long well established tradition of linguistic research: English (Levin 1993), French (Reed 1991), Japanese (Manning, Sag, and Iida 1999), Korean (Shibatani 1973) and others. However, some well-known volumes offer quite a range of studies dedicated to syntactic and semantic properties of causative constructions in individual languages (Kholodovich 1969; Shibatani 1976; Comrie and Polinsky 1993; Dixon and Aikhenvald 2000).

One of our goals is to fill the gap in bringing together formalized syntactic and semantic analyses of productive morphological causatives in typologically varied languages. Being part of a larger project[1] about valence-increasing devices, this paper suggests a method of formal modeling of constructions with morphological causatives based on transitive verbs in three languages: Nivkh (Isolate, Sakhalin island), Bashkir (Turkic, Ural region), Halkomelem (Salish, Southwestern British Columbia). We focus on constructions based on transitive verbs as we are interested in various ways of introducing an additional argument to the verb structure. We find three-argument constructions particularly challenging in this respect. The selection of languages is not supposed to be representative or typologically salient. These languages were observed to contain interesting phenomena relevant for creating comprehensive models, which was the ultimate criterion for selection.

This paper aims to present an analysis robust enough to apply to different languages, formal enough to be used for computational tasks (such as parsing and sentence generation), easily extendable and in line with modern linguistic theory.

After this ambitious statement, some disclaimers would not go amiss. Firstly, this paper manifests only a phase of a larger project. Therefore, its outcome is more of a proof of concept rather than a finished theory. Secondly, for the sake of brevity and clarity, the implementation part of the presented analysis is only briefly sketched. Thirdly, the principal focus of the paper is the development of the formal approach.

[1] This work is part of the TreeGraSP project funded by the ERC Consolidator Grant awarded to Prof. Laura Kallmeyer.

Typological data and reasoning are used insofar that it is necessary for creating an accurate model.

In Section 2, we report the necessary claims from prior work without any novel data about the languages in question. In Section 3, key concepts are explained, and the most important methods and theories are introduced. Our research's novelty is in conceptualizing the way of analyzing this data, to what Section 4 is dedicated. Conclusions are drawn, and further research paths are indicated in Section 5.

2 Data

This study can be considered pilot insofar that it accounts for a very small number of languages – only three. However, its primary aim is not to suggest a complete account of a specific construction, but to present an approach towards a formal analysis and prove its usefulness.

In this section, we describe[2] the three selected languages: Nivkh, Bashkir, and Halkomelem. The first two languages share many properties: SOV word order, accusative alignment, agglutinative morphology, dependent marking. However, Bashkir has several constructions with morphological causative in contrast to Nivkh. Halkomelem is very different in many respects and is included in this sample to show that our model can capture even these differences.

2.1 Nivkh

Causatives are formed from transitive bases by means of the suffix *-gu-*. The argument realization strategy is described by Nedyalkov, Otaina, and Kholodovich 1969, p. 195. The new subject is placed in front of all the other participants. It receives an unmarked case, which is glossed as NOM in Gruzdeva 1998. The subject of the base sentence becomes an object of the causative sentence and receives the *-aχ-* morpheme if animate (which is most often the case). All other participants are left morphologically and syntactically unchanged. Examples[3] below demonstrate that the same construction can have factual (1) or permissive (2) semantics (or, at least, translations).

(1) *ytyk p'-oγla-aχ pitγy ama-gu-d*
father POSS-child-DAT/ACC book see-CAUS-FIN
'The father showed his son the book.' (Nedyalkov, Otaina, and Kholodovich 1969, p. 192)

(2) *ytik n'-aχ mos amla-gu-d*
grandmother 1SG-DAT/ACC berry.pudding taste-CAUS-FIN
'The grandmother let me eat the berry pudding.'
(Nedyalkov, Otaina, and Kholodovich 1969, p. 192)

[2] Abbreviations: 1 = first person, 2 = second person, 3 = third person, A = actor, ABL = ablative, ACC = accusative, CAUS = causative, DAT = dative, DET = determiner, DIR = direct, FIN = finite, FUT = future, IPFV = imperfective, NMR = non-macrorole participant, NOM = nominative, OBL = oblique, POSS = possessive, SBJ = subject, SG = singular, TR = transitive, UG = undergoer, UNM = unmarked.

[3] For Nivkh, we follow Gruzdeva 1998 in transliteration and most of glossing conventions.

For our paper, we have chosen Nivkh to illustrate the simplest case of a causative construction without additional complications.

2.2 Bashkir

Bashkir has two main strategies of case marking in causative constructions. Consider examples from Perekhval'skaya 2017: the default strategy requires ablative CAUSEE marking (3a), in the other strategy dative is used (3b). The difference in case marking does not come from the verb's lexical properties but is determined by the causative construction. This can be proved by the fact that both strategies can be acceptable with the same verb stem, as demonstrated by (3).

(3) a. *Babaj ul-ə-nan xat-tə uqə-t-tər-a*
old.man son-POSS.3-ABL letter-ACC read-CAUS-CAUS-IPFV
'The old man asks his son to read the letter.' (Perekhval'skaya 2017, p. 244)

b. *Babaj ul-ə-na xat-tə uqə-t-tər-a*
old.man son-POSS.3-DAT letter-ACC read-CAUS-CAUS-IPFV
'The old man lets his son to read the letter.' (Perekhval'skaya 2017, p. 244)

The relationship between these two strategies is not easy to describe. In this paper, we would like to model only one difference between the two constructions: the ABL marking is used when the pragmatic focus is on the THEME, while with the DAT marking it is on the CAUSEE. Consider comments by Perekhval'skaya 2017, pp. 244–245: "Similar for the situation in (3a), when the old man discovers the contents of the letter by means of the son (e. g. he comes up with this solution because he has forgotten his glasses). The letter is most probably read aloud in the presence of the old man. In (3b), the son is interested in discovering the contents of the letter. He may read it on his own, in a place or time different from the situation of causation." This observation is supported by data from other languages, namely, Kalmyk (Say 2009, p. 407).

2.3 Halkomelem

Causative constructions with transitive verb bases have been believed non-existent because the causative suffix cannot be stacked on the transitive suffix in Halkomelem (D. B. Gerdts and Hukari 2006b, p. 137, footnote 7). However, some verb roots can take a causative suffix instead of a transitive and form three-argument constructions. Consider the verb root *mək̓ʷ* in a two-argument construction with a transitive suffix in (4), and the same root with a causative suffix in a three-argument construction in (5).

(4) *mək̓ʷ-ət č ceʔ tᶿə syaɬ*
pick.up-TR 2.SBJ FUT DET firewood
'You will gather firewood.' (D. B. Gerdts and Hukari 2006b, p. 137)

(5)

neḿ	*cən*	*mək̓ʷ-stəxʷ*	*tᶿə*	*sƛiʔƛqəɬ*	*ʔə*	*tᶿə*	*q̓əyeḿan,*	*neḿ*	*ʔə*	*k̓ʷaƛkʷa*	*cəwmən*
go	1SBJ	pick.up-CAUS	DET	child	OBL	DET	shell	OBL	DET	salt.water	seashore

‘I’m going to get the boy to pick up shells by the seashore.’ (D. B. Gerdts and Hukari 2006b, p. 138)

Currently, we are not going to analyze this class of Halkomelem words deeply. Nevertheless, we are interested in accounting for this construction in our formal model (see Section 4.

Halkomelem data can also illustrate another interesting phenomenon. Sometimes, additional nuances of causative meaning are present in a construction without specific marking on either of the components. Namely, D. B. Gerdts and Hukari 2006b, p. 143 suggest a translation of (6) using the English verb ‘show’ despite the lack of an overt verb ‘show’ or ‘see’ in the original sentence.

(6)

neḿ	*ʔaɬ́-stəxʷ*	*tᶿə*	*swiẃləs*	*ʔə*	*tᶿə*	*təx̌ʷaʔc*	*!*
go	stretch-CAUS	DET	young.man	OBL	DET	bow	!

‘Go show the young man how to pull the bow!’ (D. B. Gerdts and Hukari 2006b, p. 143)

Note: in causative constructions like (5) and (6), the CAUSER is always the subject, the CAUSEE is the direct object marked with DET only and the THEME is the oblique object and thus preceded by OBL.

Even though English translations are not sufficient proof that an additional meaning is proper to the construction in question, they can at least slightly indicate this possibility. In other words, there is no sufficiently persuasive evidence that (6) necessarily includes the meaning ‘show’. We would like to have the possibility of including additional semantics in a causative construction in our model once it is proven. So, we follow D. B. Gerdts and Hukari 2006b, p. 143 and their translation.

To a smaller extent, this phenomenon is also proper to Nivkh: the permissive semantics of (2) is not only reflected in translation but explicitly stated as a distinguishing property (Nedyalkov, Otaina, and Kholodovich 1969, p. 192).

3 Key concepts and methods

To approach this topic, we would need to suggest models of syntactic and semantic representations as well as provide a mechanism for linking them. We take this architecture (Van Valin 2005, Fig. 5.4 on p. 134) and some other general principles from Role and Reference Grammar. This theory has been developed with linguistic diversity in mind and thus suits our goals well.

Syntactic representations in RRG are realized as trees, where each layer corresponds to a syntactic entity. Our study will be dealing with CORE structures – syntactic entities comprising the predicate with all its arguments, but nothing more. The predicate is placed in the NUCLEUS, being the essential part of the CORE. A CLAUSE, which is a well-known unit in any linguistic paradigm, is built upon a CORE and also includes PERIPHERY (non-arguments).

We also use the concept of macroroles from the classical RRG (Van Valin 2005, pp. 60–68) and presume that any structure would have one Actor one Undergoer and one Non-macrorole participant. Using macroroles is helpful when working with various verbs that take different thematic roles, especially in several languages.

For the semantic representations, we use frames in the form of attribute-value matrices as they allow for keeping track of typed features. We follow the approach suggested by Osswald and Kallmeyer 2018 (more discussion and comparison with other solutions can be found in Kallmeyer and Osswald 2013).

In classical RRG, linking the two representations is due procedurally by a consecutive application of a number of rules (Van Valin 2005, pp. 149–150). This approach has been criticized (see Kailuweit 2013; Kallmeyer, Lichte, et al. 2016 *inter alia*). With frames, linking is possible due to typed features assigned to semantic or syntactic structures. There are also interface features, which are used for indexing components in both representations. See Kallmeyer, Lichte, et al. 2016 for more detail about argument linking.

The formalization of both representations, together with features, can be best realized as a metagrammar, which is afterward fed to a parser. We use the extensible metagrammar (XMG) formalism suggested by Crabbé et al. 2013. Due to the lack of space, we do not include a detailed presentation of how the metagrammar describing the constructions in question could look like, and focus on the first steps of modeling the linking.

By general design and purpose, the data structures that result from the metagrammar, correspond to constructional schemata as defined in Van Valin 2005, p. 132: they contain syntactic, morphological, semantic and pragmatic information about the construction and capture language-specific information, referring it to the general principles. However, we claim that any structure can be described in such a way, even the most general one.

So, what we present in Section 4 is the complex data structure comprising a syntactic tree, a semantic frame, typed features describing different properties of either of these representations and labels for constituents. We claim that some traits common for all causative constructions can be defined universally, while some others are language-specific. By designing the universal concept and then tailoring it according to each construction in individual languages, we capture the variety and, at the same time, make reasoning about generalizations legitimate. The transition between universal and language-specific models is granted by the XMG inheritance mechanism and makes the addition of other languages simple.

4 Suggested model

4.1 The universal concept

This section shows the basic architecture of linked syntactic and semantic representations together with relevant features (consider Fig. 1). The following subsections present this general concept's applicability to the chosen languages and deal with interesting phenomena listed for each of them in Section 2. The subsections are ordered from simplest to most complicated issues.

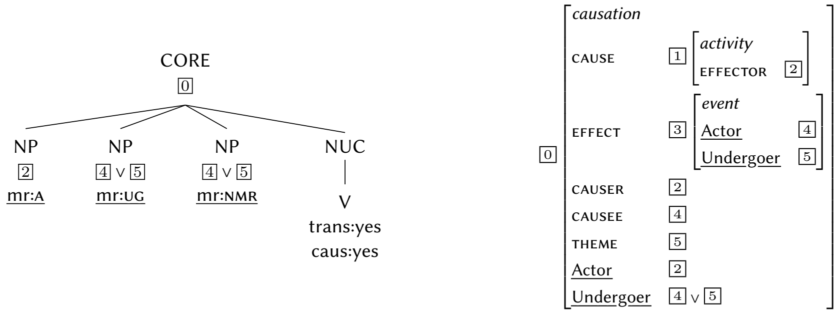


Fig. 1. Basic structure for describing constructions in question

In Fig. 1, the part on the left shows the tree-shaped syntactic representation. Features that are present in any language are already introduced into it and assigned necessary values. Namely, to describe causative constructions based on transitive verbs, one must ensure that the features 'causative' and 'transitive' are declared for a predicate and that both take positive values. Some other features (e.g., case marking) differ from one language to another and thus are not part of the general constructional schema[4].

The shape of the tree is determined by our motivation to study constructions with three core arguments. The three-argument syntactic template is not supposed to be stored in the inventory of a language nor built from scratch. We approach morphological causatives as verbal derivations and thus would like to capture the relationship between the base transitive non-causative construction and its causative counterpart. Therefore we suggest that the causative syntactic tree inherits the properties of the non-causative one but differs from it in the value of the feature 'causative' on the verb and the number of NP nodes. Inheritance is provided by the *import* method in XMG. We currently leave the detailed description of this mechanism outside of this paper.

The semantic representation is shown on the right in Fig. 1. It is a frame corresponding to the whole CORE structure (consider label $\boxed{0}$). The frame of the *causation* comprises two subframes. The CAUSE is always a general activity performed by an EFFECTOR. In all languages, this participant is also the one to bear the semantic role of the CAUSER and the Actor macrorole in the CORE construction (consider features and labels beneath the second subframe).

The EFFECT subframe corresponds to the event described by the transitive base verb. Whatever it is, it will involve two participants, one of which would bear the Actor macrorole and the other – the Undergoer. Moreover, the Actor of the EFFECT will always be the CAUSEE in the causative construction, while the Undergoer of this event will be the THEME. However, languages differ regarding which of these participants

[4] The constituent order is an individual property of each language and thus must not be accounted for in the universal structure. The present order of the constituents has been chosen for facilitating the reading of the graph.

would bear the Undergoer macrorole in the CORE construction; therefore, the value to this feature is not assigned in the general constructional schema.

Participants are assigned thematic roles as well (which in our architecture is a semantic feature). As they depend on each particular predicate, they are not part of the universal structure shown in Fig. 1.

4.2 Nivkh: the instantiation of the basic structure

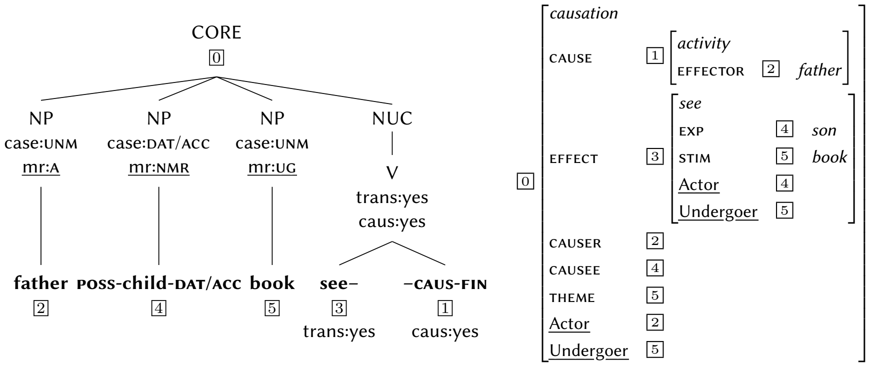


Fig. 2. A full analysis of Nivkh (1)

As already stated, example (1) from Nivkh does not have any additional syntactic or semantic peculiarities. We use it to illustrate how the general concept described in Section 4.1 is applied to a given language – consider Fig. 2.

The frame representation repeats the structure suggested in Fig. 1, but the general values of some features are substituted with specific ones. Thus, participants of the base predicate 'see' are assigned thematic roles EXPERIENCER and STIMULUS. According to the Actor-Undergoer hierarchy (Van Valin 2005, p. 54, modified by Kallmeyer, Lichte, et al. 2016), EXPERIENCER would always take the Actor macrorole when the STIMULUS is present. This is reflected in the macrorole assignment within the EFFECT subframe. Studies in Nivkh (Nedyalkov, Otaina, and Kholodovich 1969) show that the Undergoer of the embedded frame becomes the Undergoer of the whole causative construction, the CAUSEE being the Non-macrorole participant. The final macrorole assignment is presented in the last lines of the whole frame.

In our analysis, the macrorole feature is present in both syntactic and semantic representations and ensures correct linking between those. In Fig. 2, the syntactic representation has all values of the feature 'mr' determined.

Since we are dealing with a language with some morphology, features are assigned not only to complete words but also to morphemes. It is the causative suffix that makes the verb causative. So, the binary feature 'causative' receives its positive value within the suffix node. It percolates upwards to make the value of the feature 'causative' positive on higher levels and ensure the correct tree structure.

There is a new feature in the syntactic representation, which was not shown in the basic structure. The feature 'case' is language-specific and takes as its value one of the morphological cases available in a paradigm of a given language[5]. This feature is crucial for the very first step of sentence parsing (as it helps identify NPs) and the very last step of sentence generation (as it makes the sentence grammatical). The selection of cases from the language paradigm is closely related to macrorole assignment and is made on the metagrammatical level through the intersection of two XMG classes.

4.3 Halkomelem: nuances of causative semantics

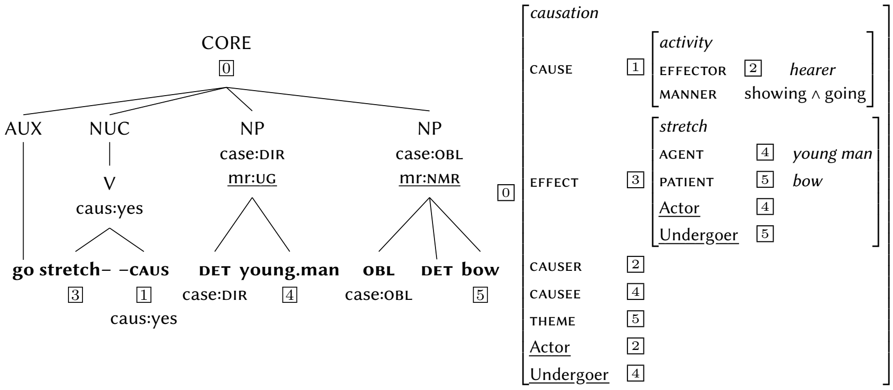


Fig. 3. A full analysis of (6) with the feature MANNER in the CAUSE subframe

Nuances of causative semantics described and discussed in Section 2.3, illustrate non-compositionality: the resulting meaning of the causative verb form does not equal to the sum of its components. Indeed, there are no overt morphemes that could bring these semantics into sentences.

To account for additional meanings, we follow the approach used by Seyffarth 2019 in a study about English lexical causatives. Designing the frame for the sentence *Sylvia laughed Mary off the stage*, Seyffarth 2019, pp. 21–22 suggests the same CAUSE frame, as in Fig. 1, enriched with a semantic feature MANNER, which takes the value *laughing*. This architecture shows the decomposition of the semantics that the verb *laugh sb.* expresses without overt marking.

We find this approach useful for describing sentences like (6) – consider Fig. 3. As it is an imperative sentence, there are only two overt NPs, although the structure requires three arguments. Thus, the participant labeled $\boxed{2}$ in the frame structure is not shown in the tree structure. The two NPs are case marked: the absence of OBL

[5] We have selected for our sample only languages that have case marking to make the demonstration of how this feature works easier. Other types of marking (verb agreement, word order, etc.) can be modeled similarly.

marker marks the direct case while its presence marks the oblique case. The macrorole assignment in Halkomelem is different from Nivkh and it is the CAUSEE (not the THEME) that becomes the Undergoer of the clause (according to various syntactic tests demonstrated by D. Gerdts 2010). The last note about Fig. 3 is the absence of the feature 'transitive', it will be discussed in Section 4.5.

The main modification of the frame is the feature MANNER in the CAUSE subframe. We show that the set of values for this feature is determined by the language, but it can remain open in principle. In this particular case, we show the possibility of combining several additional meanings. Namely, the value 'showing' reflected in the translation by D. B. Gerdts and Hukari 2006b, p. 143 and the value 'going' coming from the lexical semantics of the auxiliary verb. Once again, we use this example to illustrate the performance of our model. An in-depth investigation of different semantic augments in Halkomelem causative constructions lies outside of the present paper's scope.

4.4 Bashkir: accounting for concurring strategies

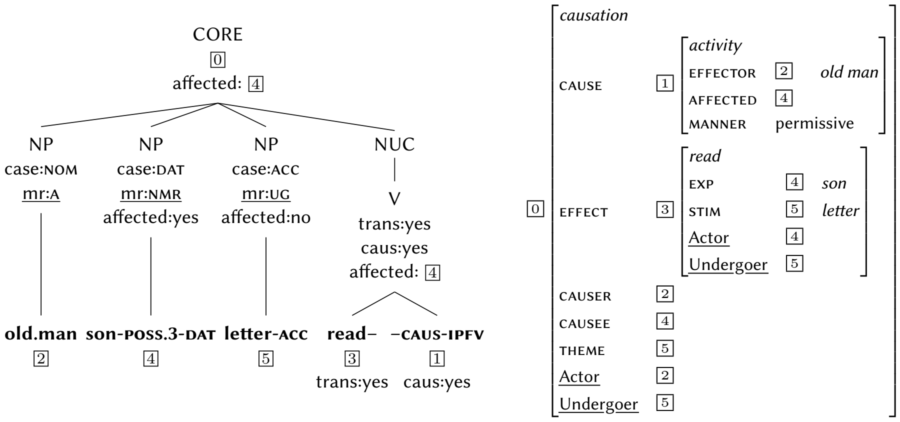


Fig. 4. A full analysis of (3b) with extensions in the CAUSE subframe

We have shown in Section 2.2 that several constructions with morphological causative can co-exist in a single language. In this section, we suggest a solution that helps to select one of the concurring constructions.

We would suggest introducing the feature AFFECTED to reflect the speaker's pragmatic intention. If the speaker intends that the THEME is affected, the ABL marking for the CAUSEE is used. If AFFECTED references the CAUSEE, it is marked with DAT. The latter construction is illustrated in Fig. 4. The CAUSE frame is enriched with the semantic feature MANNER, which has been described in the previous section. We introduced it following the translation of (3b) in Perekhval'skaya 2017, p. 244.

The feature AFFECTED appears within the CAUSE frame as it denoted which participant is pragmatically more affected by the act of causation, not by the base verb. It takes as value a label of the participant. In the depicted example, the affected participant is the causee, which is labeled $\boxed{4}$.

The value of the feature affected influences the case marking of the causee. For this reason, it must be accessible by NP nodes in the syntactic representation. This is achieved by percolating up to the CORE level and ascribing the positive value to the binary syntactic feature affected on the respective NP. In Fig. 4, the syntactic feature affected is declared for both non-nominative NPs. Since the semantic feature affected references the causee, the value of the syntactic feature on this NP becomes positive. In the model for (3a) the semantic feature affected would reference the theme and respectively the syntactic feature affected ascribed to this constituent would become positive.

The advantage of this solution is that the effect frame corresponding to the transitive base verb remains unchanged, and the causative construction can be easily built upon any stem. It also facilitates linking across dimensions involving syntactic, semantic and pragmatic reasoning.

4.5 Halkomelem: more than just 'transitive'

Halkomelem three-argument causative constructions are tricky to analyze in respect of valence. In terms of our model, it is not clear how to ascribe values to the feature 'transitive' for predicates in sentences like (5) and (6).

One way would be doing nothing special and store the value for the feature 'transitive' in the lexicon. If so, three-argument constructions would result from a combination of two arguments available for the transitive base with one additional argument added through causative verb derivation. However, non-causative transitive constructions like (4) would receive values for the feature 'transitive' twice: one coming from the stem and the other coming from the explicit suffix. This solution does not seem elegant to us.

What we suggest is a more sophisticated yet more robust solution. We suggest accounting for the semantic and the syntactic transitivity separately instead of having a single feature. Moreover, we consider the discussion about the difference between the concepts of transitivity and valence in Van Valin and LaPolla 1997, pp. 147–150 and prefer formulating further claims using terms *syntactic valence* and *semantic valence*. Following Van Valin and LaPolla 1997, p. 150, we define syntactic valence of a predicate as "the number of syntactic arguments a verb takes". The *semantic valence* in turn is defined as "the number of argument positions that a verb has in its logical structure" (*ibid*).

Given that, we posit that suffixes bearing grammatical meaning (namely, transitive and causative) index the syntactic valence, whereas lexical morphemes (verb roots) have a semantic valence. In other words, semantic valence is an invariable property of the root, and syntactic valence can be changed through verbal derivations. This claim is in line with what is commonly known about suffixes in contrast to roots. Moreover, it is in line with observations made by D. B. Gerdts and Hukari 2006a about Halkomelem. After having presented lists of verbs like *mək*^{w} *'pick up'* D. B. Gerdts and Hukari 2006a, p. 508 conclude "that transitive marking, rather than functioning as a means of deriving transitive from intransitive forms, should be viewed as inflection on roots that are already semantically transitive".

Now, the way how the causative suffix functions has to be clarified. Undoubtedly, it increases the valence of the predicate by one. But now, as we distinguish between semantic and syntactic valence, we need to add precision: the causative suffix operates on the semantic valence of a verb and makes the syntactic valence equal to a number that is higher by one. In this respect, it is different from the transitive suffix that just equals the syntactic valence to two. This discrepancy in functions can perhaps explain why transitive and causative suffixes do not stack in Halkomelem. We summarize our claims in Tab. 1.

Not only this approach helps to explain and model Halkomelem data, but it can also be useful for studying other languages demonstrating non-agglutinative traits in constructions with morphological causatives.

The more straightforward structure with the single 'transitive' feature can also be converted into the more complex one with two different 'semantic' and 'syntactic valence' features. If investigations of a larger number of languages would show that the behavior attested in Halkomelem is quite frequent, we might wish to perform the conversion for the sake of uniformity of all models.

5 Conclusion

In this paper, we have shown that constructions with morphological causatives based on transitive verbs can be syntactically and semantically decomposed in a similar way independently on the language. Features determining each particular construction can be introduced for any dimension: syntax, semantics, morphology, and pragmatics. Although only two of them seem to be critically relevant for the universal concept of a causative construction, other features are shared across languages as well. This opens a possibility to develop a complex hierarchy of constructions encountered in structurally varied languages where constructions with more modifications would automatically include features from more general constructions.

Our goal in developing a formal method of analyzing constructions with morphological causatives based on transitive verbs seems to be achieved in a sense that the prototype displayed in the present paper shows compatibility with classical RRG theory, its formalization by Osswald and Kallmeyer 2018 and language data reported by

verb	semantic valence	syntactic valence		
		bare	TR	CAUS
e.g. *yays* 'work'[6]	1	1	2	1+1=2
e. g. *məḱʷ* 'pick up'[7]	2	n/a	2	2+1=3

Table 1: Syntactic and semantic valence of Halkomelem verbs

[7] D. B. Gerdts and Hukari 2006a, p. 506, D. B. Gerdts and Hukari 2006b, p. 132

[7] D. B. Gerdts and Hukari 2006a, p. 507, D. B. Gerdts and Hukari 2006b, pp. 137–138

many other scholars. It is easily extendable and, as Section 4.5 promises, not only over other agglutinative languages.

Much was left aside from this paper and has to be done in the nearest future. First of all, the implementation of the suggested formal analysis is a challenging task *per se.* Secondly, testing our approach on a larger sample is necessary to improve both the understanding of causative constructions and the parameters relevant to modeling these linguistic data. A goal for a longer-term would be to create a formal model of not only causative but also other valence-increasing constructions.

References

Comrie, Bernard and Maria Polinsky (1993). *Causatives and transitivity.* Vol. 23. John Benjamins Publishing.

Crabbé, Benoit et al. (2013). “XMG: extensible metagrammar”. In: *Computational Linguistics* 39.3, pp. 591–629.

Dixon, Robert M W and Alexandra Y Aikhenvald, eds. (2000). *Changing valency: Case studies in transitivity.* Cambridge: Cambridge University Press.

Gerdts, Donna (2010). “Ditransitive constructions in Halkomelem Salish: A direct object/oblique object language”. In: *Studies in ditransitive constructions: A comparative handbook*, pp. 563–610.

Gerdts, Donna B and Thomas E Hukari (2006a). “A closer look at Salish intransitive/transitive alternations”. In: *Annual Meeting of the Berkeley Linguistics Society.* Vol. 32. 1, pp. 503–514.

— (2006b). “Classifying Halkomelem causatives”. In: *41st International Conference on Salish and Neighbouring Languages, University of British Columbia Working Papers in Linguistics.* Vol. 18, pp. 129–145.

Gruzdeva, Ekaterina (1998). *Nivkh.* Vol. 111. Languages of the World/Materials. München and Newcastle: Lincom Europa.

Kailuweit, Rolf (2013). “Radical Role and Reference Grammar (RRRG)”. In: *Linking Constructions into Functional Linguistics: The role of constructions in grammar* 145, p. 103.

Kallmeyer, Laura, Timm Lichte, et al. (2016). “Argument linking in LTAG: A constraint-based implementation with XMG”. In: *Proceedings of the 12th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+ 12)*, pp. 48–57.

Kallmeyer, Laura and Rainer Osswald (2013). “Syntax-driven semantic frame composition in lexicalized tree adjoining grammars”. In: *Journal of Language Modelling* 1.2, pp. 267–330.

Kholodovich, A. A. (1969). *Typology of causative constructions: Morphological causative (in Russian).* Nauka, Leningrad.

Levin, Beth (1993). *English verb classes and alternations: A preliminary investigation.* University of Chicago press.

Manning, Christopher, Ivan A Sag, and Masayo Iida (1999). “The lexical integrity of Japanese causatives”. In: *Studies in contemporary phrase structure grammar*, pp. 39–79.

Nedyalkov, V. P., G. A. Otaina, and A. A. Kholodovich (1969). “Morphological and lexical causatives in Nivkh (published in Russian)”. In: *Typology of causative constructions: Morphological causative.* Ed. by A. A. Kholodovich. Nauka, Leningrad, pp. 179–199.

Osswald, Rainer and Laura Kallmeyer (2018). “Towards a formalization of Role and Reference Grammar”. In: *Applying and expanding Role and Reference Grammar (NIHIN Studies).* Ed. by Rolf Kailuweit, Eva Staudinger, and Lisann Künkel. Freiburg: Albert-Ludwigs-Universität, Universitätsbibliothek, pp. 355–378.

Perekhval'skaya, Elena (2017). “Causative constructions in Bashkir (Published in Russian)”. In: *Acta Linguistica Petropolitana. Works by Institute of Linguistic Studies* 13.1.

Reed, Lisa (1991). “The thematic and syntactic structure of French causatives”. In: *Probus* 3.3, pp. 317–360.

Say, Sergey (2009). “Argument structure of Kalmyk causative constructions (Published in Russian)”. In: *Acta Linguistica Petropolitana. Works by Institute of Linguistic Studies* 2, pp. 387–464.

Seyffarth, Esther (2019). “Modeling the Induced Action Alternation and the Caused-Motion Construction with Tree Adjoining Grammar (TAG) and Semantic Frames”. In: *Proceedings of the IWCS 2019 Workshop on Computing Semantics with Types, Frames and Related Structures*, pp. 19–27.

Shibatani, Masayoshi (1973). “Lexical versus Periphrastic Causatives in Korean”. In: *Journal of Linguistics* 9.2, pp. 281–297.

— ed. (1976). *Syntax and semantics: The grammar of causative constructions.* Academic Press.

Talmy, Leonard (1976). “Semantic causative types”. In: *Syntax and semantics: The grammar of causative constructions.* Ed. by Masayoshi Shibatani. Academic Press, pp. 43–116.

Van Valin Jr., Robert D. (2005). *Exploring the syntax-semantics interface.* Cambridge University Press.

Van Valin Jr., Robert D. and Randy J. LaPolla (1997). *Syntax: Structure, meaning, and function.* Cambridge University Press.

A Pragmatic Account of Conventionalized Metaphors

Ioana Grosu[0000−0002−1112−643X]

New York University, New York, NY
{ioana.grosu}@nyu.edu

Abstract. I consider the process of conventionalization of metaphors, and the way in which conventionalization can be incorporated into the Rational Speech Act (RSA) based modeling approach described in Kao et al. [6]. Conventionalization of metaphors is discussed in Bowdle and Gentner [2], where it is analyzed as arising due to relationships between subjects, predicates, and abstract metaphoric categories. I take a different approach to analyzing conventionalization, building on the pragmatic account in Kao et al [6]. While Kao et al. [6] give the correct predictions for novel metaphors, they do not capture the distinction between conventional and novel metaphors. I propose an extension to the model which allows for this distinction to arise, while still maintaining a feature and goal-based analysis of metaphors.

Keywords: Metaphors · Rational Speech Act · Conventionalization.

1 Introduction

Metaphoric use of language comes in many different forms. When initially considering metaphors, one would, perhaps, come up with certain literary examples.

(1) "I'm a riddle in nine syllables" (Metaphors, Sylvia Plath)

(2) "She's a rattrap if I ever seen one" (Of Mice and Men, John Steinbeck)

These literary metaphors are used in writing to several different ends. They help make the setting more vivid, evoke unique descriptions of characters and events, and frame abstract concepts. These are not the only uses of metaphors we come across, however. We often use more mundane metaphors in our everyday conversations.

(3) My new coworker's a snake. He said the broken printer was my fault!

(4) He's a late bloomer, but eventually he'll succeed.

A difference can be observed between the metaphors in (1) and (2) and the ones in (3) and (4). Intuitively, (1) and (2) are somewhat difficult to understand. What does it mean for someone to be a rattrap? In what ways can a person be a riddle? In order to understand the two metaphors, it is necessary to provide an answer to these questions. In contrast, the metaphors in (3) and (4) are easier to grasp. We immediately understand what it means for one to be a snake, or a late bloomer, even without much contextual aid.

The difference between the above examples is one widely discussed in past analyses of metaphorical language. It is often described in terms of a process known as *conventionalization.* As predicate terms in metaphors become more conventionalized, their usage is thought to shift from metaphors such as (2), to metaphors such as (3). This shift, discussed by Bowdle and Genter [2], is the foundation of the present discussion, and will be expanded on in order to analyze the use of novel and conventional metaphors within a pragmatic model. Other empirical accounts of this difference, many of which primarily take a neurolinguistic approach to metaphor understanding (e.g., [7]), will be set aside.

The purpose of this paper is to build upon the pragmatic model proposed by Kao et al. [6], which analyzes metaphors within the context of the Rational Speech Act (RSA) framework [4]. While the model in Kao et al. [6] gives the correct predictions for several metaphors, it does not entirely capture the distinction between conventional and novel metaphors. Through a closer analysis of the effects of conventionalization, I plan to extend the Kao et al. model to account for conventionalized metaphors.

For ease of analysis, and following Kao et al. [6], I only consider metaphors of the form X is Y, such as in (5) and (6), although extensions to other metaphor constructions are possible.

(5) Beauty is a fading flower.

(6) Life is a journey.

I begin by describing some patterns which arise in examples of novel and conventional metaphors, then generalize the patterns to an informal account of the process of conventionalization. I develop this account into an extension of Kao et al.'s RSA framework, and discuss some potential further refinements.

2 From Novel to Conventional Metaphors

2.1 Empirical Observations

In order to more accurately generalize what it means for a metaphor to be conventionalized, I consider several examples of both novel and conventional metaphors. I take conventionalization to be, more precisely, the conventionalization of the predicate term. It is important to note that conventionalization is a gradient process. Certain terms are more conventionalized than others, and the degree of conventionalization is often idiosynchratic for each person. The goal here, therefore, is to consider what conventionalization means in an abstract manner, and provide a possible account of the process of conventionalization.

The crucial difference one can observe between conventional and novel metaphors is the nature of the features relevant to the predicate term, which are taken to describe the subject. Features are derived in the manner specified in feature-matching accounts such as Johnson and Malgady [5], where they are taken to be potential meaning elements. Consider the following novel metaphors, in which "sunrise" is the predicate term. Sunrises potentially have the following features applicable to them: {*happy, red and orange, new beginning*}. The sentences in (7) highlight the applicability of each of the features in a metaphor. The portions in parentheses are added context, which show how the intended metaphoric meaning can be obtained.

(7) a. That flower is a sunrise. *(Its bright red and orange hues caught my eye in the garden.)*
b. Her smile is a sunrise. *(It is always warm, and brings joy to those around her.)*
c. The start of the new school year was a sunrise. *(It was an exciting new beginning.)*

Two observations about novel metaphors can be gleaned from this example. First, several features are applicable in a metaphoric context. In the case of "sunrise", each of the features specified above can be used. Second, the metaphoric reading is fairly context dependent. The intended feature(s) (and therefore the exact metaphoric reading) is not always clear without proper context. For example, in (7a), one could assume that "sunrise" is being used in order to express that the flower looks cheerful, but contextual cues can be used in order to force the reading that the flower is colorful like a sunrise. If we consider another example of a novel metaphor ("stone") we see a similar pattern arising. We take the relevant features of "stone" to be the following: {*cold, hard, on the ground*}

(8) a. His will was stone. *(It was unbreakable.)*
b. Her face was stone. *(It was grey and sullen, when she heard the news.)*
c. The bodies on the battlefield were stones. *(They were scattered haphazardly.)*

Here, we see again that depending on the context, the metaphoric use of "stone" can be taken to convey several different features. In the case of (8a), "stone" conveys that his will is hard and unbreakable. (8b) on the other hand, conveys that her face is greyish or pale (with the added context), and potentially also that it was hard (without the added context). (8c) conveys that the bodies are on the ground. Similar patterns arise for the following examples in (9) - (10).

(9) Blanket: {*warm, covering*}
a. The freshly fallen snow is a blanket. *(It covers all in a white powder.)*
b. Their love is a blanket. *(It kept them warm through the darkest of times.)*

(10) Twig: {*brittle, thin*}
a. His arms were twigs. *(They broke quickly under pressure.)*
b. His arms were twigs. (*He had barely any muscle definition.)*

If we compare these examples to highly conventionalized metaphors, we observe that the above observations no longer hold. Take the following example of a conventionalized metaphor, involving the term "anchor". Anchor can have the following feature set applicable to it: {*heavy, metal, holding steady*}. However, only the last feature appears to be usable in a metaphoric context. If we try to force other relevant features through contextual additions (such as in (11c)), the metaphor is no longer acceptable. Additionally, if the conventionalized metaphoric use cannot arise, as in (11d), the metaphor is not acceptable, despite the validity of other relevant metaphoric features (in this case, being heavy).

(11) a. Love is an anchor. *(It keeps us all grounded.)*
b. Classwork is an anchor. *(Keeping Jane inside, when she could have been playing.)*
c. The car is an anchor *(#with its metal casing shining bright in the sun.)*
d. #That elephant is an anchor.

A similar pattern can be observed in (12), where the only relevant feature of "bombshell" is striking, and any other features cannot be used in a metaphoric context.

(12) Bombshell: {*striking, dangerous, part of an explosion*}
a. The news story is a bombshell.
b. The woman is a bombshell.
c. #The general is a bombshell. *(He is quite dangerous)*

The same observation can be seen in (13) and (14), which provide further evidence for the idea that conventionalized metaphors are used to express a specific feature in a metaphoric context.

(13) Rat: {*intelligent, sneaky, destructive*}
a. The informant for the news story is a rat.
b. My sister is a rat. *(She went behind my back and told our parents I snuck out.) / (#She broke my desk trying to sit on it.)*
c. The new scientist in our lab is a rat *(#He figures out solutions really quickly.)*

(14) Zoo: {*wild, fun, educational*}
a. The classroom is a zoo. *(#It's a place where students can learn a lot of cool things.)*
b. The meditation course was a zoo. *(#We all sat very quietly, and I had a good time.)*
c. The house was a zoo *(during the party last night.)*

A final observation to note: predicate terms of metaphors have been proposed to have a lifespan, which starts from their introduction in novel metaphors [3]. As they become more conventionalized, they become conventional metaphors.

Eventually, they may reach the point where they can no longer be considered metaphors. That is, the metaphoric use becomes fully lexicalized, and is no longer literally false. These extremely conventionalized metaphors are often referred to as *dead metaphors*.

Two dead metaphors are given in (15) and (16). As an example, the original meaning of "blockbuster" was a large explosive which could demolish a city block. Over time, that meaning got lost, and now "blockbuster" means something which is highly successful. While one can still maintain the original interpretation of "blockbuster", using the 'metaphoric' interpretation is no longer a case of non-literal language use. Similarly, "laughing stock" originated from the practice of putting people in stocks as punishment, but the meaning was later replaced.

(15) That movie was a blockbuster.

(16) He was the laughing stock *(of the class)*.

While the death of a metaphor is an interesting empirical issue on its own, I will set it aside for now, and focus on "real" metaphors.

2.2 Defining Conventionalization

Taking into consideration the empirical observations given above, it is now possible to provide an informal explanation of the effects of conventionalization. Here, I deviate from the analysis in Bowdle and Gentner [2] and Bowdle and Gentner [3]. In these studies, conventionalization is described as a shift in processing during which one no longer is required to compare the subject term to the predicate term, but in which one can take to predicate term to evoke a superordinate metaphoric category. I instead opt to take a more feature-based approach to metaphor understanding, following the account in Kao et al. [6].

The primary observable difference between conventional and novel metaphors is the fact that a conventionalized predicate term is used to convey one specific feature within the context of a metaphor. I have given examples of this observation in the above section. On the other hand, novel terms can be used to convey a wider range of features, and the usable feature is largely dependent on contextual cues alongside the relative applicability of the feature to the subject.

From this, I claim that the process of conventionalization is, in fact, a shift from the potential use of multiple applicable features in a metaphoric context, to the requirement that one feature be used. As a certain metaphoric interpretation becomes more conventionalized, the feature corresponding to that interpretation becomes more likely to be the relevant feature to be used. Conventionalization, therefore, eventually results in only one relevant feature being available for use given the predicate term of a metaphor. This diverges from the view in Bowdle and Gentner [2], since it posits that there is no abstract metaphoric category being referred to in the case of conventional metaphors, but rather a greater restriction on possible applicable features.

3 RSA Model of Metaphors

In order to account for conventionalization within a pragmatic account, I rely, as a foundation, on the RSA model proposed by Kao et al. [6]. It builds off of the basic model proposed by Frank and Goodman [4]. As in the basic RSA model, it posits listener and speaker layers which reason recursively about one another, in order to arrive at a pragmatically enriched meaning of an utterance. A speaker reasons about a literal listener (who interprets an utterance literally), and chooses the maximally informative utterance based on the expected interpretation of the literal listener. A pragmatic listener reasons about the speaker, and infers the meaning of the utterance based on the speaker's expected behavior.

The crucial difference between the metaphor model and the basic RSA model is the introduction of the speaker's goals. A speaker has certain communicative goals, and chooses utterances which help satisfy these goals. Through this addition, the Kao et al. [6] model can derive metaphoric interpretations. In order to make my contribution in Section 4 more clear, I begin by discussing each layer of the original Kao et al. [6] model, and the process through which these layers work together to allow for a derivation of metaphoric interpretations.

3.1 Literal Listener

The literal listener is modeled by the following equation,

$$L_0(c, \vec{f}|u) = \begin{cases} P(\vec{f}|c) & \text{if } c = u \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where c is the actual category of the entity being described, u is the uttered category, $\vec{f}$ is a feature vector which gives the possible applicable features for the entity being described, and $P(\vec{f}|c)$ is the prior probability that a member of the category c has the feature vector $\vec{f}$.

The literal listener, as the name suggests, interprets an utterance literally. Equation (1) shows the way in which this literal interpretation arises. In order to illustrate this, Kao et al [6] use, as an example (17):

(17) John is a shark.

In the case of (17), one can posit three features for $\vec{f}$: {*sleek, finned, scary*}. One can additionally take u to be "shark", corresponding to the predicate term. Finally, for c one can posit two categories: "shark" or "person".

The literal listener's role is to reason from the utterance to the likely category and feature vector. If the reasoned-over category matches the utterance (that is, if John is literally a shark), then the listener computes the joint probability of c and $\vec{f}$ given u to be $P(\vec{f}|c)$. This results in a literal interpretation of the utterance. In other words, if the literal listener hears the sentence in (17) they interpret it as meaning that John is a member of the category "shark", and assume that John has the corresponding features of a shark.

3.2 Speaker

The speaker is defined by the following equation.

$$S_1(u|g, \vec{f}) \propto e^{\lambda U(u|g,\vec{f})} \tag{2}$$

The speaker's role, as stated above, is to reason over the possible literal listener interpretations and choose an optimal utterance. The crucial component in deriving metaphoric interpretations is the introduction of the goal, which is denoted by g. The speaker's goal is to communicate the value of a feature. In other words, $g_i(\vec{f}) = f_i$ is taken to be the speaker's goal to communicate about the values of feature i. Following the example in (17), one could posit that the speaker's goal is to convey that John is scary. This means that the speaker wants to communicate that the value of "scary" (within the relevant feature vector) is 1.

To choose the optimal utterance, the speaker uses a utility function, which allows the speaker to act as an approximately rational planner. This utility function is the negative surprisal of the world state given an utterance. "State" in this case, refers to the speaker's goal to communicate a certain vector value. The utility of the speaker is the following.

$$U(u|g, \vec{f}) = \log \sum_{c,\vec{f}'} \delta_{g(\vec{f})=g(\vec{f}')} L_0(c, \vec{f}'|u) \tag{3}$$

Here, $\delta_{g(\vec{f})=g(\vec{f}')}$ is 0 if the literal listener's feature vector does not match the feature vector reasoned over by the speaker. The speaker utilizes the logarithm of the sum of L_0 probabilities in which the goals reasoned over by the listener match the goal of the speaker. Within the speaker equation, the utility is used in order to reason over utterances. The speaker's decision is influenced by the speaker optimality parameter, λ.

From the equations given in (2) and (3), one can therefore claim the following: the speaker knows that if they produce the utterance in (17), the literal listener will believe that John is literally a shark, and is therefore likely to be scary. The speaker will therefore be motivated to produce the utterance, since their goal is satisfied if the listener believes that John is scary.

3.3 Pragmatic Listener

The final component of the model is the pragmatic listener. The equation for the pragmatic listener is as follows.

$$L_1(c, \vec{f}|u) \propto P(c) \cdot P(\vec{f}|c) \cdot \sum_{g} P(g) \cdot S_1(u|g, \vec{f}) \tag{4}$$

The pragmatic listener's role is to marginalize over the speaker's goals, in order to determine the intended meaning. The pragmatic listener's decisions are informed by prior world knowledge. Within this equation, the following three priors are considered: $P(c)$ is the prior probability that the entity under consideration is a member of c. In the case of example (17), while it is likely that John is a human, there is a non-zero probability that John is actually a shark. $P(\vec{f}|c)$ is the prior probability that a member of c has features $\vec{f}$. This probability was obtained in Kao et al. [6] through experimental testing. Finally, $P(g)$ is the probability that the speaker has a goal g. This probability changes based on contextual information. To illustrate, if the speaker was asked "What is John like?" this likely results in a uniform prior probability distribution over goals. However, if they were asked "Is John scary?", there would be a higher probability assignment to the goal of conveying scariness.

From this model, it is, indeed predicted that the listener would arrive at the interpretation that John is not actually a shark, but that the utterance was meant to convey that John is scary. If the pragmatic listener thinks that it is likely that the speaker's goal is to convey the feature "scary", and knows that it is also unlikely that John is actually a shark, the pragmatic listener then determines that shark is being used metaphorically.

4 The Proposed Extension

Now that I have considered the basic pragmatic account of metaphors, I turn to the issues that arise for the Kao et al.[6] model in the case of conventionalized metaphors. To illustrate this, I refer back to the example in (11c), repeated below.

(18) The car is an anchor. *(#Its metal casing shines bright in the sun.)*

Here, anchor can only be used to refer to the fact that the car is a stabilizing force, and not to the fact that the car is heavy or made of metal. However, it is a priori almost certain that the car has the feature of being made of metal, and reasonably likely (although not as strictly required) that it is also heavy and is a stabilizing force (i.e., "steady").

In Figure 1, the results from the Kao et al. [6] model for the sentence in (18) are shown. As part of the implementation, I posited reasonable toy priors (given in the appendix) for the applicability of features to the given categories. As can be observed in Figure 1, since "metal" is most likely to be a feature of "car", when the listener hears "the car is an anchor" they interpret that as meaning that the speaker wished to convey the feature "metal". However, because "anchor" is a conventionalized metaphoric term, the speaker's intended interpretation given "anchor" is far more likely to be "steady", reflecting its conventional use. Currently, the Kao et al. [6] model does not capture this effect of conventionalization.

An extension to the Kao et al. [6] model is therefore required in order to account for cases such as this. I have previously posited that conventionalization arises from the features associated with the uttered category. This means that one cannot rely on the prior probabilities of the speaker's goals alone, in order to arrive at the correct interpretation of a conventionalized metaphors, since the probability that a certain feature is usable is dependent on the utterance. I

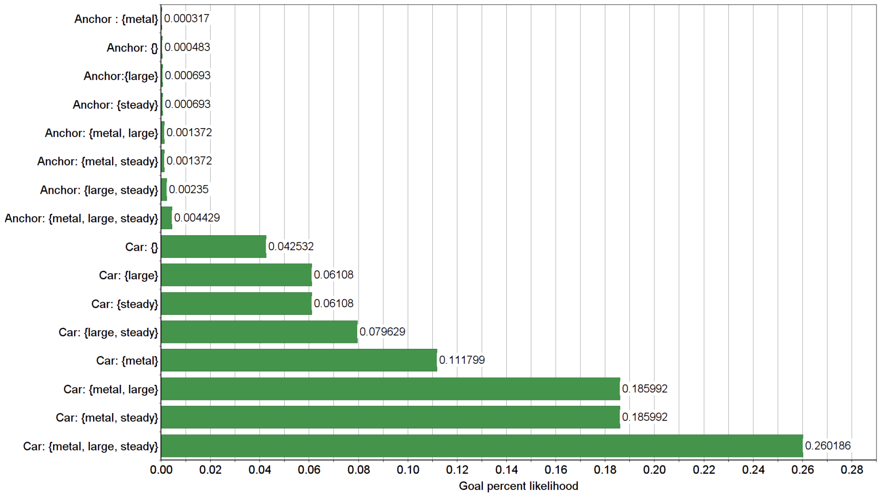


Fig. 1. Kao et al. [6] model results (L_1 output)

therefore propose the inclusion of another prior, over possible interpretations of the utterances. In this case, the interpretations are taken to be the feature vectors that the listener believes the speaker intends to convey by using a predicate term. The inclusion of this prior over feature vector sets is formally analogous to the inclusion of prior over multiple lexica in the lexical uncertainty model proposed by Bergen et al. [1]. However, there exist some crucial differences in the way the interpretations of the utterances are defined, and in the inclusion of formal aspects of the Kao et al. [6] model. Below are the equations for the extended model.

$$L_0(c, \vec{f}|u, \mathcal{F}) = \begin{cases} P(\vec{f}|c) \cdot \mathcal{F}(u, \vec{f}) & \text{if } c = u \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The literal listener in the extended model behaves similarly to the literal listener in the Kao et al. [6] model. The primary difference is the addition of the function $\mathcal{F}$, which outputs 0 if the feature vector$\vec{f}$ is not included in the set of feature vectors given a utterance u, and 1 if $\vec{f}$ is included in the set of vectors given an utterance u. Essentially, the literal listener now checks both whether the predicate term matches the category under consideration, and whether the feature vector under consideration is a part of the relevant set of applicable feature vectors given an utterance. This allows for more nuance in interpreting the utterance, since different possible interpretations of the utterance are being utilized.

$$S_1(u|g, \vec{f}, \mathcal{F}) \propto e^{\lambda U(u|g, \vec{f}, \mathcal{F})} \tag{6}$$

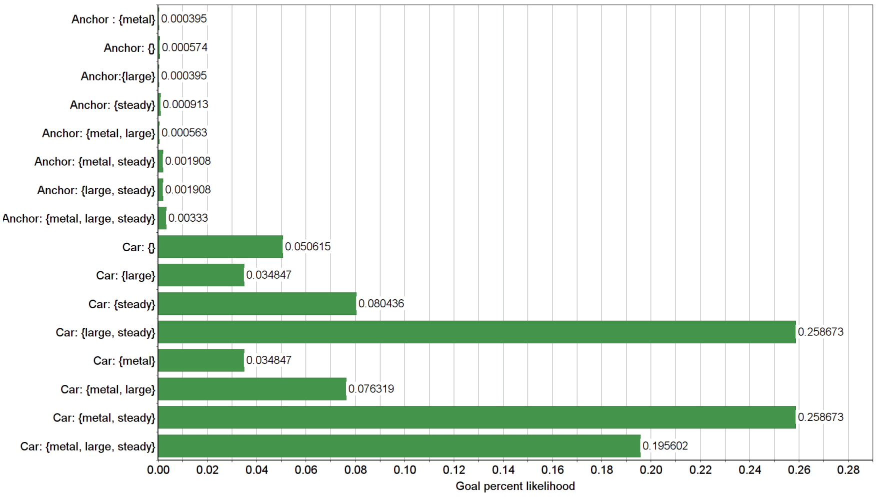


Fig. 2. Conventionalization model results (L_1 output)

$$U(u|g,\vec{f},\mathcal{F}) = \log \sum_{c,\vec{f}'} \delta_{g(\vec{f})=g(\vec{f}')} L_0(c,\vec{f}'|u,\mathcal{F}) \tag{7}$$

The speaker in the extended model is almost identical to the speaker in the original Kao et al. [6] model. The primary difference is that the speaker now outputs a probability distribution over utterances given a set of applicable features, which the speaker also takes into account when considering the literal listener model.

In sum, given a certain interpretation for the possible utterances, the speaker is likely to choose the utterance which will most optimally lead the literal listener to the correct interpretation of the speaker's goal.

$$L_1(c,\vec{f}|u) \propto P(c) \cdot P(\vec{f}|c) \cdot \sum_{g,\mathcal{F}} P(g) \cdot P(\mathcal{F}) \cdot S_1(u|g,\vec{f},\mathcal{F}) \tag{8}$$

The pragmatic listener layer is primarily where the priors over feature vector sets ($P(\mathcal{F})$) are incorporated (priors and sets of vectors are provided in the Appendix). The pragmatic listener marginalizes over goals and applicable feature vector sets, in order to determine the speaker's intended meaning.

Using these modified equations in an implementation of the metaphor model outputs the results in Figure 2, which is a much more intuitive prediction given the sentence in (18). When the listener hears the utterance "The car is an anchor", it is likely that the intended feature is that the car is steady. The results in Figure 2 show that the highest probabilities are assigned to feature sets in which "steady" is a member, and in which the intended category is "car." This is an improvement over the results in Figure 1, in which the feature sets containing "metal" are assigned the highest probabilities.

In the case of novel metaphors, one would set the weights over possible interpretations heavily skewed towards the feature vector set which contains all possible vectors (Feature vector set 1, in the Appendix), which would give the same result as the original Kao et al. [6] predictions. The model therefore accounts for both novel and conventional metaphors.

5 Conclusion

In this paper, I have outlined several empirical observations concerning conventionalized metaphors, and have provided a general account of the process of conventionalization. Crucial to this account is the idea that conventionalization requires a shift from allowing for multiple applicable features to be used in a metaphoric context, to limiting the set of applicable features to one conventionalized feature. Based on this claim, I have provided an extension to the RSA model by Kao et al. [6], which accounts for conventionalized metaphors.

Several possible further refinements to this approach would be interesting to pursue. The simplest extension would be to consider refinements within the model itself. The priors used to generate the above results were toy priors. In order to obtain more accurate results, one could run empirical tests similar to those in Kao et al. [6], in order to obtain more accurate priors. In particular, it would be an important next step to generate weights over feature vector sets based on empirical results, since the current weights are assigned based on reasonable assumptions about listener knowledge of the various interpretations. Second, a more principled way to generate applicable sets of feature vectors would be interesting to pursue. For the purposes of the current study, I posited sample sets of feature vectors, with the aim of representing different degrees of conventionalization. However, there are many possible sets of vectors one could potentially incorporate in the model, which may lead to interesting predictions.

In addition, as was mentioned in Section 2.1, one could potentially explore the process of metaphor death, and the way it relates to the account of conventionalization. Finally, and perhaps most importantly, one could explore the theoretical validity of using features-matching in an account of metaphors. While features allow for a relatively clean analysis of conventionalization, they lead to a larger issue of establishing correspondences between domain-specific properties. For example, a bombshell being "striking" is obviously different compared to a news story being "striking". In order to account for this, one would have to provide a further refinement on the conceptualization of features, and on the ways in which domain-specific properties can be generalized.

References

1. Bergen, L., Levy, R., Goodman, N.: Pragmatic reasoning through semantic inference. Semantics and Pragmatics **9**(20) (2016)
2. Bowdle, B., Gentner, D.: Metaphor comprehension: From comparison to categorization. In: Proc. of the 21st Annual Conference of the Cognitive Science Society. vol. 21, p. 90. Lawrence Erlbaum Associates (1999)

3. Bowdle, B.F., Gentner, D.: The career of metaphor. Psychological review **112**(1), 193 (2005)
4. Frank, M.C., Goodman, N.D.: Predicting pragmatic reasoning in language games. Science **336**(6084), 998–998 (2012)
5. Johnson, M.G., Malgady, R.G.: Some cognitive aspects of figurative language: Association and metaphor. Journal of Psycholinguistic Research **8**(3), 249–265 (1979)
6. Kao, J., Bergen, L., Goodman, N.: Formalizing the pragmatics of metaphor understanding. In: Proc. of the Annual Meeting of the Cognitive Science Society. vol. 36 (2014)
7. Lai, V.T., Curran, T., Menn, L.: Comprehending conventional and novel metaphors: An erp study. Brain Research **1284**, 145–155 (2009)

Appendix

Feature set priors:

- **Anchor (category prior: 0.01)**
  {metal : 1, large : 1, steady : 1}: 0.30
  {metal : 1, large : 1, steady : 0}: 0.13
  {metal : 1, large : 0, steady : 1}: 0.13
  {metal : 1, large : 0, steady : 0}: 0.05
  {metal : 0, large : 1, steady : 1}: 0.13
  {metal : 0, large : 1, steady : 0}: 0.05
  {metal : 0, large : 0, steady : 1}: 0.05
  {metal : 0, large : 0, steady : 0}: 0.05
- **Car (category prior: 0.99)**
  {metal : 1, large : 1, steady : 1}: 0.2
  {metal : 1, large : 1, steady : 0}: 0.2
  {metal : 1, large : 0, steady : 1}: 0.2
  {metal : 1, large : 0, steady : 0}: 0.2
  {metal : 0, large : 1, steady : 1}: 0.05
  {metal : 0, large : 1, steady : 0}: 0.05
  {metal : 0, large : 0, steady : 1}: 0.05
  {metal : 0, large : 0, steady : 0}: 0.05

Interpretations:

- **Feature vector set 1 (weight = 0.07)**
  'anchor' : [{metal : 1, large : 1, steady : 1}
  {metal : 1, large : 1, steady : 0}
  {metal : 1, large : 0, steady : 1}
  {metal : 1, large : 0, steady : 0}
  {metal : 0, large : 1, steady : 1}
  {metal : 0, large : 1, steady : 0}
  {metal : 0, large : 0, steady : 1}
  {metal : 0, large : 0, steady : 0}]

- **Feature vector set 2 (weight = 0.15)**
  'anchor' : [{metal : 1, large : 1, steady : 1}
  {metal : 1, large : 0, steady : 1}
  {metal : 0, large : 1, steady : 1}
  {metal : 0, large : 0, steady : 1}]

- **Feature vector set 3 (weight = 0.3)**
  'anchor' : [{metal : 1, large : 0, steady : 1}
  {metal : 0, large : 1, steady : 1}
  {metal : 0, large : 0, steady : 1}]

- **Feature vector set 4 (weight = 0.4)**
  'anchor' : [{metal : 0, large : 0, steady : 1}]

- **Feature vector set 5 (weight = 0.01)**
  'anchor' : [{metal : 1, large : 1, steady : 1}
  {metal : 1, large : 1, steady : 0}
  {metal : 1, large : 0, steady : 1}
  {metal : 1, large : 0, steady : 0}]

- **Feature vector set 6 (weight = 0.01)**
  'anchor' : [{metal : 1, large : 1, steady : 0}
  {metal : 1, large : 0, steady : 1}
  {metal : 1, large : 0, steady : 0}]

- **Feature vector set 7 (weight = 0.01)**
  'anchor' : [{metal : 1, large : 0, steady : 0}]

- **Feature vector set 8 (weight = 0.01)**
  'anchor' : [{metal : 1, large : 1, steady : 1}
  {metal : 1, large : 1, steady : 0}
  {metal : 0, large : 1, steady : 1}
  {metal : 0, large : 1, steady : 0}]

- **Feature vector set 9 (weight = 0.01)**
  'anchor' : [{metal : 1, large : 1, steady : 0}
  {metal : 0, large : 1, steady : 1}
  {metal : 0, large : 1, steady : 0}]

- **Feature vector set 10 (weight = 0.01)**
  'anchor' : [{metal : 0, large : 1, steady : 0}]

- **Feature vector set 11 (weight = 0.01)**
  'anchor' : [{metal : 0, large : 0, steady : 0}]

- **Car is the same across all feature vector sets:**

  'car' : [{metal : 1, large : 1, steady : 1}
  {metal : 1, large : 1, steady : 0}
  {metal : 1, large : 0, steady : 1}
  {metal : 1, large : 0, steady : 0}
  {metal : 0, large : 1, steady : 1}
  {metal : 0, large : 1, steady : 0}
  {metal : 0, large : 0, steady : 1}
  {metal : 0, large : 0, steady : 0}]

Hello Siri, Why Don't You Understand? – A Study on Grounding of Human and Agent Interlocutors in Dialogue

Wan Ching Ho[1]

Saarland University, 66123 Saarbrücken, Germany
wanching@coli.uni-saarland.de

Abstract. One of the reasons limiting common usage of dialogue agents is that they couldn't "understand" more complicated requests due to failures in communicative grounding attempts, which establishes mutually agreed upon knowledge. Distinct from researches that evaluate dialogue agents' performance using the rates of successfully completed tasks, this paper takes the linguistic approach of discourse analysis and investigates practical differences in how human and agent interlocutors make use of language to reach common ground in human-human and human-agent (Siri) dialogues given same tasks. Utilising the Degrees of Grounding model [10] and the modified Incremental Semantic Processing Model [5], the paper identifies four major weaknesses in Siri: (i) uninformative request repair, (ii) greedy use of grounding evidence, (iii) difficulties in interpreting resubmit and (iv) inability to understand human grounding strategies. Apart from the surface distinct language use, findings hint at a deeper challenge in optimising agents' presentation to help adjust anticipated replies. In view of Siri's expressions that had misled users about its perceptions, a number of mitigation strategies are suggested.

Keywords: Dialogue Agent · Grounding · Discourse Analysis

1 Introduction

Communication, like many other collaborative acts, requires common knowledge of participants. Grounding is the crucial process that helps to achieve mutually agreed knowledge [2]. As every interlocutor has their distinctive perception and acquired world knowledge, grounding addresses the potential discrepancies in knowledge by establishing a common ground that facilitates efficient contributions in conversations.

While grounding attempts in human conversations eventually resolves ambiguities and discrepancies in interlocutors' understanding most of the time, it is significantly more challenging for dialogue agents, often leaving miscommunication unresolved. In order to expand the applications of task-oriented agents to a wider range of contexts, successful grounding becomes an essential task when dealing with more complex commands and responses.

This paper aspires to answer the question “why aren’t ambiguities successfully resolved by grounding in human-agent dialogues as that in human-human dialogues?” from a conversational perspective. Focusing on studying the use of grounding evidence and other manners of how grounding is accomplished by human and agent interlocutors in human-human and human-agent dialogues, the paper investigates the factors that limited the effectiveness of grounding in dialogues involving agents. Since studies comparing popular task-oriented dialogue agents often rank Siri as one of the *weakest* candidates [8], this study chooses Siri to discover fundamental differences in agents’ grounding attempts from humans’.

The following first presents a background of related work in section 2 and formal background in section 2.3, then the methodologies in section 3. Section 4 presents data analysis, where section 4.1 concerns positive grounding evidence and 4.2 concerns negative grounding evidence. A number of phenomena observed in the data would be addressed in the discussion of section 5. Section 6 provides suggestions to how Siri could be optimised via enhancing its grounding attempts. Section 7 finally summarises the contribution of the paper and suggests potential extentions.

2 Related Work

2.1 Theoretical Background

Grounding was conceptually divided into a *presentation phase*, where an utterance is presented to an interlocutor, and an *acceptance phase*, where the interlocutor gives evidence to show understanding of the presented utterance [3]. By “accepting” the content, one does not necessarily have to subjectively agree with the subjective views towards the information, but instead has to acknowledge the proposition expressed, eventually including discerning a referent. However, accepting a reference and a predication may fall apart: For example, when the utterance “The movie is pretty bad” is replied with “No it’s not” in the acceptance phase, the second speaker is still accepting the reference of “the movie”, but disagrees with the attribution. In such case, at least grounding of *the movie* would be successful. Accordingly, we deal with referential grounding in the following, but also look at data concerning coherent dialogue continuation as grounding in propositional level.

Early grounding studies proposed 5 types of grounding [3], however this paper adopts the extended version with 8 types by [10], summarized in Table 1.

To establish referential identity human interlocutors tend to make use of 4 major *tactics*, namely *Alternative descriptions*, *Indicative gestures*, *Referential installments* and *Trial references* [2]. *Alternative descriptions* offer a different description to the same referent; *indicative gestures* are body gestures such as pointing and touching; *referential instalments* establish identity before further mention; and *trial references* present the reference with a try marker that allows interlocutor to confirm or deny. The tactics were found in this study often used in the acceptance phase, but more so in human interlocutors, and have shaped the interlocutors’ characteristics in request repair and resubmit (see Sec. 4.2).

Table 1. 8 Types of Grounding Evidence (Roque and Traum, 2008)

Submit	when new information is first introduced in the dialogue
Resubmit	when information is presented again by the same initiator as repair
Repeat Back	when information is presented back to the initiator as confirmation
Acknowledge	when signal of agreement does not specify content
Request Repair	when a need of resubmit is indicated
Move On	when a decision of proceeding to the next task is indicated
Use	when semantic evidence shows that previous information is understood
Lack of Response	when neither of the participants speaks, showing no objection

The degrees of groundedness could be assessed by the 9-level scale of groundedness based on the patterns of evidence as in Table 2. Roque and Traum [10] applied their *Degrees of Grounding model* to the groundedness of Common Ground Units (CGU) that contains "bits of information" defined by parameters in the domain such as evidence history, mission number and grounding criteria. As this paper focuses on the evidence of understanding and the delivery of the grounding attempts, the set of information in CGUs will not be explained in detail. However, as [10] mention, a notable limitation lies in the definition of *resubmitting* as an indicator of the degree of being *accessible*, assuming that resubmit of information always indicates lack of complete understanding. The assumption was found not necessarily true in empirical data, especially when the same speaker uses it as an emphasis of previous content.

Table 2. Degrees of Groundedness and Their Patterns (Roque and Traum, 2008)

Degree of Groundedness	Pattern / Identifier
Unknown	No information introduced
Misunderstood	(anything, Request Repair)
Unacknowledged	(Submit, Lack of Response)
Accessible	(Submit) or (anything, Resubmit)
Agreed-Signal	(Submit, Acknowledgement)
Agreed-Signal+	(Submit, Acknowledgement, other)
Agreed-Content	(Submit, Repeat Back)
Agreed-Content+	(Submit, Repeat Back, other)
Assumed	Grounded by other means

2.2 Human Behaviour in Conversations

Human interlocutors contribute to grounding with the least collaborative efforts, therefore prefer minimal signals, for example, backchaneling as acknowledgement

[3]. Human interlocutors tend to produce acknowledgements to ground smaller units of information such as phrases as updates to new information incrementally. It is therefore not necessary for one interlocutor in one turn to complete expressing a concept or context, instead interlocutors could collaborate across multiple turns towards a common goal[9]. Because of the strictly bounded real-time interaction, tactics that require collaborative contribution such as trial reference and referential installments are possible [2]. Human interlocutors tend to prefer moving on to a further task after the ambiguities are resolved, where cues were given by the interlocutors showing mutual understanding [3].

This paper takes the above behaviours found in previous works of human-human interactions as assumed features of human interlocutors. Based on these assumptions, the behaviours of Siri are analyzed in Sec. 4

2.3 Formal background

The paper adopts dynamic syntax [1] as formal framework, considering the action-based grammar formalism's strengths in interpreting partial and incremental inputs as context updates. The framework takes word-by-word inputs and constructs semantic trees with nodes representing lambda calculus formaluæ. The incremental feedback extension proposed by Eshghi et al. [5,6], uses *Type Theory with Records* (TTR) proposed by Cooper [4], giving rise to a TTR version of Dynamic Syntax (DS-TTR).

Figure 1 exemplifies how a grounded utterance would be represented as contextual update. Pointers mark nodes under development, where the self-pointer ◇ (white diamond) and the other-pointer ⧫ (black diamond) represent the two interlocutors' signalled state of comprehension and acceptance. When the location of the pointers align, the utterance up to this point is considered as grounded. Grounded contents are connected by a solid line and the contents not (yet) grounded are connected by a dotted line. The application is further extended from compound contributions, self-repairs and backchanneling of human-human interactions in previous works [5] [6] to include question-answering of human-agent dialogue in this paper. In DS, *wh*-question words are taken as place-holding devices interacting with other elements such as pronouns mentioned in Kempson et al.[7] and Cann et al. [1]. Taking that in the context of dialogue, often a dialogue act of request is triggered. As this paper does not particularly concern the detailed syntactic actions within dynamic syntax, but rather focus on the use of graph representation to derive interpretations in order to understand whether interlocutors indicate their understanding of the same sequence of updates, for simplicity "REQ" was used as a convenient abbreviation in replacement of the actual wordings to adapt to the question-answering dialogue mode and to distinguish the dialogue act of request from statements. For instance, "how do you" in "how do you go to Jona in Switzerland from Saarbrücken" would be represented with "REQ". At a later point in this dialogue, represented in Figure1, the request to "go to Jona in Switzerland" has already been grounded, indicated by the solid lines. However one interlocutor understood the latter part as "from Saarbrücken" and the other as "from Sunderland", where no evidence showed that either of the options was grounded as common knowledge between them.

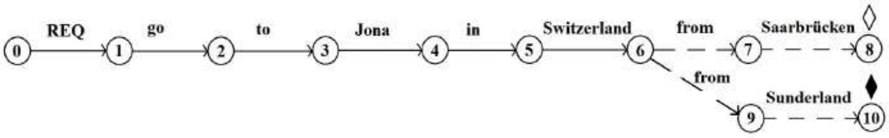


Fig. 1. Example of Graph Representation using an Incremental Semantic Processing Model according to DS [5]

3 Methodologies

Two task-oriented dialogues, one mainly between a staff and the subject at the information desk of a train station, and the other between the subject and Siri, were collected for case studies. The data was collected with voice recordings and screenshots in addition for the human-agent dialogue. The recordings have a length of 2:38 and 4:48 respectively and are partially transcribed.Here the 60 turns transcription of the human-human dialogue is referred to as excerpt 1, and the 49 turns transcription of human-agent dialogue as excerpt 2.

In order to ensure that dialogues are comparable, one student volunteered as the human subject for the data collection of both the human-human and the human-agent setting. The subject was carefully chosen to be one who has never had any interactions with Siri.

After the dialogues were collected, qualitative analysis was performed based on three frameworks (cf. Sec. 2). Based on the categorisation of *Grounding Evidence* [10] distinct distribution of usage and features of grounding evidence in human and agent interlocutors were identified. Complementarily, the *Grounding References* [2] were referred to to account for human grounding tactics, whereas the *Extended Model of Incremental Semantic Update* [6] was applied to specific cases, to further understand where and how exactly the grounding attempts fail.

4 Data Analysis

4.1 Grounding Evidence and Degrees of Groundedness

Grounding evidence includes both positive and negative evidence that indicates an interlocutor's understanding of a previous utterance. Positive evidence, for instance, *acknowledgement* and *repeat back*, are those that signal a positive understanding of the utterance. Negative evidence on the other hand, such as *request repair* and *resubmit*, reveal the occurrences of misunderstanding or miscommunication. *Request repair* and *resubmit*, despite being subcategories of grounding evidence, were separated for analysis in order to specifically account for the use of negative evidence, as well as the delivery and resolve of miscommunication through grounding. Examples of some of the grounding evidence are given below.

Submit In contrast to human-human dialogues that are strictly bounded by real-time reactions (cf. Sec. 2.2), human-agent dialogues are barely bounded by immediate response. Therefore a larger information load with multi-modal information could be found displayed beyond speech, such as web-search results in example (1).

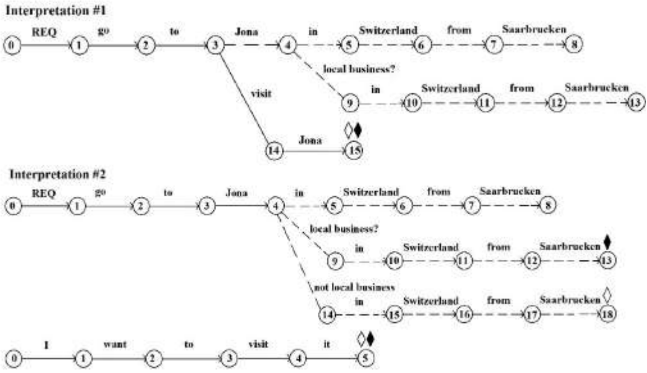
(1) a. Subject: OK. (3.0) How do you reach Jona.
b. Siri: I found this on the web: (displays Google Results text on screen:)
"Finding Jonah and defeating the Trinity army..."
"Fighting the jaguars and reaching Jonah in Peruvia n Jungle..."
Excerpt 2, Turns 45–46

Acknowledgement Unlike how human interlocutors ground smaller fragments of information due to technical limitations (cf. Secs. 1 and 2.2), Siri only acknowledges propositional levels of groundedness, as in example (2). Despite that instances of "Submit, Acknowledge" are considered "Agreed-Signal" in the degrees of groundedness, the agent's "OK" in Excerpt 2 (d) of the same sequence was followed by a request repair, indicating misunderstanding. To better understand the contents grounded by that instance of "OK", the Model of Incremental Semantic Update is applied to derive two potential interpretations of example (2) in Tab. 3.

(2) a. Subject: How do you go to Jona in Switzerland from Saarbrücken?
b. Siri: Which local business? Tap the one you want:
(in MAPS) Rapperswill Rapperswill-Jona...
c. Subject: I don't want the business? I want to visit it.
d. Siri: OK.
e. Subject: (3.2) So?
Excerpt 2, Turns 11–15

The major difference between the two interpretations of the context of example 2 lies in the information unit for grounding "I don't want the business? I want to visit it", and what the tag "OK" actually acknowledges. In Interpretation 1, the entire interaction is seen as contributing to the grounding of the earliest submitted request, where the line "I want to visit it" is taken as an update to the previously said activity in Jona. The "OK" was to acknowledge the specifics of the request, thus the request of going to visit Jona is considered grounded where the two pointers landed at the same position. In Interpretation 2, the interaction concerns the grounding of a request and a statement separately. "I want to visit it" is taken as an individual statement apart from the request. Therefore "OK" acknowledges grounding of the statement, leaving the request eventually remains ungrounded. Possibly due to the subject's rejection of the "local business" suggested by Siri, the request was discarded.

Table 3. Two Possible Interpretations of Context in Excerpt 2, Turns 11–14. Key: Black diamond: Siri's understanding, White diamond: Subject's understanding. Solid paths indicate grounded content

Interpretations	Excerpt 2, Turns 11–14
 Interpretation #1 REQ go to Jona in Switzerland from Saarbrucken local business? visit Jona Interpretation #2 not local business I want to visit it	□ Subject: How do you go to Jona in Switzerland from Saarbrucken? □ Siri: Which local business? Tap the one you want: (in MAPS) [text] Rapperswill Rapperswill-Jona... □ Subject: I don't want the business? I want to visit it. □ Siri: OK.

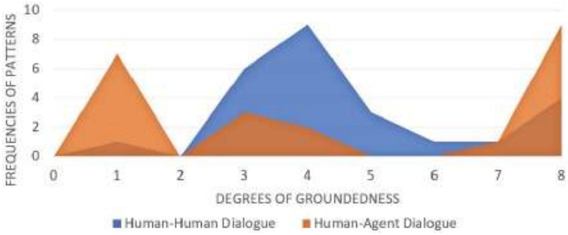
Move On, Repeat Back and Use In *Move On*, Siri was found to "jump to conclusions" and move on assuming the subject shares the same assumptions without positive grounding evidence, leading to a larger count of surface "Assumed" level of groundedness on record when the common ground was not reached. In *Repeat Back*, the agent always embeds the key terms in an utterance unlike how humans use them as trial references. Instead of repeat back, Siri tends to prefer *use*. In *Use*, when using the tactic of alternative descriptions in demonstrating comprehension, Siri uses more distantly related information from the previous utterance. As presented below, the subject paraphrased the "destinations" as "the region" in example (3), while Siri opted "Jona in Switzerland" for "ocal business", as well as providing maps references in example (2).

(3) a. Staff: Destinations where you can go with this ticket so I cannot (.) explain [And there is there (.) there exist]

b. Subject: [OK yeah yeah the region would be good]
Excerpt 1, Turns 22–23

Summary In sum, on a 9-degrees scale of groundedness(0 the lowest, 8 the highest), human-human interactions showed an accumulation of degrees of groundedness in the range of 3,4,5,8, while in human-agent interactions, groundedness was expressed in a more concentrated manner at two ends of the scale in 1,8.

Fig. 2. Degrees of Groundedness in Human-human and Human-agent dialogues



4.2 Request Repair and Resubmit

Request repair and *resubmit*, as negative evidence that signals misunderstanding, contribute to the negotiation process and how grounding moved forward. Request repair in the acceptance phase signals failure in understanding the submitted utterance. After miscommunication is expressed, usually a resubmit follows by their interlocutor in the next presentation phase as an attempt to repair or disambiguate the requested information. Apart from co-occurring with *repeat back*, *move on* and *use*, *Request Repair* and *Resubmit* are used by human and agent interlocutors drastically different with respect to *initiative*, *informativity* and *accuracy.* The observations in this section are based on the exchange in Tab. 4.

Table 4. Request Repair in Excerpt 1 by Staff and Excerpt 2 by Siri, respectively

Staff's Request Repairs	**Siri's Request Repairs**
I don't understand which ticket you mean.	Sorry I'm still not sure about that. (3×)
Ticket for five people?	I'm not sure I understand. (3×)
And then what's the name?	I don't have an answer for that.
To?	Which local business?
The name of the group	Which Joe?

Informativity *Informativity* concerns the amount of information in a request repair presented to the interlocutor as reference for their resubmit in the next turn. Humans tend to provide explicit directions to the anticipated content of repair. Taking "what's the name" in Tab. 4 as an example, the staff specified the type of answer requested as an update to the context of "group tickets for five people" in earlier exchange. One of the issues of Siri's passive template requests for repair is the low informativity and specificity of requests with little information to updating the context. As in (4), the agent did not deliver any information to whether it was the verbal reference or the command that caused the miscommunication, hence how the human interlocutor could collaborate. The human interlocutor had decided that the reference was the unclear part and resubmitted the reference. However, miscommunication was not resolved in the next turn, revealing the problem could have been from understanding the command instead. With no confirmation nor denial of any submitted information, the agent does not shift place the other-pointer ⧫ nor did it provide any new paths of interpretations.

(4) a. Subject: It's a town in Switzerland Jona h.h.h.
b. Siri: I'm not sure I understand.
c. Subject: oh my god *giggles* (1.0) JONA (.) IN (.) SWITZERLAND
d. Siri: Sorry, I'm still not sure about that.
Excerpt 2, Turns 25–27

Accuracy *Accuracy* refers to accurate expression of one's understanding, or level of groundedness. In comparison to expressions of the staff with a range of wordings, Siri's repetitive template response limited the accuracy of expressed understanding. The staff's "I don't understand which ticket you mean" expresses a lower level of understanding and "ticket for five people?" expresses a higher level of understanding. Siri's "Sorry I'm still not sure about that" and "I'm not sure I understand" in human perception should indicate two levels of understanding, however itself seemingly does not distinguish the differences between them. Even if it does, the partial information understood or doubted is not prominent to users. Such delivery is observed to have led to further hindrance in grounding due to a discrepancy in perceived understanding and fail to provide adequate information needed for repair.

Initiative *Initiative* refers to showing perceived willingness to cooperate towards a goal, therefore contributing to a more preferable response. The staff's request repair "I don't understand which ticket you mean" even though expresses miscommunication, has the initiative to move the interaction forward. In comparison, Siri's signal of negative evidences "I'm not sure I understand" and "Sorry, I'm still not sure about that" do not show efforts in cooperation. There were cases where the agent did, such as the use of "Which local business?", however were only minority cases.

5 Discussion: Siri's Weaknesses in Grounding

The weaknesses in Siri's grounding strategies can be concluded with 4 main practical challenges, elaborated in the following.

Template Responses are Undesirable Request Repairs Responses that Siri uses frequently in request repair led to inadequacy in accuracy, informativity and initiative. Inaccurate indication of the agent's understanding could alter the amount of information given by users in the next turn, resulting in insufficient knowledge for grounding hence miscommunication. When informativity is low given no specified information required in the template request repairs, human interlocutors could be confused about what to offer in the next turn, as in example (4). The low initiative perceived, from low commitment and barely any alternative provided, could frustrate users and lead to giving up of the grounding process. The success in Siri's grounding attempts is bounded by its limited knowledge of the ultimate goal and how to collaborate accordingly.

Greedy Approach in Grounding Evidence Use Siri's greedy approach in cramming multiple grounding evidence in one turn compromised the interlocutor's understanding of the most recent utterance and the predictability in their next utterance. The larger size of information, such as search results, could cause information overload hence users giving up reading and make them prone to miscommunication. Another problem is that since users may respond to any of the component mentioned, the more actions being taken, the more varied users' responses could be. That would increase the potential of unpredictable information, as well as ambiguities and discrepancies in pragmatic function, as in (b) of example 2, which could result in harder comprehension for Siri.

Difficulties in Interpreting Resubmit The difficulties Siri has in interpreting resubmits arise because the agent does not take resubmits as updates most of the time and only allows a small range of responses. Exemplified in example (2), the agent could not relate fragments of information to the ongoing task and was always looking for a new task during most of the resubmits. The agent was also bound to its own assumptions and anticipated answers, which rejecting of its suggestions could be taken as terminating the whole request.

Unable to Understand human Grounding Tactics Verbatim content register "J-O-N-A" by reading out the spelling of the word was interpreted by Siri as "Jay O NE". Meanwhile, step-by-step instalment could not be understood as related contents but instead each as a new task. Such preferences would undermine the usefulness of resubmits given by users hence slow down the progress in grounding process. On the other hand, Siri was trying to present itself as human-like by using alternative descriptions, which the descriptions ended up unfortunately causing confusion and ambiguity to the subject.

6 Suggestions

With the mentioned observations of Siri's performance, the following suggestions are put forward to potentially advance dialogue agents' strategies in grounding. The main idea of the bigger picture here is that the agent has to strike a balance between presenting itself as human-like, adapting to actual human grounding strategies, and making good use of its advantages as an agent.

For accuracy in the expression of misunderstanding, the agent should distinguish different levels and areas of understanding. For example, the information required for a request could be the area of understanding, and a confidence level could be established as the level of understanding. Based on both, agents should be using a range of wordings to specify its understanding.

Concerning informativity, when the agent is requesting repair of information, partial grounding could be incorporated into the request by repeating the known information while still required information could be explicitly named. For example, when the type of task is understood but the specifics for the task is not, the agent could first repeat or express the understanding of the task, while asking for the type of specifics with utterances such as "Where do you want me to direct you to?".

With respect to initiative, which shows Siri's devotion and willingness in cooperation, it is suggested that alternatives could be requested when linguistic miscommunication occurs. For example, when the task of navigation is known but the location is not known or is misheard, the agent could ask users to indicate the location on a map with utterances such as "Could you point to it on a map instead?". In case the task is too complicated to understand, Siri could still ask for related information then direct users to one or two websites instead of too many search results. Utterance for example "I could not assist, but maybe you can find answers about (keywords) here." could be used in such cases.

To improve Siri's comprehension, it is suggested that resubmits and submits should be distinguished and a larger range of utterances should be accepted. Even if not all the information conveyed was understood, Siri still should not give up entirely. Another important point is that Siri should allow earlier denied information to be submitted again. As in "Jona" in example 2, after being denied for once due to the agent's misunderstanding, it should still accept the keyword when the subject resubmit.

To better adapt to human communications, Siri should further introduce human grounding tactics to its own manner of communication. Essentially, the process could be broken down to smaller incremental tasks, and information should be grounded before moving on, as argued for in [9]. The grounding process is expected to be more successful if the agent could present a smaller amount of information in each turn and have each piece of information grounded before moving on. Another tactic is to adapt to human interlocutor's assumptions for displayed materials. As observed, humans take materials beside of speech as supplementary and the subject as well did not take time to look into the search results provided by Siri. To adapt to such preferences, Siri could rank the relevance of its searches, utter the title of the top few and display searches only as complementary information. Verbatim content register should be understood and could be introduced into its request repair as well. The turns required for grounding with the suggested human tactics would potentially be longer, but the possibilities in successful grounding hence completing tasks would be higher.

7 Conclusion and Possible Extensions

The paper investigated practical distinctive features of Siri in grounding attempts, as a starting point in understanding limitations in communicative grounding in human-agent dialogues. Based on the findings, challenges faced by Siri were identified and possible mitigation was suggested. However, many questions are yet left open. In future studies, it is planned to extend the study of grounding in human-agent interactions in at least three different ways: (i) a comparison between different dialogue agents (Siri, Alexa, . . .) is aimed at to pinpoint differences between the major dialogue systems today. (ii) The conversational scope of the study has to be extended in order to cover further patterns of grounding in human-agent interactions and to study about the role of expectations, and (iii) the lessons learned from studies such as those presented here have to be incorporated into implementations of dialogue agents.

References

1. Cann, R., Kempson, R., Marten, L.: Dynamics of language : An introduction. Syntax and Semantics **35**, 1–+ (01 2005)
2. Clark, H.H., Brennan, S.E.: Grounding in communication. Perspectives on socially shared cognition p. 127–149 (1991). https://doi.org/10.1037/10096-006
3. Clark, H.H., Schaefer, E.F.: Contributing to discourse. Cognitive Science **13**(2), 259–294 (1989). https://doi.org/10.1207/s15516709cog1302_7
4. Cooper, R.: Records and record types in semantic theory. J. Log. Comput. **15**, 99–112 (04 2005). https://doi.org/10.1093/logcom/exi004
5. Eshghi, A., Hough, J., Purver, M., Kempson, R., Gregoromichelaki, E.: Conversational Interactions: Capturing Dialogue Dynamics, pp. 325–349 (01 2012)
6. Eshghi, A., Howes, C., Gregoromichelaki, E., Hough, J., Purver, M.: Feedback in conversation as incremental semantic update. In: Proceedings of the 11th International Conference on Computational Semantics. pp. 261–271. Association for Computational Linguistics, London, UK (Apr 2015), `https://www.aclweb.org/anthology/W15-0130`
7. Kempson, R., meyer viol, W., Gabbay, D.: Dynamic syntax: the flow of language understanding (01 2001)
8. López, G., Quesada, L., Guerrero, L.: Alexa vs. siri vs. cortana vs. google assistant: A comparison of speech-based natural user interfaces. pp. 241–250 (01 2018). https://doi.org/10.1007/978-3-319-60366-7$_2$3
9. Poesio, M., Traum, D.: Conversational actions and discourse situations. Computational Intelligence **13**(3), 309–347 (1997)
10. Roque, A., Traum, D.: Degrees of grounding based on evidence of understanding. In: Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue. pp. 54–63. Association for Computational Linguistics, Columbus, Ohio (Jun 2008), `https://www.aclweb.org/anthology/W08-0107`

Conjunctions and Disjunctions in Probability Judgments

Cathy Agyemang

Carleton University, Ottawa Ontario, K1S 5B6

Abstract. As first introduced by Tversky and Kahneman in 1974, the conjunction error is defined an error of probabilistic reasoning where the heuristic can bias individuals to rank the conjunction or two events as more likely than one on its own. Specifically, when given a contextual description of "Linda" that is prototypically representative of a feminist, participants are more likely to chose the statement "Linda is a bank teller and a feminist" compared to the statement "Linda is a bank teller". Criticisms have challenged this notion by arguing that the comparison between an event and its conjunction is pragmatically odd, as it compares the set to its subset. The current study demonstrates that using a similar entailment between a disjunction (Linda is a bank teller or a feminist) and the conjunction significantly reduces the number of conjunction errors committed. Thus, participants are sensitive to the pragmatic circumstances of the conjunction error. Additionally, a comparison between plain disjunctions and disjunctions that only allow for an inclusive interpretation shows no difference in response patterns. A secondary analysis of the data suggests particular response strategies that participants may adopt with respect to implicature generation and probability judgments.

Keywords: Conjunction error · Implicature · Represenativeness heuristic.

1 Introduction

In their classic investigations on judgments under uncertainty, Tversky and Kahneman [24, 25] argue that individuals use cognitive heuristics to help them to make probabilistic decisions. The heuristic states that an instance that is more representative of the schematic representation of an category will be judged as more probable, regardless of the actual statistical probability of said event occurring (e.g., a robin is a highly representative instance of a bird and would therefore be more judged as a more probable instance of the bird category). Representativeness can be understood as the relative correspondence between a outcome (instance, prediction, result etc.) and the model (event, category etc.). Perceived likelihood of an outcome is influenced by the relationship between the model and a particular outcome and can be judged based on aspects such as similarity or availability [25]. Tversky and Kahneman observed the conjunction error, as a consequence of the representativeness heuristic. They found that an irrelevant context can bias individuals to misjudge the conjunction of two events as more probable than an event on its own. This was most famously demonstrated by the "Linda Problem" described in Table 1 [25, p.297]:

Table 1: The Linda Problem.

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.
Which of the following do you think is most likely to be true?
a) Linda is a bank teller.
b) Linda is a bank teller and a feminist.

They found that a higher percentage of participants chose (b) as more probable despite the conjunction of the descriptors (*bank teller and feminist*) being necessarily less probable to occur than one of its component descriptors (*bank teller*). This so-called conjunction error is attributed to the representativeness heuristic, where individuals choose instances that best conform to an overall construal of an individual or event [25].Tversky and Kahneman find further evidence to support their representativeness hypothesis, specifically that conjunction errors are virtually eliminated when a nondescript version of the problem is given [25]. Tversky and Kahneman attribute this result to the fact that the instance "Linda is a (bank teller) and feminist" is no longer representative of the overall category of feminists. That said, this finding is contested in related studies [21] among others.

The prevalence of the conjunction error is also influenced by the presentation of the problem. Hertwig and colleagues demonstrate that they can prompt a statistical interpretation of likelihood by framing the probability judgment in terms of frequencies. For example, "given the description of Linda, describe the number of women in a population of 100, who are bank tellers/bank tellers and feminists?". In this design, the relative amount of conjunction errors was significantly reduced [11, 5] Gigerenzer argues that the conjunction error is reduced in a frequency format because individuals tend represent probabilistic information in terms of frequencies [6]. This is further supported by more recent work which demonstrated when the Linda Problem is framed as a query search for a computer database, where the description is framed as a set of parameters (e.g., age = 31, college major = philosophy) conjunction errors are yet again reduced [12].

2 Criticisms and alternative views

Despite the overall appeal of the representativeness hypothesis, it has been met with various criticisms. Broadly speaking, the representativeness hypothesis has been criticized for being too vague in the requisite conditions for it to occur. Such that, there are some circumstances in which the conjunction error is robust, and other situations where individuals reason in a way consistent with probability and logic [16, 14, 23]. More specifically, there are three salient critcisms of the representative heuristic as presented in the Linda Problem: interpretations of probability [6, 11], the polysemous nature of AND in natural language [17, 10, 13] and the pragmatic considerations of implicature availability [18, 1, 22]. This paper will focus on the third criticism.

Individuals who commit the conjunction error are assumed to be observing Grice's Cooperative Principle and its subsequent maxims, which are general expectations of how interlocutors participate in conversation [7, 8]. This assumption is consequential for how individuals respond to the conjunction problem based on Grice's maxims of Quantity and Relevance, which state that conversational participants should be maximally informative and relevant in their contributions.

Adler [1] reduces the Quantity and Relevance maxims to what he coins as *selective relevance*, such that, for conjunction problems and other such contributions, participants expect that the information given in conjunction problems is true for one alternative and thereby should not be true for another or potentially all of the alternatives. The choice between "bank teller" and "bank teller and feminist" violates this expectation, as it is an odd comparison between a set and its subset [4, 18]. Adler argues that in conjunction problems, an answer that is true for more members of a set is less satisfying than an answer that applies to less members. "bank teller" would be less discriminatory based on the given description of Linda and as a result less informative, especially in comparison to the latter. Assuming an individual answering the Linda Problem is obeying these maxims, they ought to choose an answer in a way that is most informative (i.e., Linda is a bank teller and feminist) [1]. Using a similar conjunction error problem Mosconi and Macchi [15, Exp.1] demonstrate that when individuals are asked to judge the validity of the alternative that is statistically correct (e.g., Linda is a bank teller), they deemed the correct answer pragmatically inappropriate. Specifically, they either thought it was either false or reticent (partially true or informative).

Further to this effect, others argue that to reconcile the triviality of a comparison between a set and its subset by generating an implicature. Implicatures are pieces of information that are inferred above the literal meanings of the sentence [7, 8][1]. As as generalization, implicatures are presumed to arise from this

[1] For example, *Susie ate some of the cookies* implicates that she did not eat all of the cookies. Yet, it would be necessarily true that she ate some of the cookies if she did in fact eat all of them. However, since the speaker did not say that Susie ate all of the cookies, this presumption denies *all* as possibility and strengthens the meaning to *Susie ate some and not all of the cookies.*

adherence to Gricean maxims and other structural and pragmatic constraints. Implicatures deny the possibility of a more informative statement on the basis of that if one was certain about the possibility the more informative statement, they ought to state it to be maximally informative and relevant. Additionally, implicatures serve the purpose of facilitating an answer to the Question Under Discussion (QUD). An answer is the choice between relevant alternatives that sufficiently answer the QUD. Here, an answer will fully or partially satisfy a question by choosing at least one of the specified alternatives [19, 9]. Implicatures reduce the number of possibilities of answers that can address the QUD. In the context of the conjunction problem, the QUD is explicitly stated: "Which of the following statements is the most probable?". Under pragmatic constraints, it is argued that individuals generate an implicature, such that, both alternatives are on the same scale of informativity. [1, 18, 4]:

(1) Linda is a bank teller (and not feminist).
Linda is a bank teller and a feminist.

Using this inference, choosing the answer *bank teller and feminist* as more likely does not violate any rules of probability. Dulany and Hilton [4] note that this implicature *bank teller and not feminist*, referred to as the K-implicature is one of two possible implicatures, the other being *bank teller, and she may or may not be a feminist* or the so-called P-implicature. As such, the K-Implicature is a scalar implicature, such that it maximizes the quantity of information described. The K-Implicature strengthens the basic meaning *bank teller* by virtue of contextual considerations, especially in comparison to the conjunction *bank teller and feminist*. The P-Implicature is the ignorance implicature, where it is inferred that since the speaker only asserted that Linda is a bank teller, that they are not opinionated about whether Linda is or is not a bank teller.

There have been many empirical attempts to block any resultant implicatures, yielding conflicting results (see [11, 14] for more detailed reviews). To briefly summarize some key experiments, in one of their earliest studies, instead of the alternative "Linda is a bank teller" Tversky and Kahneman [25, p.299] used "Linda is a bank teller whether or not she is in the feminist movement" to make the set of bank tellers inclusive to those who are also feminists more evident. This resulted in a reduction in the number of conjunction errors. Politzer and Noveck [18, p.90] similarly found a reduction in the conjunction error when the logically entailed implicature was separated from the basic meaning as in:

(2) -Roland took an exam.
-Roland failed an exam.
-Roland passed an exam.

Conversely, Agnoli and Krantz [2] compared a standard conjunction problem, an explicit implicature version (bank teller and not a feminist) and a standard conjunction problem with a blocked implicature version (bank teller and may or may not be a feminist). They predicted that there should be less conjunction errors when the implicature was blocked, yet they did not find a significant difference. More clearly, Tentori and colleagues [22] found that conjunction error persists even when including alternatives that block the implicature or make the implicature explicit. This is particularly surprising because it found using items phrased in a frequency format, which has been demonstrated to reduce the conjunction error. Tentori and colleagues'[22] findings were replicated by Wedell and Moro [26] using the same experimental items.

Dulany and Hilton [4] scrutinized the criterion for which researchers should classify choosing "Linda is a bank teller and feminist" as constituting a conjunction error. After participants completed the standard Linda Problem, The authors asked them how they interpreted "Linda is a bank teller", specifically (1) is not a feminist, (2) is probably a feminist, (3) probably not a feminist or (4) whether she is feminist or not. They only considered those who committed the conjunction error and chose (4). They deteremined that this was the appropriate criterion because those who interpret *bank teller* as (4) are verifiably comparing the set of bank tellers to the subset of bank tellers and feminist. Using this metric, the number of conjunction errors committed was significantly reduced [4, 11].

All the findings summarized here describe the potential inferences available to a listener when one event is compared to a conjunction. This prompts related questions: does the conjunction error persist when the conjunction is compared to a disjunction? What are the available inferences when a conjunction is compared to a disjunction (i.e., bank teller or feminist)? This is of interest because disjunctions are entailed by the conjunction in a similar fashion to how the set is entailed the subset in the original construction. Specifically, it is necessarily true that Linda is a bank teller **or** a feminist, if she is a bank teller **and** a feminist. The crucial difference is that disjunctions eliminate the confound present in the original construction of the Linda Problem, where only one of the descriptors (bank teller) as a set is compared to the subset of both descriptors (bank teller and feminist).

In addition, plain disjunctions (this or that) are ambiguous in natural language between an inclusive (this or that *or both*) or exclusive (this or that *but not both*) interpretation. Scalar implicatures (inferences) can resolve this ambiguity by denying the possibility of both events being true, resulting in its exclusive interpretation. Generally speaking, exclusive disjunctions are more informative, as they reduce the number of possible alternatives that could answer the QUD, compared to inclusive disjunctions and are thus preferred. However, in this scenario an inclusive interpretation (e.g., Linda is a bank teller or a feminist or both) would be preferable because it is more statistically likely, as it allows for any of the possibilities to be true. Thus, it would be of interest to determine the types of inferences that are preferable for a listener to make use of while judging the relative likelihoods of the sentences. The current study examined these circumstances.

3 Experiment

This study examined the effect of the representativeness heuristic relative to the influence of informativity. Here, informativity is related to the entailment, where if A entails B, it is more informative (e.g., conjunction is more informative than inclusive disjunction because it entails it). This study aimed to determine if individuals would still prefer the conjunctive answer when the presented with either plain disjunctions or "or both" disjunctions (those with cancelled implicatures). A pertinent question was how participants choose between an answer that is more representative of the context (e.g., Linda is a bank teller and a feminist) compared to one that is more informative to the QUD and therefore more probable (e.g., Linda is a bank teller or a feminist (or both)). Additionally, the experiment investigated the possible inferences available in this paradigm when a scalar implicature is available in a disjunction compared to one when it is explicitly cancelled.

3.1 Design and Materials

In order to investigate our hypotheses, participants responded to 10 items from 3 possible experimental conditions, as illustrated in Table 2. Items were counterbalanced using a Latin square method to ensure that each condition was adequately represented in the data. A willingness to bet paradigm was used as an alternative to asking for a probability judgment in order to avoid any possible misinterpretations of the term "probability".

3.2 Predictions

In condition 1, when participants adopt the basic meaning (without any strengthening), *bank teller and feminist* should be judged as less probable than *bank teller* as the conjunction entails one of conjuncts. If participants strengthen the meaning of *bank teller* to *bank teller and not feminist*, the entailment relationship is broken, and neither answer is more informative than the other, therefore it is up to the participants to determine which of the two is more likely. Given the description of Linda, this should bias individuals towards choosing *bank teller and feminist.*

When participants adopt the basic meaning in condition 2, *bank teller or feminist* should be interpreted as *bank teller or feminist or both.* Here, *bank teller and feminist* entails *bank teller or feminist or both* and *bank teller and feminist* should thus be judged as less probable. If participants strengthen the meaning of*bank teller or feminist* to *bank teller or feminist and not both*, the entailment is once again broken, the relative probabilities will be at the discretion of the participant. Therefore, one has to determine if bank teller or feminist and not both (with the uncertainty of not knowing if of the two descriptions is true) is more probable than Linda being both a feminist and a bank teller.

Table 2: Sample item for original construction and disjunctive sentences of the Linda problem

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
With the aim of getting getting $10 to a local children's charity, which of the following do you think is most likely to be true?

1. Original Construction
Linda is a bank teller (p).
Linda is a bank teller and a feminist (p *and* q).
Linda is a TV salesman.
Linda is a farmer.

2. Plain Disjunction
Linda is a bank teller or a feminist (p *or* q).
Linda is a bank teller and a feminist (p *and* q).
Linda is a TV salesman.
Linda is a farmer.

3. Or-both Disjunction
Linda is a bank teller or a feminist or both (p *or* q *or both*).
Linda is a bank teller and a feminist (p *and* q).
Linda is a TV salesman.
Linda is a farmer.

Condition 3 does not allow for any strengthening, and similarly to the basic interpretation for condition 2, *bank teller and feminist* entails *bank teller or feminist or both* and *bank teller and feminist* should be deemed less probable.

Taking stock, there is an ambiguity in conditions 1 and condition 2 depending on how (i) and individual interprets the problem and the given alternatives and (ii) how they judge the associated probabilities with each answer. Based on the design of the conjunction error, the number of choices for the conjunction *bank teller and feminist* can be expected to decrease with addition of weaker scalar terms in the relative comparisons.

3.3 Participants

211 participants participated as volunteers from a community sample or from the Carleton University undergraduate research pool for partial course credit (0.25%) in an introductory Cognitive Science course. Volunteers received an invitation to the study shared via social media and did not receive any compensation for their participation. The study was approved by the Carleton University Research Ethics Board. 31 participants reported that they had previously seen items similar to the conjunction errors presented in the study, or were already

familiar with the conjunction error and were removed from subsequent analysis for a total of 180 participants.

4 Results

Response times shorter than 15s and longer than 200s were identified as extreme outliers and removed from subsequent analysis. Responses chosen for the filler/distractor items (e.g., *Linda is a farmer*) were also removed from the analysis. Response times were analyzed using a linear mixed effects model using the "lmer" function from the "lme4" package in R [3]. The random effects structure included a random intercept for subjects. More complex random effects structures either overfit the data or failed to converge. "Condition" (with respect to Table 2) was the fixed effect. A likelihood ratio test between the full and reduced model (one that does not include "condition" as a factor) reveals no significant difference between conditions, $\chi^2(2) = 0.93$, $p = 0.63$.

Response choices were analyzed using a logistic mixed effects model using the "glmer" function from the "lme4" package from R [3]. Similarly to the response times, "condition" as a include in the model as a fixed effect, and the random effects structure used "subject" as a random intercept. A likelihood ratio test determine that there was a significant difference between conditions $\chi^2(2) = 98.95$, $p < 0.0001$, where participants were significantly more likely to make conjunction errors in condition (1.) compared to (2.), $\beta = 3.44$, $z = 7.34$, $p < 0.0001$ and (3.) respectively, $\beta = 3.27$,$z = 7.21$, $p < 0.0001$. There was no difference in conjunction errors committed between conditions (2.) and (3.), $\beta = -0.17$, $z = -0.54$, $p = 0.85$. The p-values for the simple effects were adjusted for multiple comparison using the Tukey method. Figure 1 presents these results below:

4.1 Individual differences in response strategy

An exploratory analysis of the dataset was used to further probe the possible derivable inferences and response strategies that individuals could make use of in both types of disjunctions. In order to examine this, participants were grouped based on their relative likelihood of choosing the alternative sentence in those conditions. Here, participants were categorized based on whether they preferred the answer *p and q* (e.g., bank teller and feminist) or if they preferred *p or q* in condition 2 or *p or q or both* in condition 3. Mean participant response times shorter than 5s were identified as outliers based on the spread of the data and were trimmed from the analysis, resulting in a total of 114 participants. Table 3 summarizes these findings.

Table 3 suggests that participants adopted distinct response strategies. Starting with the 45 participants who chose *p and q* in condition 2 and *p or q or both* in condition 3, it can be inferred that these individuals strengthened the sentence in condition 2 and found *p and q* to be more likely. If they did not strengthen

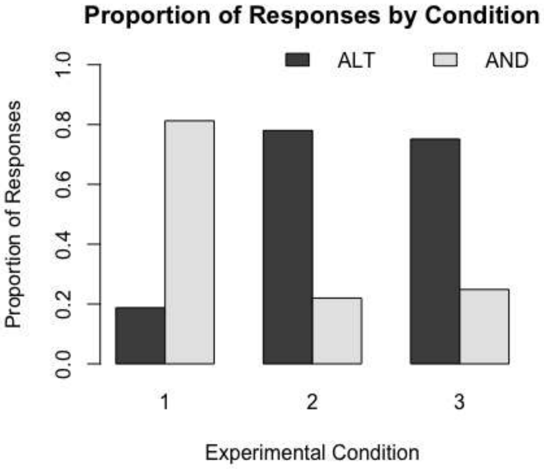


Fig. 1: Proportions of response choices for the conjunctive sentence or the alternative in each condition. The alternative in condition 1 was one of the two descriptors in the conjunction *p*, *p or q* in condition 2 and *p or q or both* in condition 3.

Table 3: Number of participants by response choices in conditions 2 and 3.

	Condition 3	
Condition 2	*p and q*	*p or q or both*
p and q	2	45
p or q	25	42

the meaning (i.e., interpreting *p or q* ⇒ *p or q or both*, they ought to choose *p or q* as the more probable answer, as it encapsulates all the possibilities associated with the asserted descriptors. Conversely, the 42 participants who chose *p or q* in condition 2 and *p or q or both* in condition 3 whether they chose to strengthen the meaning or not found *p or q (or both)* more likely than *p and.* Given that they chose *p or q or both* as more likely in condition 3, perhaps it may be speculated that they more are likely to interpret the disjunction in condition 3 literally. It is less clear what the strategy was of those participants who were in the remaining two cells in terms of strengthening. Given the small number of participants that chose *p and q* in both conditions 2 and 3 (2 participants), this may not reflect a true response strategy. An alternative explanation may be that those participants were not interpreting probability in the statistical sense. Such that, it could be the case that those participants found the conjunction the most likely answer in all conditions, whether they strengthened the meaning or not, perhaps based on representativeness. Likewise for those 25 participants who chose *p or q* in condition 2 and *p and q* in condition 3, again, whether they strengthened the meaning or not, they found *p or q* more likely than *p and q*. Then for condition 3, it is possible that they deemed the conjunction as being more probable as a full answer to the QUD, compared to the inclusive disjunction which allows for any possibility without providing any definitive indications to which one is correct. A follow-up to this secondary analysis will aim to investigate the strength of the preference within each response strategies across all of the experimental items, particularly, whether participants regularly chose one sentence over the other.

5 Conclusion

The current study successfully replicated the conjunction error in its original form, as participants were more likely to choose the conjunction of the two descriptors compared to one of the descriptors alone. Additionally, the study significantly reduces the number of conjunction errors when conjunctive sentence is presented in comparison to both plain and inclusive disjunctive sentences. Looking further to individual subject behaviours revealed that participants adopt distinct response strategies in the plain disjunction, namely whether they would exhaustify the plain disjunction or the adopt the basic meaning.

The response strategies are also revealing of the possible epistemic states of the participants and how they can make use of these beliefs when judge the probabilities associated with each outcome, similar to the experthood assumption that is proposed to be required for scalar implicature [20]. While this current study doesn't provide the level of granularity to say for certain when participants are strengthening their responses and the processes involved therein, it does suggest certain considerations that participants make in choosing their responses. Specifically, participants have to decide the likelihood of each descriptor being true as they are presented (i.e., as a conjunction or a disjunction), particularly when the description is only representative of one of the combined descriptions. This can be further compounded by how participants interpret the meaning of probability, as either representative or a description or based on statistical likelihood. Here, implicature appears to be an influence on the associated probabilities of the descriptions, although this effect is difficult to dissociate from other inferences individuals make. Although the prior research on the frequency format is met with mixed results, future studies should adapt the present study in a frequency format in order to encourage the statistical interpretation of probability. Another approach would be to consider Bayesian probabilistic reasoning, as well as the rational speech act view for both deriving implicatures and computing probabilistic judgments.

Generally, this study provides interesting results in terms of successfully reducing the conjunction error, while maintaining an entailment relationship. It also raises some relevant questions on an individual's intuitions about the component and combined probability of events under uncertainty.

6 Acknowledgments

I am deeply grateful for my supervisor, Dr. Raj Singh, this manuscript would be far worse off without his patience, guidance and careful consideration. I would additionally like to extend my thanks to the anonymous reviewers for the helpful comments made to improve the quality of this submission. All remaining errors are my own.

References

1. Adler, J.E.: Abstraction is uncooperative. Journal for the Theory of Social Behaviour **14**(2), 165–181 (1984)
2. Agnoli, F., Krantz, D.H.: Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. Cognitive Psychology **21**(4), 515–550 (1989)
3. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. Journal of Statistical Software **67**(1), 1–48 (2015). https://doi.org/10.18637/jss.v067.i01
4. Dulany, D.E., Hilton, D.J.: Conversational implicature, conscious representation, and the conjunction fallacy. Social Cognition **9**(1), 85–110 (1991)
5. Fiedler, K.: The dependence of the conjunction fallacy on subtle linguistic factors. Psychological research **50**(2), 123–129 (1988)
6. Gigerenzer, G.: Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In: Subjective probability, pp. 129–161. Wiley (1994)
7. Grice, H.P.: Logic and Conversation. In: Cole, P., Morgan, J.L. (eds.) Syntax and Semantics: Vol. 3: Speech Acts, pp. 41–58. Academic Press, New York (1975)
8. Grice, H.P.: Studies in the Way of Words. Harvard University Press (1989)
9. Groenendijk, J.A.G., Stokhof, M.J.B.: Studies on the Semantics of Questions and the Pragmatics of Answers. Ph.D. thesis, Univ. Amsterdam (1984)
10. Hertwig, R., Benz, B., Krauss, S.: The conjunction fallacy and the many meanings of and. Cognition **108**(3), 740–753 (2008)
11. Hertwig, R., Gigerenzer, G.: The 'conjunction fallacy'revisited: How intelligent inferences look like reasoning errors. Journal of behavioral decision making **12**(4), 275–305 (1999)
12. Maguire, P., Moser, P., Maguire, R., Keane, M.T.: Why the conjunction effect is rarely a fallacy: How learning influences uncertainty and the conjunction rule. Frontiers in psychology **9**, 1011 (2018)
13. Mellers, B., Hertwig, R., Kahneman, D.: Do frequency representations eliminate conjunction effects? an exercise in adversarial collaboration. Psychological Science **12**(4), 269–275 (2001)
14. Moro, R.: On the nature of the conjunction fallacy. Synthese **171**(1), 1–24 (2009)
15. Mosconi, G., Macchi, L.: The role of pragmatic rules in the conjunction fallacy. Mind & Society **2**(1), 31–57 (2001)
16. Nisbett, R.E., Krantz, D.H., Jepson, C., Kunda, Z.: The use of statistical heuristics in everyday inductive reasoning. Psychological review **90**(4), 339 (1983)
17. Partee, B., Rooth, M.: Generalized conjunction and type ambiguity. Formal semantics: the essential readings pp. 334–356 (1983)
18. Politzer, G., Noveck, I.A.: Are conjunction rule violations the result of conversational rule violations? Journal of psycholinguistic research **20**(2), 83–103 (1991)

19. van Rooy, R.: Conversational implicatures and communication theory. In: Current and new directions in discourse and dialogue, pp. 283–303. Springer (2003)
20. Sauerland, U.: The epistemic step. Experimental Pragmatics **10** (2005)
21. Stolarz-Fantino, S., Fantino, E., Zizzo, D.J., Wen, J.: The conjunction effect: New evidence for robustness. The American Journal of Psychology (2003)
22. Tentori, K., Bonini, N., Osherson, D.: The conjunction fallacy: a misunderstanding about conjunction? Cognitive Science **28**(3), 467–477 (2004)
23. Tentori, K., Crupi, V., Russo, S.: On the determinants of the conjunction fallacy: Probability versus inductive confirmation. Journal of Experimental Psychology: General **142**(1), 235 (2013)
24. Tversky, A., Kahneman, D.: Judgment under uncertainty: Heuristics and biases. science **185**(4157), 1124–1131 (1974)
25. Tversky, A., Kahneman, D.: Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological review **90**(4), 293–315 (10 1983)
26. Wedell, D.H., Moro, R.: Testing boundary conditions for the conjunction fallacy: Effects of response mode, conceptual focus, and problem type. Cognition **107**(1), 105–136 (2008)

Interaction between Intonational and Syntactic Markers of Information Structure in Mandarin

Jing Ji

Stony Brook University, Stony Brook NY 11790, USA jing.ji@stonybrook.edu

Abstract. This article applies CCG analysis to investigating the syntax-prosody interface of information structure in Mandarin. Compared with English, Mandarin makes use of both syntactic and intonational markers to convey discourse semantics. The complementary relation of pitch accent and focus projection marker 'shi...de' provides a fine-grained way of dealing with theme/rheme ambiguity, while the competition between boundary tones and discourse particles reflects the distinct competence scope of different particles.

Keywords: Syntax-prosody interface · Information structure · Mandarin.

1 Introduction

Combinatory Categorial Grammar(CCG) is a form of lexicalized grammar,[6] which assumes that logical forms can be derived directly from the surface-syntactic derivation. Steedman (2014) [5] extended this framework to discourse semantics, showing that in English, information structural distinctions are mainly conveyed by intonational prosody. However, he also mentioned that there is great cross-linguistic variation in the way the semantic distinctions are marked by grammatical devices. Since Mandarin is a tone language, intonational signals alone might not be sufficient to mark distinctions. Therefore, grammatical markers could be utilized supplementally. This article investigated syntactic and prosodic markers of discourse semantics in Mandarin, showing that compared with English, the information structure of Mandarin is conveyed by syntactic construction, discourse particles, as well as prosody, and the interaction between different signals reflects the linguistic economy and their relative significance.

According to Steedman [5], information structure is represented by semantic primitives in four dimensions, including the presence of contrast, theme/rheme, common ground realization, speaker/hearer agency. They are signaled by elementary abstract tones of Autosegmental-Metrical theory in English [1, 3], i.e., the first three dimensions are associated with the pitch accent, while the last dimension is represented by boundary tones. Specifically, contrast is indicated by the presence of pitch accent, whereas theme and rheme, common ground realization are signaled by type of pitch accent. Their relationship is summarized in Table 1 and 2.

Table 1. pitch signals of theme/rheme and success/failure

	success($\top$)	failure($\perp$)
theme(θ)	L+H*	L*+H
rheme(ρ)	H*	L*

Table 2. boundary tone of speaker/hearer agency

semantic primitive	boundary tone
Speaker(S)	LL% HL%
Hearer(H)	HH% LH%

As is shown in the table, in English, the theme is associated with bitonic pitch accents while rheme is conveyed by monotonic ones. The success of updating common ground is signaled by high tone whereas the failure is reflected in a low tone. In terms of boundary tone, falling tone and rising tone indicate speaker agency and hearer agency respectively. The following examples illustrate how to combine intonational signals to interpret information structure. In example (1), the pitch accent H^* represents rheme (ρ) and success ($\top$) of updating common ground and LL% indicates speaker agency (S). Thus the speaker's response has the meaning of 'I make that (raining) common ground successfully', which is in contrast with example (b) 'You do not suppose that (raining) to be common ground'.

(1) H: It's raining .
S: MMMM !
H* LL%

$\top$(ρ raining S)

(2) H: It's raining .
S: MMMM !
L*+H LH%

$\perp$(θ raining H)

The following two sections make a comparison of English and Mandarin markers of discourse semantics, showing that different from English, where information structural distinctions are conveyed by intonational prosody to an extreme degree, in Mandarin, however, intonational markers of information structure can be supplemented by syntactic construction and discourse particles. The interaction between syntactic and prosodic markers is illustrated with complementary and competing conditions, which are the interaction between syntactic construction and pitch accents and the interaction between discourse particles and boundary tones, respectively.

2 The Interaction between Syntactic Construction and Pitch Accent

The theme-rheme contrast in English is signaled by a bitonal/monotonal pitch accent, which is illustrated in the following minimal pair of dialogues.

(3)	Q: Who	married Bill?		
	A: Anna	married	Bill	.
	H*		L+H*	LH%

(4)	Q: Who	did	Anna	marry?	
	A: Anna	married	Bill	.	
	L+H*		H*	LH%	

In example (3), the subject 'Anna' is the rheme signaled by monotonal pitch accent H*, while in example (4), 'Bill' is rheme. The intonation signaled semantic difference results in their distinct derivations, which are shown in (5) and (6) below. Example (5) can be interpreted as 'You suppose the question of who married Bill (as opposed to anyone else) to be common ground. I make it common ground that it was Anna (as opposed to anyone else).' In contrast, the interpretation of example (6) is 'You suppose the question of who did Anna (as opposed to anyone else) marry to be common ground. I make it common ground that it was Bill (as opposed to anyone else).'

(5)

$$
\begin{array}{cccc}
\mathit{Anna} & \mathit{married} & \mathit{Bill} & \\
H^* & & L+H^* & L\dot{L}\% \\
\hline
S_{\top,\rho}/(S_{\top,\rho}\backslash NP_{\top,\rho})\ {\scriptstyle >\mathbf{T}} & (S\backslash NP)/NP & (S_{\top,\theta}\backslash NP_{\top,\theta})\backslash(S_{\top,\theta}\backslash NP_{\top,\theta})\ {\scriptstyle <\mathbf{T}} & S\$_{\phi}\backslash S\$_{\pi,\eta} \\
:\left\{\begin{array}{l}\lambda p.p\ \mathit{anna}\\ \lambda p.p\ v_{\tau_{anna}}\end{array}\right\} & :\lambda x.\lambda y.\mathit{married}\ xy & :\left\{\begin{array}{l}\lambda p.p\ \mathit{bill}\\ \lambda p.p\ v_{\tau_{bill}}\end{array}\right\} & :\{\lambda f.\pi(\eta f S)\} \\
\cline{1-1}\cline{2-3}
S_{\phi}/(S_{\phi}\backslash NP_{\phi})\ {\scriptstyle <\%} & \multicolumn{2}{c}{S_{\top,\theta}\backslash NP_{\top,\theta}\ {\scriptstyle <}} & \\
:\top(\rho\left\{\begin{array}{l}\lambda p.p\ \mathit{anna}\\ \lambda p.p\ v_{\tau_{anna}}\end{array}\right\}S) & \multicolumn{2}{c}{:\lambda x.\mathit{married\ bill}\ x} & \\
\cline{2-4}
 & \multicolumn{3}{c}{S_{\phi}\backslash NP_{\phi} : \top(\theta\{\lambda x.\mathit{married\ bill}\ x\}S)\ {\scriptstyle <}} \\
\hline
\multicolumn{4}{c}{S : \left\{\begin{array}{l}\mathit{married\ bill\ anna}\\ \mathit{married}\ v_{\tau_{bill}}\ v_{\tau_{anna}}\end{array}\right\}\ {\scriptstyle >}}
\end{array}
$$

(6)

$$
\begin{array}{ccccc}
\mathit{Anna} & \mathit{married} & & \mathit{Bill} & \\
L+H^* & & LH\% & H^* & L\dot{L}\% \\
\hline
S_{\top,\theta}/(S_{\top,\theta}\backslash NP_{\top,\theta})\ {\scriptstyle >\mathbf{T}} & (S\backslash NP)/NP & S\$_{\phi}\backslash S\$_{\pi,\eta} & S_{\top,\rho}\backslash(S_{\top,\rho}/NP_{\top,\rho})\ {\scriptstyle <\mathbf{T}} & S\$_{\phi}\backslash S\$_{\pi,\eta} \\
:\left\{\begin{array}{l}\lambda p.p\ \mathit{anna}\\ \lambda p.p\ v_{\tau_{anna}}\end{array}\right\} & :\lambda x.\lambda y.\mathit{married}\ xy & :\lambda f.\pi(\eta f H) & :\left\{\begin{array}{l}\lambda p.p\ \mathit{bill}\\ \lambda p.p\ v_{\tau_{bill}}\end{array}\right\} & :\{\lambda f.\pi(\eta f S)\} \\
\cline{1-2}
\multicolumn{2}{c}{S_{\top,\theta}/NP_{\top,\theta} : \left\{\begin{array}{l}\lambda x.\mathit{married}\ x\ \mathit{anna}\\ \lambda x.\mathit{married}\ x\ v_{\tau_{anna}}\end{array}\right\}\ {\scriptstyle >\mathbf{B}}} & & & \\
\cline{1-3}\cline{4-5}
\multicolumn{3}{c}{S_{\phi}/NP_{\phi} : \top(\theta\left\{\begin{array}{l}\lambda p.p\ \mathit{anna}\\ \lambda p.p\ v_{\tau_{anna}}\end{array}\right\}H)\ {\scriptstyle <}} & \multicolumn{2}{c}{S_{\phi}\backslash(S_{\phi}/NP_{\phi}) : \top(\rho\left\{\begin{array}{l}\lambda p.p\ \mathit{bill}\\ \lambda p.p\ v_{\tau_{bill}}\end{array}\right\}S)\ {\scriptstyle <}} \\
\hline
\multicolumn{5}{c}{S : \left\{\begin{array}{l}\mathit{married\ bill\ anna}\\ \mathit{married}\ v_{\tau_{bill}}\ v_{\tau_{anna}}\end{array}\right\}\ {\scriptstyle <}}
\end{array}
$$

However, intonation in English has limitations in indicating theme/rheme. For instance, sentences with unmarked themes lack an accent, which can result in ambiguity concerning the information-structural division into theme and rheme. According to Selkirk(1995) [4], accented words are F-marked and the F-marking of the head of a phrase or an internal argument of the head licenses the F-marking of the entire phrase. Therefore, the sentence is ambiguous between a structure with object-NP focus and VP focus. In the following examples, the same sentence with H^* on ‘Bill’ can be the answer to different questions with various phrasing.

(7) Q: What will Anna do?
A: Anna will (marry Bill) .
H^* LL%

(8) Q: What about Anna?
A: Anna (will marry Bill) .
H^* LL%

(9) Q: What’s new?
A: Anna will marry Bill .
H^* LL%

Compared with English, Mandarin makes use of both syntactic and intonational markers to signal theme/rheme. The role of pitch accent can be partially substituted by syntactic construction. A typical syntactic device is the construction ‘shi...de’. ‘shi’ is a modal adverb and ‘de’ serves as the complementizer of a relative clause. The well-formedness of a sentence would not be affected when omitting ‘shi...de’. While ‘de’ occurs after the main verb or VP, ‘shi’ is used before the emphasized component in a sentence. Therefore, ‘shi’ is traditionally supposed to be a focus particle used to update common ground [2]. The emphasized component can be an internal argument of the main verb, such as subject and object, or peripheral argument, just as time and manner. Therefore, ‘shi’ indicates rheme in a similar way as pitch accent H^* . As the counterpart of (5) and (6), the following examples illustrate the similar function of ‘shi...de’ indicating theme/rheme with CCG derivations. It can be seen from (10) and (11) that the argument of the verb immediately after ‘shi’ is denoted as rheme while other arguments are taken as the theme.

(10)

$$\begin{array}{ccccc}
\textit{shi Anna} & \textit{zuotian} & \textit{gen Bill} & \textit{jiehun} & \textit{de LL\%} \\
\textit{shi Anna} & \textit{yesterday} & \textit{with Bill} & \textit{married} & \textit{de LL\%} \\
\hline
{>}\mathsf{T} & & {>}\mathsf{T} & & \\
S_{\top,\rho}/(S_{\top,\rho}\backslash NP_{\top,\rho}) & (S_{\top,\theta}\backslash NP_{\top,\theta})/(S_{\top,\theta}\backslash NP_{\top,\theta}) & PP^{\uparrow}_{\top,\theta} & (S\backslash NP)\backslash PP & S\$_{\phi}\backslash S\$_{\pi,\eta} \\
: \left\{\begin{array}{l} \lambda p.p\ anna \\ \lambda p.p\ v_{\tau_{anna}} \end{array}\right\} & : [\lambda p\lambda y.p^{0}\ y]^{yesterday} & : \left\{\begin{array}{l} \lambda p.p\ bill \\ \lambda p.p\ v_{\tau_{bill}} \end{array}\right\} & : \lambda x.\lambda y.marry\ x\ y & \lambda f.\pi(\eta f S)
\end{array}$$

$$\overline{\qquad\qquad}{<}\%\quad S_{\phi}/(S_{\phi}\backslash NP_{\phi}) : \top(\rho \left\{\begin{array}{l} \lambda p.p\ anna \\ \lambda p.p\ v_{\tau_{anna}} \end{array}\right\} S)$$

$$\overline{\qquad\qquad}{>}\quad S_{\top,\theta}\backslash NP_{\top,\theta} \left\{\begin{array}{l} \lambda y.marry\ bill\ y \\ \lambda y.marry\ v_{\tau_{bill}}\ y \end{array}\right\}$$

$$\overline{\qquad\qquad}{>}\quad S_{\top,\theta}\backslash NP_{\top,\theta} \left\{\begin{array}{l} \lambda y.marry\ bill\ y \\ \lambda y.marry\ v_{\tau_{bill}}\ y \end{array}\right\}^{yesterday}$$

$$\overline{\qquad\qquad}{<}\quad S_{\phi}\backslash NP_{\phi} \quad \top(\theta \left\{\begin{array}{l} \lambda y.marry\ bill\ y \\ \lambda y.marry\ v_{\tau_{bill}}\ y \end{array}\right\}^{yesterday} S)$$

$$\overline{\qquad\qquad}{>}\quad S : \left\{\begin{array}{l} married\ bill\ anna \\ married\ v_{\tau_{bill}}\ v_{\tau_{anna}} \end{array}\right\}^{yesterday}$$

(11)

$$\begin{array}{cccccc}
\textit{Anna} & \textit{zuotian} & \textit{LH\%} & \textit{shi gen Bill} & \textit{jiehun} & \textit{de LL\%} \\
\textit{Anna} & \textit{yesterday} & \textit{LH\%} & \textit{shi with Bill} & \textit{married} & \textit{de LL\%} \\
\hline
{>}\mathsf{T} & & & {>}\mathsf{T} & & \\
S_{\top,\theta}/(S_{\top,\theta}\backslash NP_{\top,\theta}) & NP^{\uparrow}\backslash NP^{\uparrow} & S\$_{\phi}\backslash S\$_{\pi,\eta} & PP^{\uparrow}_{\top,\rho} & (S\backslash NP)\backslash PP & S\$_{\phi}\backslash S\$_{\pi,\eta} \\
: \left\{\begin{array}{l} \lambda p.p\ anna \\ \lambda p.p\ v_{\tau_{anna}} \end{array}\right\} & : [\lambda np\lambda p.np^{0}\ p]^{yesterday} & \lambda f.\pi(\eta f S) & : \left\{\begin{array}{l} \lambda p.p\ bill \\ \lambda p.p\ v_{\tau_{bill}} \end{array}\right\} & : \lambda x.\lambda y.marry\ x\ y & \lambda f.\pi(\eta f S)
\end{array}$$

$$\overline{\qquad\qquad}{<}\quad S_{\top,\theta}/(S_{\top,\theta}\backslash NP_{\top,\theta}) : \left\{\begin{array}{l} \lambda p.p\ anna \\ \lambda p.p\ v_{\tau_{anna}} \end{array}\right\}^{yesterday}$$

$$\overline{\qquad\qquad}{>}\quad S_{\top,\rho}\backslash NP_{\top,\rho} \left\{\begin{array}{l} \lambda y.marry\ bill\ y \\ \lambda y.marry\ v_{\tau_{bill}}\ y \end{array}\right\}$$

$$\overline{\qquad\qquad}{<}\quad S_{\phi}/(S_{\phi}\backslash NP_{\phi}) \quad \top(\theta : \left\{\begin{array}{l} \lambda p.p\ anna \\ \lambda p.p\ v_{\tau_{anna}} \end{array}\right\}^{yesterday} H)$$

$$\overline{\qquad\qquad}{<}\quad S_{\phi}\backslash NP_{\phi} \quad \top(\rho : \left\{\begin{array}{l} \lambda p.p\ anna \\ \lambda p.p\ v_{\tau_{anna}} \end{array}\right\} S)$$

$$\overline{\qquad\qquad}{>}\quad S_{\phi} \quad S : \left\{\begin{array}{l} married\ bill\ anna \\ married\ v_{\tau_{bill}}\ v_{\tau_{anna}} \end{array}\right\}^{yesterday}$$

Similar to the limitation of pitch accent, ‘shi...de’ can also result in ambiguity. Since ‘shi’ can be used either before a certain argument (narrow focus) or before the whole predicate (broad focus), sentences with ‘shi’ in front of the predicate is ambiguous. However, Mandarin has a way of dealing with ambiguity by combining syntactic construction and pitch accents. As is shown in the following examples, when ‘shi’ is used before the whole predicate containing more than one argument, the emphasized argument should bear a pitch accent.

(12) Q: *Anna natian gen Bill jiehun*?
Anna when with Bill marry
‘When did Anna marry Bill?’

A: *Anna shi zuotian gen Bill jiehun de.*
Anna FP yesterday with Bill marry REL
H*
‘It was yesterday that Anna married Bill.

(13) Q: *Anna gen shei jiehun?*
Anna with whom marry

‘Who did Anna marry?’

A: *Anna shi zuotian gen Bill jiehun de.*
Anna FP yesterday with Bill marry REL
H*

‘It was Bill that Anna married yesterday.

(14) Q: *Anna zenmeyang?*
Anna how

‘What about Anna?’

A: *Anna shi hui gen Bill jiehun de.*
Anna FP will with Bill marry REL

‘Anna will marry Bill’

The flexible position of ‘shi’ shows that instead of marking focus, ‘shi’ is used to mark the focus projection boundary. There are two pieces of evidence in support of this argument. First, according to focus projection theory, an F-marked constituent that is not a focus is interpreted as new [4]. If the question in example (13) is ‘Who did Anna marry yesterday?’, ‘shi’ can not occur before ‘zuotian’ (yesterday). It is because ‘yesterday’ is neither the focus nor the new information, which can only come before the focus projection boundary marked by ‘shi’.

Second, if ‘shi’ marks narrow focus instead, there would be a mismatch between syntactic and intonational markers in the answer of example (13) since ‘zuotian’ (yesterday) after ‘shi’ and ‘Bill’ with pitch accent can both serve as focus, even though the component after ‘shi’ is not as prominent as ‘Bill’ with nuclear pitch accent [7].

It can be seen that ‘shi...de’ can only be partially regarded as a theme/rheme indicator, which is further evidenced by ‘shi’ used before the subject of a sentence. In the case of the complex subject shown in example (15), positional constraint of ‘shi...de’ is reflected since ‘shi’ can only occur before the subject of the main clause rather than coming immediately before the focus ‘Marry’.

(15) *shi Bill shanghai Marry de yaoyan rang ta nanguo de.*
FP Bill hurt Marry REL rumor made him upset REL
H*

‘The rumor that Bill hurt Marry (as opposed to anyone else) made him upset.’

* *Bill shanghai shi Marry de yaoyan rang ta nanguo de.*

Given the limitation of both syntactic construction and pitch accent, Mandarin combines them in a complementary relationship to indicate theme-rheme contrast.

3 Interaction between Discourse Particles and Boundary Tone

In English, the speaker-hearer agency is signaled by boundary tones, with falling tone standing for speaker agency and rising tone for hearer agency. Therefore, different intentions and speech acts are reflected by the combination of pitch accent and boundary tones, as is illustrated in the following examples. Example (13) is a declaration, indicating that ‘I noticed you did that’. Although both (14) and (15) are questions, example (15) is less aggressive since it does not call for consistency maintenance activity, whereas example (14) requires the hearer to do so by claiming that the hearer fails to make a supposition common ground.

(16) You put my trousers in the microwave !
H* H* LL%

$$\top(\rho\left\{\begin{array}{l} put(in\ microwave)trousers\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{trousers}}\ \ H \end{array}\right\} S)$$

‘I make it common ground that you put my trousers in the microwave.’

(17) You put my trousers in the microwave ?
L* L* LH%

$$\perp(\rho\left\{\begin{array}{l} put(in\ microwave)trousers\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{trousers}}\ \ H \end{array}\right\} H)$$

‘You do not make it common ground that you put my trousers in the microwave.’

(18) You put my trousers in the microwave ?
H* H* LH%

$$\top(\rho\left\{\begin{array}{l} put(in\ microwave)trousers\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{trousers}}\ \ H \end{array}\right\} H)$$

‘You make it common ground that you put my trousers in the microwave.’

As the counterpart of the intonational markers, discourse particles in Mandarin have a similar function given the following reasons. First, the particle used at the end of a Mandarin sentence can signal sentence type. For instance, the sentence ending with ‘ma’ or ‘ne’ is interrogative while sentence ending with ‘le’ is declarative. Besides, they also convey speakers’ attitudes and supposition towards the utterance, such as ‘a’ indicating exclamation and ‘ba’ representing speculation. Therefore, they can be associated with the speaker or hearer agency similarly as boundary tones.

Analog to the patterns in (16)-(18) represented by boundary tone and pitch accent, in the following examples, the sentence ‘ni ba ta fang zai weibolu le’ with a default boundary tone LL% for (19) and LH% for (20)-(21) could express corresponding meanings when followed by different discourse particles.

(19) *ni ba ta fang zai weibolu le a* !
you ba it put in microwave ASP EXC
LL%

'You put it in the microwave!'

$$\top(\rho \left\{ \begin{matrix} put(in\ microwave)it\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{it}}\ H \end{matrix} \right\} S)$$

'I make it common ground that you put it in the microwave.'

(20) *ni ba ta fang zai weibolu le ma* ?
you ba it put in microwave ASP QUE
LH%

'Did you put it in the microwave?'

$$\perp(\rho \left\{ \begin{matrix} put(in\ microwave)it\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{it}}\ H \end{matrix} \right\} H)$$

'You do not make it common ground that you put it in the microwave.'

(21) *ni ba ta fang zai weibolu le ba* ?
you ba it put in microwave ASP QUE
LH%

'You put it in the microwave?'

$$\top(\rho \left\{ \begin{matrix} put(in\ microwave)it\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{it}}\ H \end{matrix} \right\} H)$$

'You make it common ground that you put it in the microwave.'

Different from the complementary relation between pitch accent and syntactic construction 'shi...de', there exists competition between boundary tones and discourse particles because although they convey similar meanings, boundary tone is obligatory in a sentence, which might result in redundant information. If it is true, the relative significance of different markers can be revealed by mismatched cases since the meaning of strong markers can be maintained when there is a conflict in between. Examples (22)-(24) are formed by replacing the default boundary tone of (19)-(21) with an alternative boundary tone.

(22) *ni ba ta fang zai weibolu le a* ?
you ba it put in microwave ASP EXC
LH%

'Did you put it in the microwave?'

$$\perp(\rho \left\{ \begin{matrix} put(in\ microwave)it\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{it}}\ H \end{matrix} \right\} H)$$

'You do not make it common ground that you put it in the microwave.'

(23) *ni ba ta fang zai weibolu le ma* .
you ba it put in microwave ASP QUE
LL%

'You didn't put it in the microwave, did you?'

$$\perp (\rho \left\{ \begin{array}{l} put(in\ microwave)it\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{it}}\ H \end{array} \right\} S)$$

'I do not make it common ground that you put it in the microwave.'

(24) *ni ba ta fang zai weibolu le ba* .
you ba it put in microwave ASP QUE
LL%

'You put it in the microwave?'

$$\top(\rho \left\{ \begin{array}{l} put(in\ microwave)it\ H \\ put(in\ v_{\tau_{microwave}})v_{\tau_{it}}\ H \end{array} \right\} H)$$

'You make it common ground that you put it in the microwave.'

Sentences of (22)-(24) differ from (19)-(21) in both speaker/hearer agency and success/failure of common ground realization when the default boundary tones were replaced. Comparing the meaning of corresponding sentence pairs, we can see that both (19) and (20) changed the meaning with the replacement of boundary tone, whereas the meaning the (21) remains unchanged irrespective of boundary tone. Specifically, sentences ending with 'a' with the default and alternative boundary tone differ in both dimensions, while sentences ending with 'ma' only differ in the speaker/hearer agency. It turns out the 'a' plays no role in either realization of common ground or speaker-hearer agency, while 'ma' serves as the indicator of failure in updating common ground. In contrast, the sentence ending with 'ba' is associated with hearer agency and success in updating common ground regardless of boundary tones.

Therefore, the strength of marking information structure among three discourse particles can be ranked as 'ba' > 'ma' > 'a'.

4 Conclusion

In conclusion, this paper applies surface-compositional semantics of English intonation to Mandarin discourse semantics. From the analysis above, it can be seen that Mandarin makes use of both syntactic and intonational markers of information structure. As the indicator of theme-rheme contrast, 'shi...de' can be used to indicate the focus projection domain in the main clause. The combination of syntactic and intonational theme-rheme markers provides a fine-grained way of dealing with the ambiguity of unmarked-theme in English. The competition between discourse particles and intonation in indicating the realization of common ground and the speaker-hearer agency reflects the relative significance of discourse markers. While ‘ba’ is an indicator of both hearer agency and successful realization of common ground, ‘ma’ representing the failure of common ground realization, ‘a’ lacks independent role in information structure.

References

1. Liberman, M.Y.: The intonational system of English. Ph.D. thesis, Massachusetts Institute of Technology (1975)
2. Paul, W., Whitman, J.: Shi... de focus clefts in mandarin chinese. The Linguistic Review **25**(3-4), 413–451 (2008)
3. Pierrehumbert, J.B.: The phonology and phonetics of English intonation. Ph.D. thesis, Massachusetts Institute of Technology (1980)
4. Selkirk, E.: Sentence prosody: Intonation, stress, and phrasing. The handbook of phonological theory **1**, 550–569 (1995)
5. Steedman, M.: The surface-compositional semantics of english intonation. Language **90**(1), 2–57 (2014)
6. Steedman, M., Baldridge, J.: Combinatory categorial grammar. Non-Transformational Syntax: Formal and explicit models of grammar pp. 181–224 (2011)
7. Welby, P.: Effects of pitch accent position, type, and status on focus projection. Language and Speech **46**(1), 53–81 (2003)

Reducing Homogeneity to Distributivity[⋆]

Alexandros Kalomoiros

University of Pennsylvania, Philadelphia PA 19104, USA
akalom@sas.upenn.edu

Abstract. This paper examines the semantics of homogeneity, being specifically concerned with the question whether or not homogeneity can be reduced to distributivity. Recent influential accounts of homogeneity in Križ 2015, 2019 have argued that such a reduction is not possible, as there are collective predicates that show homogeneity. We argue that in fact the empirical landscape is more complicated: while true that some collective predicates show homogeneity, not all collective predicates have this property. Collective activities and accomplishments show homogeneity, whereas collective states and achievements do not. Interestingly, collective accomplishments and activities have been analyzed as being able to host a D operator in their structure, while this is not possible for collective states and achievements (Brisson 2003). Therefore, once we control for the aktionsart of a collective predicate, it emerges that the collective predicates that allow homogeneity are exactly those that allow distributivity. We therefore conclude that we can reduce homogeneity to distributivity.

Keywords: Homogeneity · Distributivity · Collective Predicates · Aktionsart.

1 Introduction

Definite plurals require the verbal predicate to hold either of all or none of the parts of the plural individual they denote (Fodor 1970, Schwarzschild 1994, Križ 2015 a.o.):

(1) a. The knights died in battle.
 b. The knights did not die in battle.

Assume a model where there are 5 knights. (1-a) is True iff **all** of the five knights died in battle. (1-b) is true iff **none** of the five knights died in battle. In a mixed situation where three knights did in battle and two knights did not,

[⋆] Many thanks to Florian Schwarz for his generous help and encouragement with this project, as well as his detailed comments on the ideas presented in this paper. Thanks also the members of the semantics lab at Penn for helpful feedback on the ideas presented in this paper. Special thanks to Andrea Beltrama for his tremendous help in reading and commenting on the structure of multiple versions of an abstract based on this paper. All errors are my own.

the sentences have been claimed to be neither true not false; rather they are undefined (modulo non-maximality, which we leave aside in the present paper). This property of definite plurals is called homogeneity.

The rest of this paper is organised as follows: Section 2 briefly reviews previous approaches to homogeneity. Section 3 attempts to probe the empirical landscape of which collective predicates license homogeneity and which do not. Section 4 connects the results of section 3 to work in Brisson 2003, and section 5 uses the tools developed in Brisson's work to develop an analysis where homogeneity arises via distributivity. Section 6 examines the extent to which our account captures various cases of homogeneity. Section 7 concludes.

2 Accounts of homogeneity

2.1 Homogeneity tied to distributivity

One approach to homogeneity takes it to be associated in some way with distributive predicates (Schwarzschild 1994, Gajewski 2005) (see also the discussion in chapter 1 of Križ 2015 for more details). Gajewski 2005 for instance models homogeneity directly as an excluded middle presupposition in the meaning of the D(istributivity) operator, Link 1983[1]:

(2) $||D|| = \lambda P.\lambda x : (\forall y \leq_\alpha x : P(y)) \vee (\forall y \leq_\alpha x : \neg P(y)).\forall y[y \leq_\alpha x \rightarrow P(y)]$

He starts from the assumption that distributive predicates are primitively defined just for atoms and that in order to apply to pluralities, a D operator (Link 1983) needs to apply to them. So, in (1), the function denoted by 'die' cannot apply directly to the plural individual denoted by 'the knights' (assuming that 'the knights' denotes the maximal plural individual that is a knight in the model). Instead, 'die' combines with the D operator, yielding [D [die]]', which then combines with 'the knights'. Thus, the LF of (1) is as in (3):

(3) [The knights [*D* [died-in-battle]]

Here is how Gajewski's operator gives us homogeneity: In models where all atomic knights either died or did not die, the presupposition of the D operator is satisfied. Thus, (1-a) is predicted to be true iff all atomic knights died, and false iff all atomic knights did not die.

However, imagine model with the following extensions

(4) a. $||knight|| = \{a, b\}$
b. $||knights|| = {}^*||knight|| = \{a, b, a \oplus b\}$
c. $||the\ knights|| = a \oplus b$
d. $||king]|| = \{c\}$

[1] Throughout the paper $\oplus$ represents the summation operation on the domain of individuals, D_e. $\leq$ represents the part-of relation on individuals defined in the usual way. $\leq_\alpha$ represents the atomic-part relation. * is the star operator which takes a predicate and closes it under $\oplus$. See Link 1983 for further details.

e. $||die\ in\ battle]|| = \{a, c\}$

The LF in (3) will have the following truth conditions (based on (2)):

(5) $\forall y[y \leq_{\alpha} a \oplus b \rightarrow died - in - battle(y)]$

It will also have the following presupposition:

(6) $(\forall y \leq_{\alpha} a \oplus b : died - in - battle(y)) \vee (\forall y \leq_{\alpha} a \oplus b : \neg died - in - battle(y))$

In this case knight a is in $||die\ in\ battle||$, but b is not. Therefore, the presupposition in (6) cannot be satisfied and presupposition failure arises, which we can think of as undefinedness (see e.g. Heim 1983).

2.2 Homogeneity as an irreducible property

Another approach is that homogeneity is not a reducible property of predicates (Križ 2015, 2019). Importantly, it cannot be reduced to distributivity, because there are collective predicates that show homogeneity[2]. Rather, homogeneity has to be taken as 'a fundamental property of lexical predicates'[3] (Križ 2019) (i.e. non-derived elements that are listed in the lexicon are born homogenous.).

Križ's main argument that homogeneity is not reducible to distributivy stems from the existence of collective predicates that show homogeneity:

(7) a. The students performed *Hamlet.*
b. The students did not perform *Hamlet*

The sentences in (7) seem to require that either all, (7-a), or none, (7-b), of the students participated in a performance of *Hamlet.* They also seem undefined in a mixed situation where only half of the students for instance performed/ did not perform *Hamlet.* But 'perform *Hamlet*' is a collective activity. Thus, homogeneity is found beyond distributive predicates. We will consider this argument more deeply in sections 4 and 5, where we will argue that distributivity is in fact involved in predicates like 'perform *Hamlet*'.

Since homogeneity cannot be identified with distributivity, one needs to state what it means for a predicate to be homogeneous in a way that captures both distributive homogeneous predicates and collective homogeneous predicates. Križ 2019 suggests the following:

[2] Furthermore, Križ argues that the gap associated with homogeneity cannot be reduced to presupposition projection. Since, this is point is orthogonal to the concerns of our paper, we do not go into the details.

[3] This formulation is somewhat ambiguous. It could be taken to mean either (i) that elements of the lexicon need to be specified for their homogeneity, i.e. marked [+/-homogeneous]), or (ii) that every element of the lexicon is homogeneous by default. In light of the fact that Križ views non-homogeneous collective predicates like 'numerous' as exceptions to the idea that homogeneity is a property of lexical predicates, we adopt the second interpretation. Under the first interpretation, one could take predicates like 'numerous' to be simply specified as [-homogeneous], and thus unexceptional.

(8) **Homogeneity Generalization:** A homogeneous predicate P that is not true of a plurality *a* is undefined of *a* if it is true of some plurality *b* that overlaps (i.e. has parts in common)[4] with *a*.

This distinguishes the following cases of homogeneity[5]:

(9) P is not true of *a* and it is true of *b* and ...
   a. *b* is properly contained in *a* (Downward Homogeneity)
   b. *a* is properly contained in *b* (Upward Homogeneity)
   c. *a* and *b* overlap, but neither contains the other (Sideways Homogeneity)

Applying the generalization in (8) to various predicates, Križ identifies two exceptions to the claim that homogeneity is a lexical property:

First, there are (lexical) collective predicates that resist homogeneity. These tend to be measure expressions, like 'be numerous' or 'be heavy'.

(10) The knights were heavy/ numerous

(10) can be true, and not undefined, in a situation where the various subgroups of knights are not heavy, but the plurality of all the knights is heavy.

Second, there are derived predicates that show homogeneity. These are the predicates that are lexically collective but are shifted to a distributive interpretation via the addition of a D operator:

(11) a. The students received a gift.
   b. [The students [D received a gift]].

Under the collective interpretation of (11-a), the students received one common gift as a group. However, the sentence also has an interpretation where each individual student received a gift, and this is homogeneous, since in a situation with five students where only three of them got a pen each, (11-b) appears undefined. Assuming that the latter interpretation is derived via the addition of a D operator, then we have **derived** predicates that are systematically homogeneous. Križ 2019 takes the D operator to be responsible for introducing the homogeneity in these cases.

These exceptions open the possibility that homogeneity is not lexically specified, but rather predictable on the basis of distributivity, with cases like (7) involving hidden distributivity. In the rest of this paper, we explore this idea.

3 Broadening the empirical landscape

In this section we want to broaden the empirical domain by examining which collective predicates exhibit homogeneity. As Križ notes, measure phrases are

[4] Two entities x and y overlap iff there is z such that $z \leq x$ and $z \leq y$.

[5] See section 6 for more discussion on this.

collective predicates that behave in this way. Our claim is that in fact all collective states behave in this way, with measure phrases being just one example of a state. Furthermore, collective states share this behavior with collective achievements, which also do not exhibit homogeneity. This contrasts with collective activities and accomplishments, which do show homogeneity effects.

Consider the following collective states:

(12) a. The students are a productive team.
b. The books constitute a famous series.

Let us apply Križ's homogeneity generalization to (13-a). Imagine a situation where there are three students that are a productive team. Now, sometimes a fourth student joins the team, but the new team is not at all productive. Thus, we have a situation where (12-a) is not true of a plurality *a* (the four unproductive students), true of a plurality *b* (the three productive students) that overlaps with *a*, and in this situation (12-a) is plainly false, and not undefined, with 'the students' referring to *a*.

The same observation holds for (12-b), which can be true in a situation where there is a famous series of five books (plurality *a*), consisting fundamentally of three famous books that form the main series (plurality *b*), and two prequel novels that are not well-known at all. Nonetheless, the five books together (plurality *a*) can be truthfully said to constitute a famous series. No undefinedness arises.

Collective achievements pattern like collective states:

(13) a. The students elected a president
b. The senators passed the bill.

For (13-a) to be true, it is not required that all of the students elected a president. All that is required is that enough students participated in the electoral process in order for a president to be elected. For instance, imagine a class of 20 students who vote to elect a president. 15 students vote for John and 5 students abstain. 'elect a president' can be truthfully applied to the collectivity of 20 students that comprise the class, without any undefinedness arising. However, it is false to say only the 15 students who voted for John, or the 5 students who abstained, elected a president. The election is an effect of everyone participating. The same holds for (13-b): if a certain number of senators vote for the bill, then the bill passes even if not all of them vote.

Conversely, collective accomplishments, (14), pattern together with collective activities (i.e. 'perform *Hamlet*') in showing homogeneity:

(14) The girls built/did not build a raft.

Example (14) is true iff all the girls were involved or were not involved in the building of a raft. In a situation where 5 out of 10 girls were involved in the (collective) building of a raft, (14) appears undefined.

In sum, these additional data highlight a systematic connection between homogeneity and aktionsart:

(15) **Homogeneous Collective Predicates Generalization (version 1)**:
Collective activities/accomplishments show homogeneity;
collective states/achievements do not.

We now turn to the issue of connecting this aktionsart split to distributivity.

4 Brisson 2003 on Taub's generalization

Interestingly, homogeneity is not the only domain where the state/achievement vs accomplishment/activity split matters. The same generalization appears when we try to characterize which collective predicates allow 'all' (this is known as Taub's generalization, (Taub 1989)):

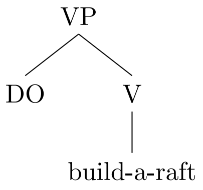
(16) a. All the students performed *Hamlet*
b. All the girls built a raft.
c. *All the students are a productive team.
d. *All the students elected a president.

(17) **Taub's Generalization:** The collective predicates that allow 'all' are collective accomplishments and collective activities. Collective states and colletive achievemnets disallow 'all'.

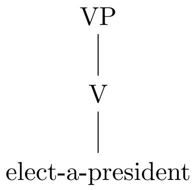
Brisson 2003 treats 'all' as being dependent on the presence of a D operator. If the predicate can host a D operator, then it can license 'all'. But how do we get 'all' to apply to collective achievements and activities, if we think that one of the hallmarks of collectivity is that collective predicates lack a D operator in their structure?

To capture the pattern in (16), Brisson makes a simple move. She claims that collective activities and accomplishments have more structure than collective achievements and states: they have an aspectual DO predicate (McClure 1994) which can host a D operator. States and achievements lack this predicate and hence cannot host a D operator.

(18)



(19)



Brisson adopts an event-based framework, where VPs are predicates of events[6]. She also assumes that the domain of events is structured via a part-of relation, $\leq$. This leads to the following extensions:

(20) a. $||DO|| = \lambda x_e.\lambda e.DO(e) \wedge Ag(e,x)$
b. $||build - a - raft|| = \lambda e.build - a - raft(e)$
c. $||elect - a - president|| = \lambda e.elect - a - president(e)$

[6] Events are of type v

$||DO||$ and $||build\ a\ raft||$ combine via an operation that Brisson terms event composition[7]:

(21) If α is a branching node, $\{\beta, \gamma\}$ the set of its daughters, and $||\beta||$ is function of the form $\lambda e[P(e)]$ (type $\langle v, t\rangle$) and $||\gamma||$ is a function of the form $\lambda x_e.\lambda e[Q(x)(e)]$ (type $\langle e, vt\rangle$), then $||\alpha|| = \lambda x_e.\lambda e.[P(e) \wedge \exists e'[Q(x)(e')] \wedge e' \leq e]$

Applying (22) to (18) leads to the following result:

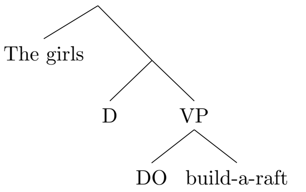
(22) $\lambda x_e.\lambda e.build - a - raft(e) \wedge \exists e'[DO(e') \wedge Ag(e', x) \wedge e' \leq e]$

We also assume a version of the D operator that can handle predicates of events[8]:

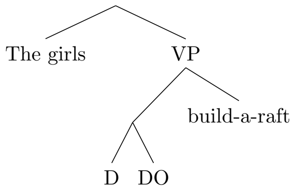
(23) $||D|| = \lambda P_{e,vt}.\lambda x_e.\lambda e.\forall y[y \leq_a x \rightarrow \exists e'[P(y)(e') \wedge Ag(e', y) \wedge e' \leq e]]$

By assumption, the D operator attaches to things that take a plural DP as a first argument. So, it can either attach to DO, or to the VP (i.e. to the result of event composition):

(24)



(25)



Interpreting (25) and (24), we get the truth conditions in (26-a) and (26-b) respectively[9]:

(26) a. $\exists e[\forall y[y \leq_\alpha \iota x.girls(x) \rightarrow \exists e''[build - a - raft(e'') \wedge \exists e'[DO(e') \wedge Ag(e', y) \wedge e' \leq e''] \wedge Ag(e'', y) \wedge e'' \leq e]]]$

b. $\exists e[build - a - raft(e) \wedge \exists e''[\forall y[y \leq_\alpha \iota x.girls(x) \rightarrow \exists e'[DO(e') \wedge Ag(e', y) \wedge e' \leq e''] \wedge e'' \leq e]]$

The truth conditions in (26-a) are the ordinary distributive interpretation, where each girl is required to have built her own raft (since for every girl there is raft-building event of which she is the agent). The truth conditions in (26-b) express a collective reading, where each girl was the agent of some DO-ing subpart of the raft-building event. All the girls still did something, but it was different sub-events in an overall raft-building event. This then gives us a way of formalizing the intuition that events like 'build-a-raft' have distributive sub-entailments (Dowty 1987): each girl participated in the raft-building by doing something.

[7] We formulate event composition assuming a Heim & Kratzer 1998 system.

[8] Brisson uses a Generalized D operator from Lasersohn 1998; We slightly adapt some things here to suit our own purposes. Nonetheless, Brisson's point is retained.

[9] A background assumption in (26), (37) is that every event has a unique agent.

Collective accomplishments and activities lack this DO predicate and therefore cannot host a D operator (and if one attempts to apply one on the VP level, then one ends up with distinctly odd truth conditions; see Brisson 2003 for details)

Therefore, the reason collective accomplishments and activities allow ‘all’ is the D operator that these predicates can host. Notice here that the analysis does not commit us exactly to Taub’s generalization, (17). Brisson’s analysis links the presence of ‘all’ to the presence of distributivity. Therefore, we need to revise (17) as follows:

(27) **Taub’s Generalization (revised):** The collective predicates that allow ‘all’ are those that can host a D operator somewhere in their structure.

Therefore, if we find collective states/achievements that allow ‘all’, we have not necessarily falsified (27). However, we do predict that collective states/achievements that allow ‘all’ involve the presence of a D operator. Reciprocals are a class of collective states that seem to function in this way:

(28) All the students look alike.

Reciprocals like ‘look alike’ are typically taken to involve quantification over parts[10]. If we take the source of this quantification to be some hidden distributivity operator, then such examples do not falsify (27), but rather confirm it.

5 Applying Brisson 2003 to homogeneity

The point that comes out of sections 3 and 4 is that distributivity and homogeneity have the same distribution. Applying Brisson’s idea that collective accomplishments/activities can host a D operator (on the DO part of their structure), whereas collective achievements/states cannot (because they lack this DO) leads us to revise the generalization in (15):

(29) **Homogeneous Collective Predicates generalization (revised):** The collective predicates that show homogeneity are those that can host a D operator in their structure.

What is homogenised in a collective accomplishments/activities such as ‘The students performed *Hamlet*’ is the DO-ing sub-events that each student is undertaking in the performance, e.g. performing the role of Hamlet for student a, being in charge of staging for student b etc.

Moreover, if Brisson is right that the presence of ‘all’ depends on the presence of distributivity, then by reducing homogeneity to distributivity we have a nice connection between the presence of ‘all’ and the presence of homogeneity, whereby only homogeneous predicates license ‘all’. This accords well with the claim that ‘all’ is a homogeneity remover (Kriz 2015, 2019):

[10] See Brisson 2003 and references therein for more details.

(30) The knights all died in battle.

The sentence in (43) is plainly false in a situation where only some of the knights died in battle, and not undefined. Thus, 'all' removes the undefinedness associated with homogeneity. This then constitutes another argument for the reduction we are pursuing: Homogeneity can be removed only from predicates that have it as a property. If 'all' is licensed only by predicates that involve distributivity, then homogeneity must only be present in predicates that involve distributivity.

As with the generalization in (27), our generalization in (29) is not limited to activities and accomplishments. Any collective predicate that involves distributivity qualifies. Consider reciprocals (see previous section):

(31) The students look alike.

(31) is homogeneous, as it is undefined in a situation where some students look alike, but others do not.

One issue that arises is how we can capture the way homogeneity works in negated sentences, where the requirement is that the property expressed by the VP not hold of any part of the subject:

(32) a. The knights did not die in battle.
b. The girls did not build a raft.

We propose to do this by making the following syntactic assumptions: First, in negated sentences, the D operator can only attach above negation[11]. Second, existential closure happens low, on the level of aspect (cf. Hacquard 2009).

These assumptions lead to LFs like the following[12]:

(33) [The girls [D [not [closure [build a raft]]]]]

To properly interpret these LFs, we need another D operator (call it D_2) in addition to the one in (23) (which we call D_1), that will combine with a negated predicate (of type *et*). We also need to define a closure operator that applies low. Finally, we need a meaning for negation. These additions, together with the rest of the lexical entries we need to interpret the sentence in (33), are included below:

(34) a. $||D_1|| = \lambda P_{et}.\lambda x_e.\forall y[y \leq_\alpha x \rightarrow P(y)]$
b. $||D_2|| = \lambda P_{e,vt}.\lambda x_e.\lambda e.\forall y[y \leq_\alpha x \rightarrow \exists e'[P(y)(e') \wedge Ag(e',y)$
$\wedge\, e' \leq e]]$
c. $||closure|| = \lambda P_{e,vt}.\lambda x_e.\exists e[P(x)(e)]$
d. $||not|| = \lambda P_{et}.\lambda x_e.\neg P(x)$

(35) a. $||DO|| = \lambda x_e.\lambda e.DO(e) \wedge Ag(e,x)$
b. $||build\ a\ raft|| = \lambda e.build - a - raft(e)$

[11] We consider only predicate negation, leaving sentential negation to future research.

[12] We follow Gajewski 2005 in assuming that a distributive predicate has to combine with a D operator when it applies to a plurality.

c. $||the\ girls|| = \iota x.girls(x)$

We assume the lexical entries above, together with Brisson's system as presented in the previous section. We give the truth conditions for the following sentences:

(36) a. [The girls [closure [[D_2 DO] [build a raft]]]
b. [The girls [D_1 [not [closure [build a raft]]]]]

(37) a. $\exists e[build - a - raft(e) \wedge \exists e''[\forall y[y \leq_\alpha \iota x.girls(x) \rightarrow \exists e'[DO(e') \wedge Ag(e', y) \wedge e' \leq e''] \wedge e'' \leq e]]$
b. $\forall y[y \leq_a \iota x.girls(x) \rightarrow \neg\exists e[build - a - raft(e) \wedge \exists e'[DO(e') \wedge Ag(e', y) \wedge e' \leq e]]]$

The truth conditions in (37-a) say that for every atomic girl, there is a DO-ing event of which she is the agent, and that DO-ing event is part of a building event. (37-b) says that for no atomic girl is it the case that there is a raft-building event in which this girl did something. These are the truth conditions we want.

Finally, one might wonder how we capture the gappines of these sentences in mixed situations. We will say that a sentence α of the form [NP [(D_1) [closure VP]]][13] has the following truth conditions:

(38) a. α is True in a model M iff || $[NP\ [(D_1)\ [closure\ VP]]]$ || = 1 in M and || $[NP\ [(D_1)\ [not\ [closure\ VP]]]]$ || = 0 in M.
b. α is False in a model M iff || $[NP\ [(D_1)\ [closure\ VP]]]$ || = 0 in M and || $[NP\ [(D_1)\ [not\ [closure\ VP]]]]$ || = 1 in M.
c. α is undefined iff it is neither True nor False.

Consider now (36-a) in a situation where there are 5 girls, three of which participated in the building of a raft, while the other two did not. The truth conditions in (37-a) are not satisfied since it is not the case that for every atomic there is a DO-ing event of which that girl is agent, since 2 girls did nothing. Neither are the truth conditions in (37-b) satisfied, since it is not the case that no atomic girl is the agent of a DO-ing event that is part of a building event. Therefore, neither the positive nor the negative version of (36-a) is true, and the sentence is undefined in this scenario.

6 Comparison with Križ's homogeneity generalization

It is interesting to compare between the cases of homogeneity predicted by our approach and Križ's constraint in (8). Recall that (8) distinguished three cases of homogeneity: (i) downward homogeneity, (ii) upward homogeneity, and (iii) sideways homogeneity. Our approach predicts downward and sideways homogeneity, but not upward homogeneity. Let us go back to the raft-builidng example:

(39) The girls built a raft.

[13] The reason D_1 is in parentheses is that not every sentence will have a D_1 attached above closure, e.g. sentences with collective predicates.

(40) **Context:** Only a subgroup of the girls built a raft. (Downwards Homogeneity)

As we have seen, our approach predicts that (39) should be undefined in (40), since neither (39) nor its negation is true in (40).

(41) **Context:** Some of the boys together with some of the girls built a raft. (Sideways Homogeneity)

Our approach predicts that (39) should be undefined in (41): (39) is not true because the semantics in (37-a) requires that every individual girl participated in the raft-building. (37-b) is not true because there are individual girls that participated in the raft-building.

(42) **Context:** Both the boys and the girls participated in the raft-building. (Upward Homogeneity)

Our approach predicts that (39) should be true, since for every individual girl, that girl participated in the raft-building (which makes (37-a) true and (37-b) false).

Therefore, our approach groups together downward and sideways homogeneity, to the exclusion of upwards homogeneity.

While further research is required into the status of these three homogeneity types, it should be mentioned that there is at least one case where downward homogeneity contrasts with upward homogeneity. While 'all' removes downward homogeneity, it does not remove upward homogeneity (Križ 2015):

(43) All the girls built a raft.

Even though (43) is false in (40), it is still undefined in (42). On the other hand, the addition of 'only' removes upward homogeneity, but not downward homogeneity:

(44) Only the girls built a raft.

(44) is plainly false in (42), but still undefined in (40)[14]. This suggests a parallel with scalar implicatures:

(45) **Context:** All of the students of a class ran in a race.
a. #Some of the students ran in the race.
b. Only some of the students ran in the race.

While (45-a) seems weird in the given context, it is not false, merely under-informative. (45-b) on the other hand, is clearly false. One could imagine then that the reason (39) is odd in (42) is that it is under-informative to claim that the girls built a raft in a context where all the boys and all the girls built a raft.

[14] Interestingly, (43) and (44) seem both false in (41), suggesting that sideways homogeneity has a somewhat mixed status between upward and downward homogeneity.

7 Conclusion

In this paper we have argued for a reduction of homogeneity to distributivity. We noticed that while collective activities and accomplishments license homogeneity, collective states and achievements do not. The same aktionsart split is found with the collective predicates that allow ‘all’. We followed Brisson 2003 who argues that collective activities and accomplishments have a DO predicate as part of their structure that can host a D operator. Collective states and achievements lack this DO and hence cannot host a D operator. We used this to formulate the generalization that the collective predicates that allow homogeneity are those that can host a D operator and developed a semantics for homogeneity (following again Brisson) whereby homogeneity arises as an effect of distributvity (together with certain assumptions about the syntax of D). Finally, we showed that while our account captures downward and sideways homogeneity, it cannot capture upward homogeneity, and made the tentative suggestion that upward homogeneity might be due to Gricean reasoning.

References

1. Brisson, C.: Plurals, all and the Nonuniformity of Collective Predication. Linguistics and Philosophy **26**. 129–184. (2003)
2. Dowty, D.: Collective Predicates, Distributive Predicates, and All. In: F. Marshall (ed.), Proceedings of the 3rd ESCOL, Ohio State University, Ohio (1987)
3. Fodor, J. D.: The linguistic description of opaque contexts. MIT dissertation (1970)
4. Gajewski, J.: Neg-raising: Polarity and Presupposition. MIT dissertation (2005).
5. Hacquard, V.: On the Interaction of Aspect and Modal Auxiliaries. Linguistics and Philosophy, **32**, 279–315 (2009)
6. Heim, I.: On the projection problem for presuppositions. In: M. Barlow, D. Flickinger and N. Wiegand (eds.), Proceedings of WCCFL 2, pp. 114–125. Stanford University (1983)
7. Heim, I., Kratzer, A.: Semantics in Generative Grammar. Blackwell, Oxford (1998)
8. Križ, M.: Aspects of homogeneity in the semantics of natural language. University of Vienna dissertation (2015)
9. Križ M.: Homogeneity effects in natural language semantics. Language and Linguistic compass. (2019)
10. Lasersohn, P: Generalized Distributivity Operators, Linguistics and Philosophy, **21**, 83–93 (1998)
11. Link, G: The Logical Analysis of Plurals and Mass Terms: A Lattice Theoretical Approach, In: Bauerle et al. (eds.), Meaning, Use, and Interpretation of Language, DeGruyter, Berlin (1983)
12. McClure, W.: Syntactic Projections of the Semantics of Aspect, Cornel dissertation (1994).
13. Schwarzschild, R.: Plurals, presuppositions and the sources of distributivity. Natural Language Semantics **2**(3). 201–248 (1994)
14. Taub, A.: Collective Predicates, Aktionsarten and All, In: Emmon Bach, Angelika Kratzer, and Barbara Partee (eds.), Papers on Quantification, University of Massachusetts at Amherst (1989)

The Falsity of the Consequent in Contrastive Conditionals

Hayley Ross

Brandeis University, Waltham MA 02453, USA
hayleyross@brandeis.edu

Abstract. This paper presents novel empirical observations showing that counterfactuals and conditionals involving a contrast between events generate the implicature that their consequent is false, in addition to the well-known implicature that the antecedent is false. I show that existing compositional theories which predict the falsity of the antecedent do not predict the falsity of the consequent, and propose an extension of Starr's unified semantics [22] to capture it. I also examine the effect of this new implicature on minimal model generation for conditionals, providing algorithms for generating minimal models capturing the falsity of the antecedent for counterfactuals, the consistency of presuppositions with the actual world [9,22] for indicative and future subjunctive conditionals, and this newly observed falsity of the consequent.

Keywords: Counterfactuals · Conditionals · Implicature · Computational Semantics · Minimal Model Generation.

1 Introduction

In this paper we will be concerned with three types of conditionals, which I will call *indicative conditionals*, *subjunctive conditionals* and *counterfactuals* (note that in other papers, the term *subjunctive* often includes counterfactuals):

(1) If he takes this syrup, he will get better. (indicative)

(2) a. If Mary knew the answer, she would be the only one. (present subj.)

b. If you went tomorrow, you would see Ed. (future subj.)

(3) If your plants had died, I would have been very upset. (counterfactual)

Previous work on conditionals focuses on establishing a semantics for counterfactuals (first formulated using possible worlds by Stalnaker [20] and Lewis [14]) and establishing a unified compositional semantics for all types of conditionals ([12,8,9,2,22] inter alia). Four aspects of this problem become apparent: how to define the similarity relation between possible worlds ([20,14,15,13,10] inter alia), the falsity of the antecedent for counterfactuals (shown to be an implicature by [1,11,5] inter alia), the consistency of presuppositions of indicative [22] and subjunctive [9] conditionals with the actual world, and the chaining together of multiple counterfactuals in sequences such as modal subordination or Sobel sequences [19,6,23,22]. In this paper, I will focus on the second and third issues: implicatures and presuppositions. I will show that an additional implicature arises for “contrastive” conditionals, namely the falsity of the consequent. I will discuss the unified semantics for conditionals proposed by Ippolito [9] and Starr [22] which account for the falsity of the antecedent, show why neither of these nor the scalar implicature of conditionals [17] accounts for the falsity of the consequent, and propose an extension of Starr’s proposal as a first account for it. Finally, I will discuss a computational model which captures the implicatures for both the antecedent and the consequent as well as the consistency of the antecedent’s presuppositions using minimal model generation.

2 Novel Data: The Falsity of the Consequent

2.1 The Existence of the Implicature

Consider the following counterfactual[1]:

(4) If Charlie had taken his test tomorrow, he would have passed.

If we utter this sentence with no other preamble, it appears to convey two things: (I) that Charlie took his test at some time in the past (instead of tomorrow), and (II) that when he took it, he didn’t pass. Observation (I) can be derived from the combination of the temporal adverb with the falsity of the antecedent (assuming that Charlie may only take his test once), but observation (II), namely that the consequent is also false in the actual world, is new. We can test that this information is in fact conveyed: the following combination is infelicitous.

(5)# Charlie took his test last week and failed. If Charlie had taken his test tomorrow, he wouldn’t have passed.

Because the conclusion sentence in (6) is non-redundant, the falsity of the consequent is not asserted:

(6) If Charlie had taken his test tomorrow, he would have passed. In fact, he took his test yesterday, and he didn’t pass.

Further, the falsity of the consequent can be canceled as in (7), so cannot be a presupposition:

(7) Charlie took his test last week and failed. If Charlie had taken his test on Monday, he would have failed. But if he had taken it on Tuesday, he would have passed.

[1] I will use what Ippolito [9] calls “mismatched past counterfactuals”, i.e. counterfactuals with a future temporal adverb, to allow a direct analogy between counterfactuals and future subjunctive counterfactuals. Some native speakers take issue with future counterfactuals; firstly, the same judgments can be replicated by replacing *tomorrow* with *yesterday* throughout; secondly, counterfactuals in fact show this behavior without any temporal contrast: *If Charlie had brought his calculator to his test, he would have passed* yields the same implicature for the consequent.

The third sentence (the “conclusion”, like the Anderson-style conclusions for the falsity of the antecedent [1]) greatly improves the felicity of the counterfactual in (7). Without it, there is a sense of awkwardness due to the seeming lack of new information. (Some speakers resolve this without the conclusion sentence by adding an implicit *still* (*still would have failed*) to the counterfactual.)

What is happening here is that (4) is essentially *contrasting* a possible test-taking event and its outcome with another implicit test-taking event in the past. (If the consequent were not false, there would be no contrast between the two events, yielding the awkwardness we saw when canceling the implicature.) This contrast is equivalent to the combination of (I) and (II). We can replicate this contrast in future subjunctive conditionals if we change the situation to allow repetition: imagine that Charlie can now re-take his test. That is, we can analogously show that the falsity of the consequent also occurs in (8) and is an implicature by applying the tests (9-11):

(8) If Charlie re-took his test tomorrow, he would pass.

(9)# Charlie took his test last week and failed. If Charlie re-took his test tomorrow, he wouldn’t pass.

(10) If Charlie re-took his test tomorrow, he would pass. When he took his test yesterday, he didn’t pass.

(11) Charlie took his test last week and failed. If Charlie re-took his test on Thursday, he would fail. But if he re-took it on Friday, he would pass.

Interestingly, we cannot replicate it in indicative conditionals, even with event repetition. The following example is felicitous, if a little odd:

(12) Charlie took his test last week and failed. If Charlie re-takes his test tomorrow, he won’t pass.

For the purpose of this paper, I will not discuss indicative conditionals in any more detail and assume that “anything goes” (all combinations are felicitous and the falsity of the consequent does not occur). However, initial experiments with native speakers indicate that it is not quite so simple and that this aspect of indicative conditionals is open for further research.[2].

2.2 Comprehensive Data Collection

The examples above show that for at least one counterfactual and one future subjunctive conditional involving *passing/failing a test*, the falsity of the consequent arises and is an implicature. This section describes a comprehensive collection of data for each combination of *(not) passing/failing* for both counterfactuals and future subjunctive conditionals, as well as controlling for the inherent polarity and normativity of that example with a second example: *getting vanilla/strawberry ice-cream* (from selection of at least three flavors). This includes sentence pairs such as the following.

[2] It appears, for example, that indicative conditionals with false consequents are preferred over those with consequents true in the actual world, however this preference is not significant enough to cause infelicity.

(13) Charlie went to the ice-cream parlor last week and got vanilla ice-cream. If Charlie had gotten ice-cream tomorrow, he would have gotten strawberry.

(14) Charlie went to the ice-cream parlor last week and didn't get vanilla ice-cream. If Charlie got ice-cream tomorrow, he would get vanilla.

The judgments summarized in Tables 1 and 2 were gathered from three native speakers of English, with additional judgments from a further 15 native speakers for a representative selection of controversial cases (indicated by a shaded cell background)[3]. The rows correspond to Charlie's test outcome or ice-cream choice in the set-up sentence, while the columns correspond to the test outcome or ice-cream choice in the conditional. I will use $\tilde{\checkmark}$ to indicate that the combination is felicitous but awkward in the sense discussed above.

Table 1. Felicity for *pass/fail*, counterfactuals and future subjunctive conditionals.

		Counterfactual				Subjunctive (future)			
		pass	fail	not pass	not fail	pass	fail	not pass	not fail
Set-up	passed	$\tilde{\checkmark}$	✓	✓	#	$\tilde{\checkmark}$	✓	✓	✓*
	failed	✓	$\tilde{\checkmark}$	#	✓	✓	$\tilde{\checkmark}$	#	✓
	didn't pass	✓	#	$\tilde{\checkmark}$	✓	✓	#	$\tilde{\checkmark}$	✓*
	didn't fail	#	✓	✓	$\tilde{\checkmark}$	#	✓	✓	$\tilde{\checkmark}$
	not taken	#	#	#	#	✓	✓	✓	✓*

Note that in Table 1, the last row for counterfactuals is infelicitous throughout due the falsity of the antecedent implicature. We notice our first irregularity for subjunctive conditionals: based on the pattern for counterfactuals, we would expect the combination *passed – wouldn't fail* to be infelicitous. However, this conditional is rescued by a secondary reading of *not failing*: we may interpret *pass*, *not fail* and *fail* to be on a scale, in the sense of *"Well, I didn't fail the test, but I wouldn't really say I passed it either"*. This reading is indicated by ✓*. Thus, the outcome in the conditional differs from the set-up sentence after all. Curiously, this reading does not occur for the corresponding counterfactuals.

Table 2 shows the same pattern of infelicity across the top right-bottom left diagonal (*x* and *not y* pairs) for the ice-cream example, but also show additional infelicity in the bottom right (*not x* and *not y*). This infelicity arises from the underspecification of Charlie's choice of ice-cream in the set-up sentence, which is not resolved by the conditional: since there are more than two flavors, we don't know in either case what flavor he chose.

In summary, we see that we have felicity whenever the conditional has a different outcome to the set-up sentence, felicity with awkwardness when the outcomes are identical (top left-bottom right diagonal), and infelicity when the outcomes are either the same but described differently (e.g. *fail* and *not pass*) or are not mutually exclusive (e.g. *vanilla* and *not strawberry*).

[3] For the full results and analysis, see [18]. The responses were clear in all cases except the subjunctive *vanilla/not strawberry* pair, where half of the respondents were able to accommodate context to make it felicitous. Since the other half could not, I judged this ultimately infelicitous.

Table 2. Felicity for *get vanilla/strawberry ice-cream*, counterfactuals and future subjunctive conditionals.

		Counterfactual				Subjunctive (future)			
		strawb.	vanilla	not str.	not vanilla	strawb.	vanilla	not str.	not vanilla
Set-up	strawb.	˜✓	✓	✓	#	✓	✓	✓	#
	vanilla	✓	˜✓	#	✓	✓	✓	#	✓
	not strawb.	✓	#	˜✓	#	✓	#	✓	#
	not vanilla	#	✓	#	˜✓	#	✓	#	✓
	no ice-cream	#	#	#	#	✓	✓	˜✓	˜✓

I will generalize this as follows:

> Generalization. Counterfactuals and subjunctive conditionals contrasting with another event carry the implicature that their consequent is false; that is, that the event in their consequent differs in some way from the first event. This can be canceled, but only if the consequent uses the exact same, salient predicate used to describe the other (previous or actual world) event.

3 Theoretical Accounts

In this section, I will discuss three possible accounts of the falsity of the consequent: the scalar implicature of conditionals [17] and two unified compositional accounts of conditionals by Ippolito [9] and Starr [22], which are the only unified semantics to explicitly account for the falsity of the antecedent. I will show that none of these theories account for the falsity of the consequent directly, and propose an extension of Starr's account which does predict it.

3.1 Scalar Implicature of Conditionals

A simple explanation for the falsity of the consequent could be the "scalar implicature of conditionals": that (15a) implies (15b) (similarly for counterfactuals).

(15) a. If you mowed the lawn, I would pay you ten dollars.

b. If you didn't mow the lawn, I wouldn't pay you ten dollars.

In other words, $p \rightarrow q$ implies $\neg p \rightarrow \neg q$ [17]. Given that we know that counterfactuals generate the implicature that their antecedent is false, $\neg p$, with this second implicature we should be able to derive the falsity of the consequent $\neg q$ using modus ponens. However, this scalar implicature does not arise in the conditionals and counterfactuals with temporal contrast used here, so we cannot use it to explain the falsity of the consequent. For example, (16a) and (17a) do not imply (16b) and (17b) respectively: passing on a certain day does not imply failing on all the other days[4].

[4] Note that we need to use *yesterday* rather than *tomorrow* in the counterfactuals, otherwise the negated version is infelicitous due to the falsity of the antecedent.

(16) a. If Charlie took his test tomorrow, he would pass.
b. If Charlie didn't take his test tomorrow, he wouldn't pass.

(17) a. If Charlie had taken his test yesterday, he would have passed.
b. If Charlie hadn't taken his test yesterday, he wouldn't have passed.

Thus, while the scalar implicature of conditionals may account for some simple cases of the falsity of the consequent, it is not suitable as a general account.

3.2 Ippolito's Account

Ippolito [9] derives the falsity of the antecedent from the speaker's choice of a counterfactual over a subjunctive conditional. Earlier in her paper, she sets out that the presuppositions for subjunctive conditionals must be consistent with the context set at utterance time. For counterfactuals, the presuppositions must only be consistent with the context set at some salient past time. Thus, asserting a subjunctive conditional is stronger than asserting a counterfactual. This asymmetry creates a Gricean scalar implicature, which Ippolito exploits to derive the implicature of the falsity of the antecedent. Given that a counterfactual is chosen, she derives that one of the antecedent's presuppositions must be false and thus the antecedent is undefined in the actual world. (This is not strictly the same as the antecedent being (defined and) false, but that does not matter for our argument.) Crucially, Ippolito's derivation rests on the choice of a counterfactual over a subjunctive conditional. We observe the falsity of the consequent for both counterfactuals and certain subjunctive conditionals, therefore, the distinction between them cannot motivate it.

3.3 Starr's Account

Starr [22] also provides an account of conditionals which derives the falsity of the antecedent. Starr distinguishes indicative from "subjunctive" conditionals, by which he means both subjunctive conditionals and counterfactuals. His account of the differences between the two centers around the following principle, which he draws from [21]:

> Stalnaker's Distinction: An indicative conditional focuses solely on ante-cedent-worlds among the contextually live possibilities. A "subjunctive" conditional focuses on antecedent-worlds that need not be among those possibilities, that is they may be counterfactual from the perspective of the discourse.

Much like Ippolito's account, Starr's derivation depends on the speaker's choice of conditional, only this time of a "subjunctive" (subjunctive or counterfactual) conditional over an indicative conditional. Specifically, Starr explains the infelicity of (18) as follows:

(18) # Bob always danced. If Bob had danced, Leland would have danced.

The first sentence limits the live possibilities to the ones where Bob danced. Following Stalnaker's Distinction, the second sentence "suggests"[5] that some relevant antecedent-worlds (where Bob dances) may be outside the live possibilities. However, Starr's update semantics for the conditional in conjunction with his Centering Axiom show that the only worlds considered by the conditional are the same live worlds where Bob dances. This is "at odds" with the speaker's selection of a subjunctive conditional, which suggested that some relevant worlds were counterfactual.

Can we extend Starr's account to capture the falsity of the consequent? Suppose we extend Stalnaker's Distinction to handle consequent-worlds like antecedent-worlds for subjunctive conditionals and counterfactuals.

> Stalnaker's Distinction, Extended: An indicative conditional focuses solely on antecedent-worlds among the contextually live possibilities. A subjunctive conditional or a counterfactual focuses on antecedent-worlds and consequent-worlds that need not be among those possibilities.

Consider our original infelicitous example:

(5) # Charlie took his test last week and failed. If Charlie had taken his test tomorrow, he wouldn't have passed.

The first sentence limits the live possibilities to the ones where Charlie took his test last week and failed it. The counterfactual, by the Extended Distinction, "suggests" that that the antecedent-world and the consequent-world need not be among the live possibilities. Indeed, the antecedent is not among the live possibilities because Charlie took his test yesterday (which we are assuming is not last week), and we presuppose that Charlie can only take his test once. However, strictly speaking, the consequent is not among the live possibilities either, because the live possibilities involve *Charlie failing / not passing last week*, while the consequent describes *Charlie failing / not passing yesterday.* (Since Charlie may only take his test on one day, he may only fail it on one day as well.) Thus, Starr's reasoning with our Extended Distinction incorrectly predicts this counterfactual to be felicitous.

3.4 Modifying Stalnaker's Distinction: A First Account

We observed that the difference in time in (5) caused the consequent to be "automatically" false, even when the observed judgment is that the combination is infelicitous. Of course, we can't ignore time, because it drives the falsity of the antecedent, but it is as if the difference in time is "checked off" when determining the falsity of the antecedent, and that same difference in time no longer counts as different when evaluating the consequent. I will propose the following modification of Stalnaker's Distinction, incorporating this "checking off" and also Ippolito's observation on the antecedents of future subjunctive conditionals.

[5] Stalnaker uses the notion of "suggestion" throughout, avoiding the explicit notion of implicature; however, the behavior of his "suggestion" is essentially the same as that of a generalized implicature, namely that it defaults to being true but can be canceled, and creates infelicity if it is violated but not canceled.

> New Distinction: (I) Indicative and future subjunctive conditionals focus solely on antecedent-worlds among the contextually live possibilities. Present subjunctive conditionals and counterfactuals focus on antecedent-worlds that need not be among those possibilities. (II) Subjunctive conditionals and counterfactuals focus on consequent-worlds that, discounting differences logically entailed by the antecedent, need not be among the contextually live possibilities.

The stipulation of entailment in the New Distinction is crucial. In general, consequents of conditionals follow causally from their antecedents; such is the crux of the substantial body of work on counterfactuals and causality ([7,4] and many others). We want to exclude "uninformative" consequences that can be derived logically without causal relations. In this case, we suppose (according to the conditional) that (i) Charlie's test-taking happens tomorrow, and (ii) Charlie's test outcome is that he failed. We also know that (iii) Charlie's test-taking and Charlie's test outcome must happen at the same time (for the sake of argument, as an inherent property of tests). Therefore, we can derive without invoking causality that: (iv) Charlie's test outcome happens tomorrow, and therefore (v) Charlie's failing happens tomorrow.

Consider example (5) again. The first sentence limits the live possibilities to the ones where Charlie took his test last week and failed it. The counterfactual, by the New Distinction, "suggests" that that the antecedent-world need not be among the live possibilities. Indeed, it is indeed not among them because Charlie may only take his test on one day. For the consequent, we first modify the consequent to remove the changes entailed by the antecedent, namely that the test and thus the *not passing* happen tomorrow. By the New Distinction, this modified consequent need also not be compatible with the live possibilities. However, it is: *Charlie doesn't pass* (with no time attached) is compatible with the live possibilities (the actual world) where *Charlie fails/failed.* This is "at odds" with the choice of a counterfactual over an indicative conditional, and thus infelicitous – as is indeed correct.

We can apply the same reasoning to contrastive subjunctive conditionals.

(19) # Charlie went to the ice-cream parlor last week and got vanilla ice-cream.
If Charlie got ice-cream tomorrow, he wouldn't get strawberry.

The first sentence limits the live possibilities to the ones where Charlie got vanilla ice-cream last week. The future subjunctive conditional, by the New Distinction, requires that the antecedent-world be among the live possibilities; this is true (there are no restrictions on Charlie getting ice-cream again tomorrow). Moving to the consequent, we again first modify it to remove the entailed change that the *ice-cream-getting* happens tomorrow. Now, the remaining consequent proposition, *Charlie doesn't get strawberry ice-cream*, is fully compatible with the live possibilities where *Charlie gets/got vanilla ice-cream.* The New Distinction says that it should be incompatible, again yielding the required infelicity.

4 Minimal Model Generation

Much work on counterfactuals is concerned with the possible world or worlds that counterfactuals should be evaluated in. Minimal model generation [3] offers an alternative angle on the same problem: What must be true in that possible world, however it is selected, for the counterfactual to be felicitous? What must be true in the actual world? If a speaker utters a counterfactual or contrasting conditional without the set-up sentence describing the past event, then discourse participants will essentially accommodate a suitable set-up sentence: the minimal model for contrasting conditionals includes additional facts about the past. This section will present an overview of algorithms for minimal model generation which capture these implicatures and accommodations. Notably, the model generation will fail if the accommodation required for the conditional and its implicature(s) is inconsistent with the model created from the previous discourse. This captures and predicts the infelicity of examples such as (5). Conversely, successful model generation implies felicity.

The full model generation algorithm is implemented in Haskell and is publicly available on GitHub[6]. In brief, the end-to-end algorithm takes a list of syntactic trees representing the sentences in the discourse, converts the trees to intermediate data structures containing the information necessary for model generation, and then generates the minimal model. The full set of algorithms is described extensively in [18]. In this paper, I will focus on the actual model generation. The parsing of trees into intermediate structures is done straightforwardly by tree traversal and lookup tables; the curious reader is welcome to read [18] or explore the Haskell implementation on GitHub in the `Parsing` module.

I will use an LTL-style interpretation of time, namely that time has a linear structure (no branching futures) in conjunction with a broadly Davidsonian view of events and the standard computational view that worlds are simply the set of propositions which are true in them.

Definition 1. *A **proposition** consists of an event, a boolean indicating negation, and a boolean indicating whether it is cancelable (i.e. an implicature).*

Definition 2. *A **minimal model** consists of two maps from times to worlds. One map contains actual worlds, the other contains possible worlds. A **world** is a set of propositions plus a boolean indicating if the world is actual or possible.*

4.1 Model Generation Algorithms

At a high level, the model generation algorithm has three cases: simple sentences, conditionals with no time contrast and conditionals with time contrast. For each of these, it generates a model just for the sentence, then invokes the algorithm CombineModels to combine the sentence model with the existing discourse model. This step may fail and yield `Nothing`, indicating that the sentence is infelicitous / inconsistent with the previous discourse. The majority of

[6] https://github.com/rossh2/counterfactual-model-generation

CombineModels is just simple recursion over the minimal model data structure; the crucial logic happens in the function AddProp (Algorithm 1) which adds a new proposition to a given world, or returns `Nothing` if this would cause inconsistency. This includes canceling implicatures if applicable and checking for antonyms (making sure that combining e.g. *pass* and *not fail* invoke either the implicature check or cause failure). Observe that AddProp will crucially fail on the case where earlier in the discourse, *Charlie passed* was asserted, and we add the counterfactual *Charlie wouldn't have failed.* This creates the implicature *Charlie failed.* When trying to add this implicature, it matches on antonym with *Charlie passed* but describes a different event (the variable *sameEvent* is False). Thus the variable *tryToCancel* is False and the model returns `Nothing`.

Algorithm 1 Add a proposition to a world, if consistent

```
function AddProp($q, w$)
    $matching$ ← all $p$ in $w$ s.t. $p, q$ have the same event
    $antonymMatching$ ← all $p$ in $w$ s.t. $p, q$'s events are antonyms
    if $matching$ and $antonymMatching$ are empty then
        return $w$ with $q$ inserted
    else                        ▷ Uniqueness of $p$ is guaranteed by consistency of $w$
        $p$ ← the unique matched proposition in $matching$ or $antonymMatching$
        $sameEvent$ ← True iff $p$ has the opposite negation value to $q$
        $tryToCancel$ ← True iff $matching$ is not empty or $sameEvent$ is True
        $cancelExisting$ ← True iff $p$ is cancelable
        $nothingToDo$ ← True iff $q$ is cancelable or $p = q$ or $sameEvent$ is True
        if $tryToCancel$ and $cancelExisting$ then
            $v$ ← Delete $p$ from $w$
            return $v$ with $q$ inserted
        else if $tryToCancel$ and $nothingToDo$ then
            return $w$
        else
            return Nothing
        end if
    end if
end function
```

For the per-sentence model generation, the two conditional cases are laid out in Algorithms 2 (non-contrasting) and 3 (time contrast); the case for simple sentences trivially adds the proposition and its presuppositions to the actual world. These algorithms capture the falsity of the antecedent for counterfactuals, any presuppositions of the antecedent, and the newly observed falsity of the consequent for contrastive conditionals. To check conditionals for time contrast, the algorithm checks if a) the predicate of the antecedent p has the `Repetition` property in the lexicon (such as *retake*), b) p's event matches any event in the actual worlds of m, or c) if the conditional is a counterfactual, returning true if any of those conditions are true. (In the below algorithms, the utility function WorldsFromProps handles the details of building a map of times to worlds from a list of propositions.)

Algorithm 2 Generate a minimal model for a non-contrasting conditional

```
function GenerateConditionalModel($p, q$)
    $antePresupps \leftarrow$ presuppositions of $p$ with time set to Present
    $props \leftarrow [antePresupps, p, q,$ presuppositions of $p, q]$
    $actualWorlds \leftarrow$ WorldsFromProps($props$)
    $possibleWorlds \leftarrow$ empty map
    return $(actualWorlds, possibleWorlds)$
end function
```

Algorithm 3 Generate a minimal model for a conditional with time contrast

```
function GenerateContrastModel($p, q$)
    if $q$ is Subjunctive then
        $antePres \leftarrow$ presuppositions of $p$ with time set to Present
        $contrP \leftarrow p$ with time set to Past
        $contrQ \leftarrow \neg q$ with time set to Past
        $actualProps \leftarrow [antePres, p, q, contrP, contrQ,$ pre's of $p, q, contrP, contrQ]$
        $actualWorlds \leftarrow$ WorldsFromProps($actualProps$)
        $possibleWorlds \leftarrow$ empty map
        return $(actualWorlds, possibleWorlds)$
    else $q$ is Counterfactual
        $actualP \leftarrow \neg p$ with time set to Past
        $actualQ \leftarrow \neg q$ with time set to Past
        $actualProps \leftarrow [actualP, actualQ,$ presuppositions of $actualP, actualQ]$
        $actualWorlds \leftarrow$ WorldsFromProps($actualProps$)
        $possibleWorlds \leftarrow$ WorldsFromProps($[p, q,$ presuppositions of $p, q]$)
        return $(actualWorlds, possibleWorlds)$
    end if
end function
```

5 Conclusion

While the antecedent of conditionals has been studied extensively, their consequent has received little attention in previous literature. In this paper, we saw empirically that counterfactuals and contrastive subjunctive conditionals yield an additional implicature: the falsity of the consequent. We discussed possible accounts for this and concluded by extending Starr's account to give the New Distinction. This correctly predicts the falsity of the consequent. While this account is effective, it must be noted that both Stalnaker's Distinction and the New Distinction are somewhat stipulative: we stipulate which possibilities different types of conditionals "focus on' and then derive the desired implicatures. Notably, this preliminary account does not use the contrast inherent to these conditionals. Further research could explore a more compositional theory which might combine the lexical and event semantics of this contrast with the semantics of conditionals to derive the implicature directly, perhaps adapting the role of contrast from [16]. The minimal model generation algorithms in this paper thus serve a dual purpose: to lay out explicitly how these implicatures are generated and processed in minimal models, but also to serve as inspiration for how the conditional semantics interact with the event and lexical semantics involved.

Many thanks to the anonymous reviewers, my thesis supervisor James Pustejovsky, and my advisor Lotus Goldberg for all the guidance and feedback.

References

1. Anderson, A.R.: A note on subjunctive and counterfactual conditionals. Analysis **12**(2), 35–38 (1951). https://doi.org/10.2307/3327037
2. Arregui, A.: When aspect matters: the case of would-conditionals. Natural Language Semantics **15**(3), 221–264 (2007). https://doi.org/10.1007/s11050-007-9019-6
3. Blackburn, P., Bos, J.: Computational semantics. Theoria: An International Journal for Theory, History and Foundations of Science **18**(1(46)), 27–45 (2003)
4. Ciardelli, I., Zhang, L., Champollion, L.: Two switches in the theory of counterfactuals. Linguistics and Philosophy **41**(6), 577–621 (2018). https://doi.org/10.1007/s10988-018-9232-4
5. von Fintel, K.: The presupposition of subjunctive conditionals. The Interpretive Tract **25**, 29–44 (1998)
6. von Fintel, K.: Counterfactuals in a dynamic context. Current Studies in Linguistics Series **36**, 123–152 (2001)
7. Galles, D., Pearl, J.: An axiomatic characterization of causal counterfactuals. Foundations of Science **3**(1), 151–182 (1998). https://doi.org/10.1023/A:1009602825894
8. Iatridou, S.: The grammatical ingredients of counterfactuality. Linguistic Inquiry **31**(2), 231–270 (2000). https://doi.org/10.1162/002438900554352
9. Ippolito, M.: Presuppositions and implicatures in counterfactuals. Natural Language Semantics **11**(2), 145–186 (2003)
10. Ippolito, M.: How similar is similar enough? Semantics and Pragmatics **9**, 6–1 (2016). https://doi.org/10.3765/sp.9.6
11. Karttunen, L., Peters, S.: Conventional implicature. In: Presupposition, pp. 1–56. Brill (1979). https://doi.org/10.1163/9789004368880_002
12. Kratzer, A.: The notional category of modality. Words, Worlds, and Contexts: New Approaches in Word Semantics **6**, 38 (1981). https://doi.org/10.1515/9783110842524-004
13. Kratzer, A.: An investigation of the lumps of thought. Linguistics and Philosophy pp. 607–653 (1989). https://doi.org/10.1007/BF00627775
14. Lewis, D.: Counterfactuals. Blackwell (1973)
15. Lewis, D.: Counterfactual dependence and time's arrow. Noûs pp. 455–476 (1979). https://doi.org/10.2307/2215339
16. Ogihara, T.: Counterfactuals, temporal adverbs, and association with focus. In: Semantics and Linguistic Theory. vol. 10, pp. 115–131 (2000)
17. Prince, E.F.: Grice and Universality: A Reappraisal. Manuscript (1982)
18. Ross, H.: The Falsity of the Consequent in Contrastive Conditionals. Master's thesis, Brandeis University (Jun 2020), http://bir.brandeis.edu/handle/10192/37528
19. Sobel, J.H.: Utilitarianisms: Simple and general. Inquiry **13**(1-4), 394–449 (1970). https://doi.org/10.1080/00201747008601599
20. Stalnaker, R.: A theory of conditionals. In: Ifs, pp. 41–55. Springer (1968). https://doi.org/10.1007/978-94-009-9117-0_2
21. Stalnaker, R.: Presuppositions. In: Contemporary research in philosophical logic and linguistic semantics, pp. 31–41. Springer (1975). https://doi.org/10.1007/BF00262951
22. Starr, W.B.: A uniform theory of conditionals. Journal of Philosophical Logic **43**(6), 1019–1064 (2014). https://doi.org/10.1007/s10992-013-9300-8
23. Veltman, F.: Making counterfactual assumptions. Journal of Semantics **22**(2), 159–180 (2005). https://doi.org/10.1093/jos/ffh022

The 'In' and 'According to' operators[⋆]

Merel Semeijn

University of Groningen
m.semeijn@rug.nl
https://merelsemeijn.wordpress.com/

Abstract. Semanticists and philosophers of fiction that formulate analyses of reports on the content of media – or 'contensive statements' – of the form 'In/According to s, ϕ', usually treat the 'In s'-operator (**In**) and the 'According to s'- operator (**Acc**) on a par. I argue that **In** and **Acc** require separate semantic analyses based on three novel linguistic observations: (1) preferences for **In** or **Acc** in contensive statements about fictional or non-fictional media, (2) preferences for **In** or **Acc** in contensive statements about implicit or explicit content and (3) tense preferences in contensive statements with **In** and **Acc**. To account for these three observations I propose to adopt the Lewisian possible world analysis for contensive statements with **In** and to analyse contensive statements with **Acc** as indirect speech reports.

Keywords: 'According to' · contensive statements · fiction operator · 'In' · parafictional statements · speech reports

1 Introduction

Semanticists of fiction distinguish between, in Recanati's [20] terminology, 'fictional' statements, i.e. statements that are part of a fictional narrative such as (1) below taken from *The Hobbit*, and 'parafictional' statements, i.e. statements about the content of some fictional narrative. Parafictional statements can feature either an 'In s' operator (**In**) as in (2a) or an 'According to s' operator (**Acc**) as in (2b):[1]

(1) In a hole in the ground there lived a hobbit.

(2) a. In *The Hobbit*, Bilbo travels to the Lonely Mountain.
b. According to *The Hobbit*, Bilbo travels to the Lonely Mountain.

On the face of it, the **In** and **Acc** operators also exist in other languages such as Dutch (e.g. (3) and (4)) and Spanish (e.g. (5) and (6)):

(3) In *De Hobbit* reist Bilbo naar de Eenzame Berg.
In *The Hobbit* travels Bilbo to the Lonely Mountain.

[⋆] This research is supported by NWO, Vidi Grant 276-80-004. Many thanks to Natasha Korotkova and Emar Maier for valuable comments and discussions.

[1] Most theorists take parafictional statements to also have implicit variants where the fiction operator is covert (e.g. "Bilbo travels to the Lonely Mountain").

(4) Volgens *De Hobbit* reist Bilbo naar de Eenzame Berg.
According-to *The Hobbit* travels Bilbo to the Lonely Mountain.

(5) En *El Hobbit*, Bilbo viaja a la Montaña Solitaria.
In *The Hobbit*, Bilbo travels to the Lonely Mountain.

(6) Según *El Hobbit*, Bilbo viaja a la Montaña Solitaria.
According-to *The Hobbit*, Bilbo travels to the Lonely Mountain.

Whereas fictional statements *determine* what is true in the fiction (i.e. The fact that (1) is part of *The Hobbit* makes it true in *The Hobbit* that a hobbit lived in a hole in the ground), parafictional statements *report* on what is true in the fiction (i.e. (2a) and (2b) and their translations are true statements because it is true in *The Hobbit* that Bilbo travels to the Lonely Mountain). One of the central objectives of semantics of fiction is to provide a semantic analysis of fictional and parafictional statements that takes into account this difference in function. In providing these analyses many philosophers (e.g. Zucchi [26]; Recanati [20]; Zalta [23]) and semanticists (e.g. von Fintel & Heim [9]) treat **In** and **Acc** on a par, i.e. (2a) and (2b) receive the same truth conditions.

One of the main objectives of this paper is to establish that there is in fact a relevant semantic difference between **In** and **Acc**. These differences have probably remained largely unrecognized or glossed over because semanticists of fiction traditionally focus on providing analyses for reports on the content of *fictional* media (i.e. parafictional statements) only. Instead, I adopt a broader perspective and consider reports on the content of media whether fictional or non-fictional – or 'contensive statements'. First, I propose that **In** receives the widely adopted Lewisian [16] possible world analysis (section 2.1). Roughly: '**In** s, ϕ' is true iff in worlds compatible with s, ϕ. In line with Krawczyk's [15] analysis of 'According to s', contensive statements with **Acc** are analysed as indirect speech reports (section 2.2). Roughly: '**Acc** s, ϕ' is true iff s asserts that ϕ. Second, I will explore three novel observations concerning the divergent linguistic behaviour of **In** and **Acc** that a uniform treatment of the operators *cannot* but that the proposed semantic analyses *can* explain. These observations add to existing observations in recent linguistic literature that show that there is a crucial difference between **Acc** and other intensional operators (Krawczyk [15], Kaufmann & Kaufmann [13]). The novel observations relate to the fictionality of the medium that is reported on (section 3.1), reporting explicit and implicit content (section 3.2) and tense use in contensive statements (section 3.3).

2 Semantic analyses

2.1 True in the worlds of the story

First, a simplified representation of Lewis' [16] possible worlds analysis of **In**:

$[\![\text{In } s, \phi]\!] = 1$ iff in all possible worlds compatible with s, ϕ

I adopt Lewis' final analysis (analysis 2) of fictional truth according to which the possible worlds that are compatible with a story s are the worlds where s is told as known fact that are most similar to – or in Lewis' terms 'closest to' – our conception of the actual world. Here 'our conception of the actual world' consists in the overt beliefs in the community of origin of the relevant fiction, i.e. beliefs that are generally and openly shared. For instance, it is true in *The Lord of the Rings* that Frodo lives in the Shire because in worlds where *The Lord of the Rings* is told as known *fact* this is true. In addition, it is true in *The Lord of the Rings* that water is H_2O because we believe this to be actually true and nothing in *The Lord of the Rings* contradicts it. Hence worlds where *The Lord of the Rings* is told as known fact that are closest to our conception of the actual world are worlds in which water is H_2O.

2.2 What the story asserts

Contensive statements that feature the operator **Acc** are analysed as a type of indirect speech report, i.e. reports on what a medium asserts[2]:

$$[\![\text{According to } s,\ \phi]\!] = 1 \text{ iff } s \text{ asserts that } \phi$$

This analysis of contensive statements with **Acc** is in line with Krawczyk's [15] and Kaufmann & Kaufmann's [13] analysis of the general (i.e. also outside of contensive statements) use of the phrase 'According to *s*'. These semanticists treat **Acc** not as a simple intensional operator (cf. von Fintel and Heim [9]) but rather treat statements with this phrase as indirect speech reports. Indeed, such an analysis fits the use that **Acc**, unlike **In**, has outside of contensive statements; **Acc** can be used to report not only on the content of a medium but also on what some *person* asserted:

(7) a. According to Joe, seagulls are the worst.
b. # In Joe, seagulls are the worst.

As Anand and Korotkova [4] note, such reports behave like regular indirect speech reports. For instance, whereas belief reports can be followed by a denial of the embedded content having been said, speech reports cannot:

(8) a. Joe thinks that seagulls are the worst. He never said that, though.
b. # Joe asserted that seagulls are the worst. He never said that, though.

Likewise, (7a) cannot be followed by a denial of the embedded content having been said:

[2] It is possible to formulate the analysis with speech verbs that are similar in meaning such as "say". I use "assert" because I want to restrict the analysis to reports on speech acts that are clearly commitment inducing (see Anand & Hacquard [3]). Possibly, "say" is too generic (e.g. fictional statements or presuppositions may be *said* but are not *asserted*). I leave an investigation into the exact differences between "say" and "assert" to future research.

(9) # According to Joe, seagulls are the worst. He never said that, though.

Anand and Korotkova [4] argue that this analysis of **Acc** can not only apply to speakers but also to inanimate objects as long as they are repositories of propositional information (or 'ROI subjects', see Anand et al [2]) such as books, theories, films or lecture notes. Hence **Acc** can feature in contensive statements which as a result are interpreted as reports on what some medium (e.g. *The Lord of the Rings* or a news report) asserts – rather than reports on what the author of the medium asserts.[3]

Before moving on it is instructive to highlight two features of the speech act of assertion. First, since assertions are non-fictional statements, when a asserts ϕ this means that a states that ϕ is true in the *actual world*, i.e. a communicates that the actual world is in the set of ϕ worlds. Likewise, when some medium is reported on as making an assertion, this means that it is approached as stating something about the actual world. In other words, it is reported on as if it is non-fiction (cf. Murday [18] who argues that use of **Acc** in parafictional statements relates the content of the fictional narrative to the actual world).

Second, indirect speech reports are, unlike simple intensional operators, generally not closed under logical entailment. For instance, suppose Anne asserts/says/claims/yells/mutters/whispers that Chrissy is cool and suppose that Chrissy being cool implies that there is at least one cool person in town. Does it follow that Anne asserts/says/claims/yells/mutters/whispers that there is at least one cool person in town? As Maier [17] notes, for many 'descriptive communication verbs' (e.g. yells/mutters/whispers) the entailment is off but for less descriptive verbs (e.g. say/assert/claim) the entailment will sometimes seem acceptable. For the latter type of verbs we can follow von Stechow & Zimmerman's [22] suggestion to analyse indirect speech reports with "say" as ambiguous between a strict reading – where they are *not* closed under entailment – and a non-strict reading – where they *are* closed under entailment. In the above semantic analysis of **Acc** "asserts" is to be read on a non-strict reading, i.e. s asserts that ϕ iff s explicitly states ϕ or ϕ is entailed by what s explicitly states.

3 Linguistic observations concerning In and Acc

Now that I have presented my semantic analyses of **In** and **Acc**, I turn to three linguistic observations concerning the diverging linguistic behaviour of **In** and **Acc** (and some qualifications to them). Current analyses of contensive statements do not distinguish **In** from **Acc** and therefore do not explain these observations. I will argue that the analyses proposed above can account for them.

3.1 Fiction/non-fiction

A central observation concerning **In** and **Acc** is that whereas contensive statements about fiction can be formulated with both **In** and **Acc**, contensive state-

[3] This semantic definition is akin to Zalta's [24] analysis of parafictional statements in general (with **In** or **Acc**) as reporting on what a fictional narrative asserts.

ments about non-fiction with **In** rather than **Acc** are typically unacceptable. Consider the following minimal pairs of statements:

(10) a. In the *Star Wars* saga, Darth Vader is a Sith Lord.
b. ? According to the *Star Wars* saga, Darth Vader is a Sith Lord.

(11) a. # In *Ludwig Wittgenstein: The Duty of Genius*, Wittgenstein was Austrian.
b. According to *Ludwig Wittgenstein: The Duty of Genius*, Wittgenstein was Austrian.

Whereas use of **Acc** seems appropriate to report on the content of fictional and non-fictional media, use of **In** seems restricted to reports on the content of fictional media. Even stronger, even though **Acc** can be (and is) used to report on the content of fictional media, often **In** will be more appropriate, e.g. (10a) and (10b) are both acceptable but (10a) is a more natural way of talking about the content of the *Star Wars* films. Thus the general picture that is sketched is that the canonical use of operators links **In** to fiction and **Acc** to non-fiction.

The observation made above can be qualified in several ways. For instance, use of **In** is not in fact unequivocally wrong for contensive statements about non-fictional media.[4] Zucchi provides the following example of a contensive statement featuring **In** about Woodward's biography *Shadow*:

(12) a. In *Shadow*, Clinton only cares about sex and golf. [25, p.350]
b. According *Shadow*, Clinton only cares about sex and golf.

Not only use of **Acc** but also use of **In** is acceptable in this non-fiction contensive statement. However, I argue that such use of **In** is restricted to reports on subjective viewpoints or portrayals that are expressed by some medium rather than objective facts. Use of **In** here seems to signal distancing from the reported content. Likewise, a contensive statement with **In** that reports on an objective fact expressed by *Shadow* sounds as odd as (11a):

(13) a. # In *Shadow*, Clinton is the 42nd president of the U.S.
b. According to *Shadow*, Clinton is the 42nd president of the U.S.

The provided analyses account for these observations. First, when we report on the content of some non-fictional source s (e.g. a biography, news report or encyclopedia entry), we will report on the medium as telling us (or asserting) something about the actual world – not as some story that is compatible with some set of worlds that may or may not include the actual world. Hence we have a strong preference for **Acc** in contensive statements about non-fiction. By contrast, when talking about the content of a fictional medium s it is appropriate to consider what is true in the set of s worlds without reporting on s as asserting anything about the actual world. Hence **In** is appropriate whereas use of **Acc** (i.e. reporting on the content of a fiction story as if it relates to the actual world)

[4] Moreover, sometimes **In** may even seem *less* appropriate than **Acc** for reports on the content of fictional media, e.g. when the embedded content of the contensive statement is to be taken as also *actually* true (see section 4).

is less natural. Thus there is a general preference to use **In** for reports on fiction and to use **Acc** for reports on non-fiction.

As I have shown, however, although there may be a preference for **In**, **Acc** is in fact generally acceptable for contensive statements about fiction (e.g. (2b), (10b)). The semantic analysis of **Acc** suggests that this is because it is often considered appropriate to talk about the content of a fictional medium by reporting on it as if it is non-fiction, i.e. as something that asserts something about the actual world. Hence the analysis provided is in line with Friend's [11] claim that *all* fictional narratives are essentially to be interpreted as being about the actual world, even when the described events take place in an outlandish magical realm where for instance Earth does not even exist.

The analyses also account for the fact that sometimes **In** may be appropriate for contensive statements about non-fiction as in (12a). According to our semantic analysis of **In**, (12a) roughly means that in the worlds compatible with *Shadow*, Clinton only cares about sex and golf. In other words, the medium is not presented as telling us something about the actual world. Rather, because we are reporting on subjective content it is acceptable to report on what the worlds compatible with the medium are like (i.e. report on *Shadow* as if it is fiction). The perceived distancing from the reported content by the speaker of the contensive statement seems to be the result of pragmatic implication (i.e. given that the relevant medium is non-fictional, why doesn't the speaker report on its content as asserting something about the actual world?)

3.2 Explicit/implicit content

The second observation about the difference between **In** and **Acc** relates to whether the reported content is explicit or implicit in the medium. Semanticists of fiction often assume some version of Lewis' [16] Reality Principle (see Franzén [10] for an in depth discussion): We assume the fictional worlds to be as much like the actual world as the story permits. In other words, we can distinguish two types of fictional truths: 'Explicit fictional truth', i.e. propositions that are explicitly stated in a story (or follow directly from what was explicitly stated) and 'implicit fictional truth', i.e. propositions that are assumed to be fictionally true because we consider them to be actually true and the story has not forced us to revoke them. For instance, it is explicit fictionally true in *The Lord of the Rings* that Frodo inherits Bag End because this follows directly from some of the statements in the novels. On the other hand, it is implicit fictionally true in *The Lord of the Rings* that water is H_2O because we believe this to be actually true and nothing in the novels contradicts this information.

Semanticists of fiction generally allow for both implicit and explicit fictional truths to feature in parafictional statements. This type of approach ignores important differences in linguistic behaviour between **In** and **Acc**. Whereas **In** is appropriately used to report on both implicit and explicit fictional truth, **Acc** can only appropriately be used to report on explicit fictional truth. Consider the following statements:

(14) a. In *The Lord of the Rings*, Frodo inherits Bag End.
b. According to *The Lord of the Rings*, Frodo inherited Bag End.

(15) a. In *The Lord of the Rings*, water is H_2O.
b. # According to *The Lord of the Rings*, water is H_2O.

Use of **Acc** is thus restricted to parafictional statements that report content that is explicitly stated in the medium or follows directly from what was stated.

This observation generalizes to contensive statements about non-fiction. Consider the following contensive statements about a news report that reports on a drought (but does not state anything about the molecular structure of water):

(16) According to this news report, there was a terrible drought.

(17) # According to this news report, water is H_2O.

Although the fact that water is H_2O may be assumed to be true (by speaker and addressee alike) when engaging with this news report, such ‘implicit truths’ cannot feature in contensive statements with **Acc**. Again, **Acc** is only appropriate to report on what was explicitly stated in the medium or what follows directly from this.

The proposed analyses can account for the above observations concerning implicit and expicit content. First, the Lewisian analysis of **In** was formulated so as to include implicit fictional truths. The worlds compatible with s are the worlds where s is told as known fact that are *as similar as possible* to our conception of the actual world. In other words, everything that we believe to be actually true will be true in the worlds compatible with f unless f contradicts it. So even though the fact that water is H_2O is never stated explicitly (nor follows from anything that was stated) in *The Lord of the Rings*, still it is true in the worlds compatible with *The Lord of the Rings* because the worlds where *The Lord of the Rings* is told as known fact that are *closest* to our conception of the actual world are worlds in which water is H_2O. Thus **In** can appropriately be used to report on such implicit content.

Second, the analysis of contensive statements with **Acc** as indirect speech reports excludes reports on implicit content. Under the non-strict reading that we adopt of “asserts” in the semantic definition of **Acc**, s asserts only those things that are explicitly stated by s or follow from what is explicitly stated. Information that is merely assumed by s but that is neither said nor entailed by what was said cannot feature in indirect speech reports (e.g. From the fact that Anne asserts that Chrissy is cool we cannot derive that Anne asserts that Chrissy plays basketball even though it may be common ground that she does). Likewise, it is not appropriate to report on ‘content’ that was not stated explicitly (or follows from what was stated) in some medium (e.g. *The Lord of the Rings* or a news report on a drought) with **Acc** even though this information may arguably be part of what is assumed to be true by the medium. [5]

[5] This semantic difference between **In** and **Acc** suggests that the proper parafictional testcase sentences to check our intuitions against about fictional truth should be formulated with **In** rather than **Acc**.

3.3 Tense use

The third and last observation concerning **In** and **Acc** that I will discuss relates to tense use preferences in contensive statements. As has been observed by Zucchi [25], parafictional statements with **In** display a preference for present tense use while past tense, although often acceptable, sounds awkward and future tense simply sounds wrong.[6] Parafictional statements with **In** trigger this preference for present tense independently from whether the embedded content includes an eventive or stative verb and independently from when the events described in the fictional narrative supposedly take place. Consider for example the following contensive statements about the Harry Potter novels, the *Star Wars* saga and the *Star Trek* series for which the time of the relevant fictional events and states described respectively overlap, precede and succeed the fictional counterpart of the utterance time of the contensive statement:

(18) In the Harry Potter novels, there **are/? were/# will be** wizards in England. (*stative/overlap*)

(19) In the *Star Wars* saga, Luke **destroys/? destroyed/# will destroy** the Death Star. (*eventive/precede*)

(20) In the *Star Trek* series, Earth **conolizes/? colonized/# will colonize** Mars in the year 2103.(*eventive/succeed*)

This preference for present tense does not generalize to parafictional statements with **Acc**. Rather, preferences for tense use within these statements seems to depend on the time of the events described in the narrative relative to the utterance time of the contensive statement, i.e. whether, at the time of utterance, the fictional events took, take or will take place:

(21) According to the Harry Potter novels, there **are/# were/# will be** wizards in England. (*stative/overlap*)

(22) According to the *Star Wars* saga, Luke # **destroys/destroyed/# will destroy** the Death Star. (*eventive/precede*)

(23) According to the *Star Trek* series, Earth # **colonizes/# colonized/will colonize** Mars in the year 2103. (*eventive/succeed*)

In fact, this is true for contensive statements with **Acc** in general, i.e. tense use in contensive statements with **Acc** about non-fictional media also seems to depend on the time of the events described in the medium relative to the utterance time of the contensive statement. Consider tense use in the following statements about the content of news reports that report on respectively protests going on at this moment, a robbery last night and tomorrow's weather:

[6] The prohibition against past and future tense in parafictional statements is not absolute. Consider for instance: "In Patrick O'Brian's first novel, Jack Aubrey was a post captain, in his new novel, he is a commodore, in the next novel he will be an admiral".

(24) According to this news report, there **are/# were/# will be** protests in Amsterdam.

(25) According to this news report, masked men **# rob/robbed/# will rob** the Regio Bank in Erp.

(26) According to this weather forecast, it **# is/# was/will be** extremely dry.

The proposed semantic analyses can account for these observations. First, the analysis of **In** predicts a preference for present tense in contensive statements.[7] To see why let's first consider tense use in other intensional operators such as believe:

(27) Adeela believes that Sara was nervous.

Because this propositional attitude report reports on a current belief (i.e. the attitude verb is in present tense), the tense use in the embedded clause tells us whether Adeela believes Sara to be nervous before, during or after the time of utterance of (27).[8] In the above example: if (27) is uttered at t_1 then (27) is true iff in worlds compatible with what Adeela believes at t_1, Sara *was* nervous at t_1 (i.e. *is* nervous at some t_x where $t_x < t_1$).

In, although also an intensional operator, functions somewhat differently. Whereas someone's beliefs may change over time (e.g. Adeela might change her mind about whether Sara is in fact nervous), the content of a story or medium (e.g. the Harry Potter novels) consists in an abstract set of statements or system of axioms that is timeless. The Harry Potter story today is not going to differ from the Harry Potter story tomorrow; It is eternally the same abstract object. Hence, although we report on what some agent's beliefs are at some time in (27), we do not report on what the Harry Potter novels are like at a certain point in time in contensive statements. Reconsider the present tense version of (18):

(18) In the Harry Potter novels, there are wizards in England.

Even though (18) is uttered at a specific point in time t_1, (18) does not mean that in worlds compatible with the Harry Potter novels at t_1, wizards are in England at t_1. Rather, (18) uttered at t_1 is true iff in worlds compatible with the Harry Potter novels at some t_x, there are wizards in England at t_x. Hence, because it is true that there are wizards in England at a specific point on the fictional timeline of the Harry Potter novels, (18) is true. Indeed, given this fact, the past and future tense versions of (18) are also strictly true:

(28) In the Harry Potter novels, there were/ will be wizards in England.

[7] See Zucchi [25] for an alternative possible world analysis of **In** that accounts for this present tense preference.

[8] Reports with past or future tense attitude verbs (e.g. 'Adeela believed/will believe that Sara is nervous') pose additional complications since tense in these reports does not necessarily switch the time of events as it does outside of these contexts (see Abusch [1]; Ogihara & Sharvit [19]).

It is true on some point in the timeline of the Harry Potter worlds that there *were* wizards in England and similarly there is such a point where there *will be* wizards in England. I suggest that the acceptability of all three tenses licenses a gnomic or generic use of the present tense (see e.g. Carson [5]) that is similar to that in scientific statements that express timeless truths (e.g. The fact that whales are, were and will be mammals can be expressed as "Whales *are* mammals"). The same reasoning applies to contensive statements with **In** that report on fictions about past or future events (e.g. it is true at some point on the fictional timeline of *Star Wars* that Luke destroys the Death Star) and hence these will also display a preference for present tense.

Second, the proposed analysis of **Acc** accounts for tense use in contensive statements with this operator. Contensive statements with **Acc** are analysed as indirect speech reports (i.e. on what a medium 'asserts'). Hence tense use in such contensive statements mirrors that of indirect speech reports. If an indirect speech report reports on a 'current' speech event (i.e. the speech verb is in present tense), then the tense use in the embedded clause mirrors that of the reported speech act. The reported speaker's tense use in turn depends on whether the time of the events described coincides, precedes or succeeds the utterance time of her statement, i.e. whether she is telling us what things *are*, *were* or *will be* like. Hence, tense use in indirect speech reports on current speech events shifts depending on whether the time of the described events coincides, precedes or succeeds the utterance time of the contensive statement. For instance, if Adeela says "Sara will be nervous" at t_1, a speech report at t_1 will mirror her tense use:

(29) Adeela asserts that Sara will be nervous.

(29) uttered at t_1 is true iff Adeela asserts at t_1 that Sara will be nervous at t_1 (i.e. *is* nervous at some t_x such that $t_x > t_1$).[9]

A contensive statements with **Acc** is analysed as a report on what a medium *asserts*. Hence it is a report on a current speech event.[10] In other words, unlike contensive statements with **In**, contensive statements with **Acc** are essentially time bound; They report on what the medium asserts *now*. Likewise, tense use in contensive statements with **Acc** shifts depending on whether the events described by medium overlap, precede or succeed the utterance time of the contensive statement. For example, since the *Star Wars* saga is about events that supposedly took place a long time ago (in a galaxy far, far away), we report

[9] I assume a simple analysis of "will" as a tense marker. (e.g. Kissine [14], Salkie [21]). Under a modal analysis (Abusch [1]; Condoravdi [6]; Enç [8]) "will" still has a temporal dimension and hence a modal analysis can also be incorporated into my analysis.

[10] A complication for this comparison to indirect speech reports is that whereas the speech report about Adeela's assertion mirrors her tense use, tense use in contensive statements does not necessarily mirror the tense use in the medium itself. For instance, although a science fiction novel may be written from the point of view of the year 4020 and include the past tense statement "Mars was inhabited in 3020", it currently (in 2020) asserts that Mars *will be* inhabited in 3020.

on its content using past tense, e.g. *Star Wars* asserts that Luke *destroyed* the Death Star. Hence (22) displays a preference for past tense:

(22) According to the *Star Wars* saga, Luke destroyed the Death Star.

(22) uttered at t_1 is true iff *Star Wars* asserts at t_1 that Luke *destroyed* the Death Star at t_1 (i.e. *destroys* the Death Star at some t_x such that $t_x < t_1$). Likewise, since a medium like the news report on protests in Amsterdam in (24) reports on events that currently take place and the *Star Trek* series is (amongst other things) about events that supposedly will take place in the future, we report on the content of these media using respectively present and future tense.

4 Conclusions and further research

In this paper I have argued that the **In** and **Acc** operators require separate semantic analyses to account for three linguistic observations. These concern preferences for using **In** for contensive statements about fiction and **Acc** for non-fiction; the unacceptability of using **Acc** to report on implicit content (whereas **In** is fine for implicit and explicit content); and preferences for present tense in contensive statements with **In** and tense use in contensive statements with **Acc** depending on whether the events described by medium overlap, precede or succeed the utterance time of the contensive statement.

I have proposed to adopt the Lewisian possible world analysis of parafictional statements for contensive statements with **In**. Roughly: '**In** s, ϕ' is true iff in the worlds compatible with s, ϕ. I have proposed to analyse contensive statements with **Acc** as indirect speech reports. Roughly: '**Acc** s, ϕ' is true iff s asserts that ϕ. Lastly, I have explained how the proposed analyses account for the three described linguistic observations.

A potential direction of future research is to see whether the three linguistic observations described in this paper can be experimentally confirmed. Moreover, it is an open question to what extent the observations generalize to other languages. Another potential direction of future research is to relate this discussion to the literature on 'export' of fictional truth, i.e. learning truths (empirical facts such as 'Nassau is the capital of the Bahamas' or general truths such as 'love conquers all') about the actual world from fiction. A possible explanation of export (e.g. Currie [7], Carcía-Carpintero [12]) is that these truths are (indirectly) asserted in the fictional narrative. This analysis suggests an increase in acceptability of **Acc** in contensive statements that report on fictional truth viable for export (e.g. 'According to the Harry Potter novels, love conquers all').

References

1. Abusch, D.: Sequence of Tense and Temporal De Re. Linguistics and Philosophy, **20**(1) pp. 1-50 (1997)
2. Anand, P., Grimshaw, J., Hacquard, V.: Sentence embedding predicates, factivity and subjects. In: Condoravdi, C. (ed.), Lauri Karttunen FestSchrift. CSLI (2019)

3. Anand, P., Hacquard, V.: Factivity, belief and discourse. In: Crnic, L., Sauerland, U. (eds.) The Art and Craft of Semantics: A Festschrift for Irene Heim. pp. 69-90 (2014)
4. Anand, P., Korotkova, N.: Speech reports (lecture notes) (2019)
5. Carson, G. N.: Generic terms and generic sentences. Journal of Philosophical Logic **11** pp. 145-181 (1982)
6. Condoravdi, C.: Temporal interpretation of modals. Modals for the present and for the past. In: Beaver, D.I., Martinez, L.D.C., Clark, B.Z., Kaufmann, S (eds.) The construction of meaning. pp. 59-88, Stanford: CSLI Publications (2002)
7. Currie, G.: The nature of fiction. Cambridge University Press, Cambridge (1990)
8. Enç, M.: Tense and modality. In: Lappin, S. The handbook of contemporary semantic theory. pp. 345-358,Oxford: Blackwell (1996)
9. Fintel, K., Heim, I.: Intensional Semantics (lecture notes) (2011)
10. Franzén, N.: Truth in fiction: In defense of the Reality Principle. In: Maier, E., Stokke, A. (eds.) The Language of Fiction, Oxford:OUP (forthcoming)
11. Friend, S.: The Real Foundation of Fictional Worlds. Australasian Journal of Philosophy **95**(1) pp. 29-42 (2017)
12. García-Carpintero, M.: Assertions in Fictions. Grazer Philosophische Studien **96**(3) pp. 445-62 (2019)
13. Kaufmann, M., Kaufmann, S.: Talking about Sources. in Proceedings of NELS 50 (forthcoming)
14. Kissine, M.: Why will is not a modal. Natural Lang Semantics **16**, pp. 129-155 (2008)
15. Krawczyk, E. A.: Inferred Propositions and the Expression of the Evidence Relation in Natural Language. PhD thesis, Georgetown University (2012)
16. Lewis, D.: Truth in fiction. American Philosophical Quarterly **15**(1) pp. 37-46 (1978)
17. Maier, E.: On the exceptionality of reported speech. Linguistic Typology **23**(1) pp. 197-205 (2019)
18. Murday, B.: Two-Dimensionalism and Fictional Names. In: Lihoreau, F. (ed.) Truth in Fiction, pp.43-76 (2011)
19. Ogihara T., Sharvit, Y.: Embedded tenses. In: Binnick, R. I. (ed.) The Oxford Handbook of Tense and Aspect, Oxford: OUP (2012)
20. Recanati, F.: Fictional, metafictional, parafictional. Proceedings of the Aristotelian society **118**(1) pp. 25–54 (2018)
21. Salkie, R.: Will: Tense or modal or both? English Language and Linguistics **14**(2), pp. 187-215 (2010)
22. von Stechow, A., Zimmerman, E.: A problem for a compositional treatment of de re attitudes. In: Carlson, G., Pelletier, F. (eds.) Reference and Quantification: The Partee Effect pp 207–228, CSLI Publications, Stanford (2005)
23. Zalta, E.N.: Abstract objects: An introduction to axiomatix metaphysics. Springer, New York (1983)
24. Zalta, E.N.: Erzählung als taufe des helden: Wie man auf fiktionale objekte bezug nimmt. Zeitschrift fur Semiotik **9**(1-2) pp. 85-95 (1987) [2003]
25. Zucchi, S.: Tense in Fiction. In: Cecchetto, C., Chierchia, G., Guasti, M.T. (eds.) Semantic Interfaces: Reference, Anaphora and Aspect, pp. 320-355. CSLI Publications, Stanford (2001)
26. Zucchi, S.: On the generation of content. In: Maier, E., Stokke, A. (eds.) The Language of Fiction. Oxford University Press, Oxford (forthcoming)

A gamified approach to teaching the expression of emotion in English to native speakers of tone and stress languages

Natalie Mosseri, Jenny Ortega

CUNY Kingsborough Community College, CUNY Brooklyn College
Mosserinatalie@gmail.com, Ortegajenny808@gmail.com

Abstract. There are few studies which have explored the way paralinguistic information varies across different languages, specifically languages of tone versus languages of stress. It has been shown that speakers of tone use less F0 related cues in the production of verbal expressions of emotions than speakers of stress do. This study proposes a new method of teaching English intonation patterns, specifically those that aid in the expression of emotion, to speakers of tone languages. A total of twelve participants were tested, all who spoke English as their second language, eight whose first language was tone based. One portion of the participants reacted to and mimicked an exaggerated group of recordings while the other portion reacted to the non-exaggerated version of these same recordings. Two scoring methods were used during the analysis portion of our experiment. The first was a visual scoring process. Each of the participants' pitch contours were compared directly against the model's using the Praat software. The second method was a perceptual one. During this scoring process we had the two scorers listen to the participants' recordings and both indicated the emotion they perceived for each one. Our results show that the participants who responded to the exaggerated version of the model's recordings received more accurate and consistent perceptual ratings. Anger was the most difficult emotion for the participants to imitate and for both scorers to recognize. Identifying effective language learning strategies will have important implications in language teaching.

Keywords: Suprasegmentals · Tonal languages · Stress timed languages · Praat · Intonation · Motherese · EFL · ESL · Second language acquisition.

1 Introduction

Tonal languages such as Mandarin, Chinese and Vietnamese rely heavily on aspects of tone in order to convey word meaning. Stress timed languages such as English or Russian do not use tone to determine meaning. Rather, these languages rely not on the pitch of the syllables being spoken, but on the stress placed on those syllables. Syllable length holds paralinguistic information, for instance by conveying which emotions are experienced by the speaker. Tone languages use contrastive pitch specification at every level of the phonological hierarchy compared to non-tonal languages that have gaps in contrastive use of pitch at the segmental level [2].

To acquire English as a second language (L2), the mastery of intonation units, stress, tone and pitch ranges is necessary, as these are all essential features of the language [4]. Even when a native speaker of a tonal language goes on to learn a stress timed language and becomes linguistically proficient, it can still be difficult for them to properly express their emotions [5]. It has been reported that tone language speakers use less F0 related cues in the production of verbal expressions of emotions [Annoli et al. 2008, Chong et al. 2015]. This restricted pitch makes it difficult for native speakers of stress timed languages to properly comprehend the intended emotions of native speakers of tonal languages. A restricted pitch range has also been associated with a variety of psychological conditions such as depression and schizophrenia [Ellgring Scherer 1996] and can often be misinterpreted. This can result in reduced responsiveness or a negative response on the part of conversational partners of ESL speakers of tonal languages. Although English is very widespread worldwide, the paralinguistic aspects of the language are not a commonly taught feature in educational settings. Consequently, introducing such aspects in interactive teaching methods where the teacher asks students to appropriately express their emotional input, would increase the likelihood of fluency via short dialogue exchange [4].

This study proposes and explores a new method of teaching English intonation patterns specific to various common emotions to speakers of tonal languages. Acquiring these specific patterns of the target language will enhance communication between native speakers of stress timed languages and speakers of tonal languages. Teaching paralinguistic and suprasegmental patterns in ESL classrooms in an interactive manner will help ensure that emotional undertones are not lost in conversation and that speakers will not lack intelligibility due to having a restricted pitch range. Our approach is justified by the fact that exaggerated intonation shares certain similarities with Motherese / Parentese [8], which have been proven to be more enjoyable to listen to for extended periods of time as well as being easier to comprehend.

2 Methodology

2.1 Hypothesis

It has been proven that exaggerated intonation patterns share common vocal patterns with motherese. This is believed to aid non-native English speakers, specifically those whose native language is one that is tonal based, in their capability of grasping suprasegmental aspects of the English language [13]. We believe that through an imitation teaching method, one that includes exaggerated sentences combined with an appropriate emoticon, these non-native speakers will be able to mimic and produce common intonation patterns associated with emotion on their own.

2.2 Participants

For this study we tested a total of 12 participants between the ages of 19 to 45. Eight of those participants were female and the other four were male. Eight participants' native language was one that was tonal and four spoke a stress-timed language other than English as their native language (Mandarin, Vietamese, Korean, Russian, Uzbek) . All participants spoke English as their second language. While all were sufficiently familiar with the English language to be able to read the prompts given during the experiment without any assistance, their levels of fluency varied, possibly as a result of the age of first exposure to English as well as the length of time residing in the United States.

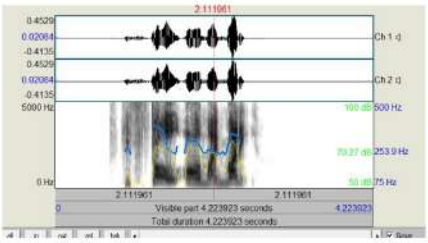
2.3 Stimuli and Procedure

The stimuli were recorded by a trained native speaker of English with professional acting experience. They consisted of two types of sentences, neutral (i.e. whose content was not associated with a specific emotion, e.g. Sally came by last night, I won the lottery!) and emotionally-charged (i.e. whose content was most consistent with one of three specific emotions: happiness, sadness, and anger). Each sentence was exactly 6 syllables long. The full list of sentences is provided in the appendix, Table 2. The speaker recorded the list twice, once using exaggerated intonation (similar to Motherese), and once in a very natural manner, without any exaggeration. The latter recordings served as a control condition.

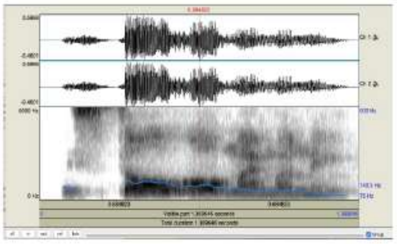
The participants were all tested individually via Zoom [14] due to the social isolation imposed to prevent the spread of Covid-19. The experimenter shared their screen with each participant and prompted them to follow the instructions, providing any necessary clarifications. The sentences were shown to the participants using the Powerpoint software. Each sentence was presented in isolation in the center of the screen, and accompanied by a smiley face displaying one of the three emotions tested: happiness, sadness, and anger. The procedure consisted of 4 distinct stages: a warm-up stage during which the participants read several sentences to get used to the format of the experiment, a baseline stage during which the participants read each of the sentences displayed as they normally would, a training stage during which they listened to the pre-recorded stimuli and were instructed to imitate them immediately after hearing them, and a testing/learning phase, in which they had to read again the sentences from the baseline block. Half of the participants were presented with the exaggerated version, and half with the natural, non-exaggerated one. The goal for the testing phase was to see if the participants would apply what they had learned during the training phase, using appropriate intonation patterns for producing emotionally-charged sentences.

2.4 Scoring Process

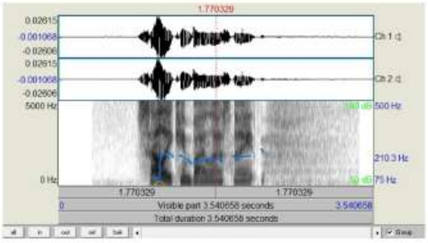
Our experiment underwent two scoring processes. First, we performed perceptual scoring. All the recordings taken from each participant during the baseline, imitation and learning process, along with all of the model speaker's recordings, were gathered and randomized before being compiled into an audio file. Two native English speakers were used as independent scorers. The first was a 21-year old female and the second scorer was an 18-year old male. Both speakers had lived in the United States their entire lives and were monolingual. Upon listening to a recording, each scorer had to select the appropriate emotion perceived in a forced-choice task presenting them with the following options: happy, sad, angry or none of the above. The scorers were asked to focus not on the words spoken by the participant, but only on the way their voices sounded. The second process was one of visual scoring, assessing the similarity between the pitch tracks of each participant during the imitation portion of the experiment to the pitch tracks of the model speaker. We visualized the pitch tracks using the Praat software [10]. The reason for doing so is to supplement the native speakers' perceptual judgements. Such judgement can be biased and certain transitory details can go unnoticed. Comparing pitch tracks visually can help pinpoint exactly how a participant and the model speaker vary in their production. The visual scoring was performed by one of the researchers. Each statement spoken by every participant in the imitation phase was given a score from 0 to 2. A score of 0 meant that the participant's pitch contour did not appear similar at all compared to the model speaker's pitch contour. A score of 1 meant that the pitch contour was roughly similar, but did diverge in some ways. The highest score given was 2, indicating very close similarity between a participant's pitch contour and the model's pitch contour.



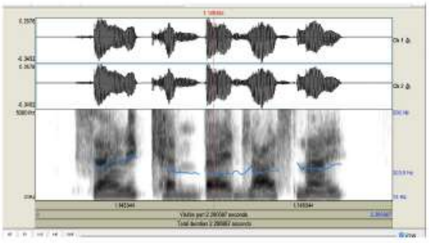
Model's *angry* exaggerated pitch contour



Subject 8 *angry* imitation pitch contour scoring a 0



Model's *sad* exaggerated pitch contour



Subject 7 *sad* imitation pitch contour scoring 1

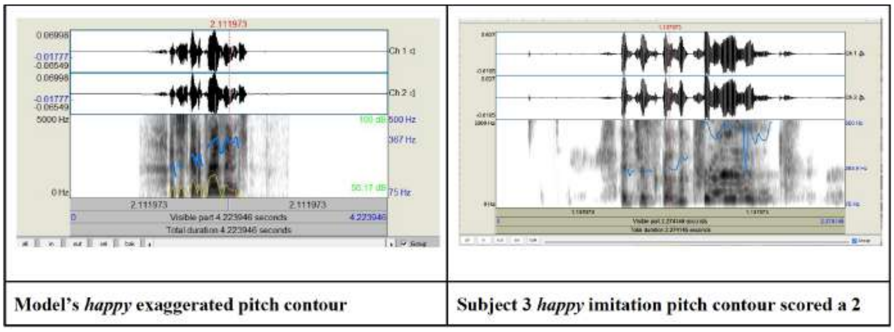

Model's *happy* exaggerated pitch contour
Subject 3 *happy* imitation pitch contour scored a 2

Fig. 1. The spectrograms showed above are from participants in the exaggerated condition, When compared to the model speaker's spectrogram and each subject score.

3 Results

We focused here on two distinct measures: *inter-rater agreement* (that is, did the raters consistently recognize the same emotion for a given sentence) and *accuracy* of perceived emotions (that is, was the emotion recognized by the scorers the same as the intended emotion). Using these data, we conducted analyses of variance to compare the means of the different groups. Because our results are preliminary, we discuss the patterns observed separately from the results of statistical analyses. (The means for the different categories are provided in Appendix A)

3.1 Visual scoring of pitch contours

The visual scores are similar to the accuracy results, suggesting that *angry* is the most difficult emotion to mimic, especially in the exaggerated condition, whereas *sad* and *happy* are easier to imitate when exaggerated. *Happy* obtained the lowest visual similarity scores in the non-exaggerated condition (Figure 5). We conducted a univariate ANOVA with the visual similarity scores as the dependent variable and condition (exaggerated/non-exaggerated), emotion (happy/angry/sad), and block (model speech/baseline/training/testing) as independent variables. Emotion significantly affected the visual scores, $F(2, 102) = 6.17$, $p < .05$ (*sad* was significantly different from *happy*) and the interaction between emotion x condition also had a significant effect, $F(2, 102) = 5.24$, $p < .05$. As the figure shows, *happy* and *sad* had higher scores in the exaggerated condition, whereas *angry* had lower scores in the same condition compared to the non-exaggerated one.

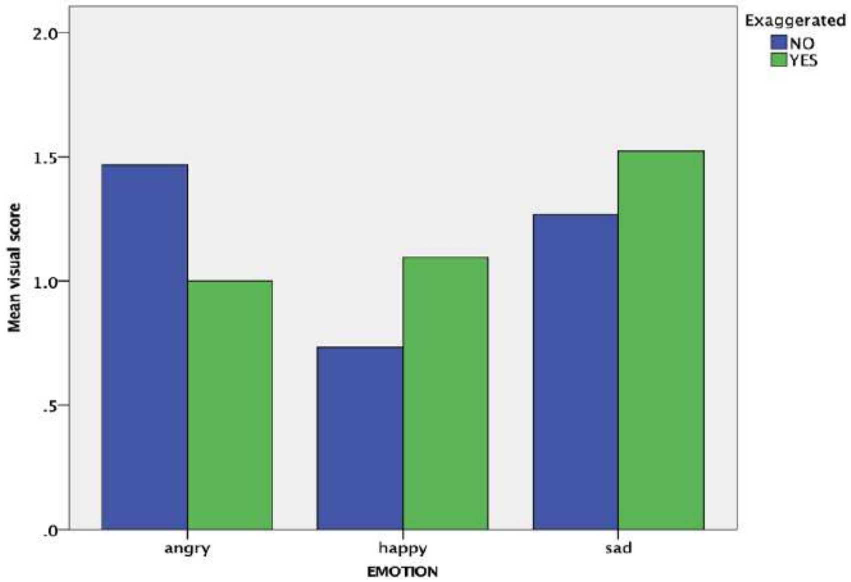


Fig. 2. Table comparing the visual scoring results between the participants who imitated the exaggerated version of the models' recordings versus the participants who imitated the non-exaggerated version.

3.2 Perceptual Scoring

With respect to inter-rater agreement, we have found that while the model native English speaker obtained over 80% inter-rater agreement, the agreement for the non-native subjects was lower across the board. However, if we only consider the Exaggerated condition, we note that the model speaker had 100% agreement. In the non-exaggerated condition, there appears to be no learning improvement in the testing phase (from 73% agreement in baseline to 62% in learning), and a large decrease is observed in the raters' agreement for the happy and sad emotions. The same does not apply in the exaggerated condition (for which the agreement in baseline was 62% and in testing 69%). This suggests that the expression of emotion may have been more consistent and easier to recognize by native speakers.

Turning to emotion accuracy (summarized in Figure 4), the model speaker obtained 78% accuracy, while the non-native speakers received less across the board, but relatively close to the model speaker (highest 75%). If we only consider the Exaggerated condition, the model speaker did have 100% accuracy. In the non-exaggerated condition, there seems to be no learning improvement for the non-native speakers (from 67% accuracy in baseline to 66% in testing). We note a large decrease in accuracy for the sad emotion, and a large increase for happy. While not very pronounced, we note some potentially positive effects of the exaggerated condition (baseline 74% and testing 81% overall). This indicates that ESL speakers' expression of emotion became more accurate as a result of the training. We also note an unexpected tendency for the participants to perform consistently worse in the imitation phase (compared to both baseline and testing). This may be due to increased cognitive demands of imitating the various correlates of emotionally-charged statements.

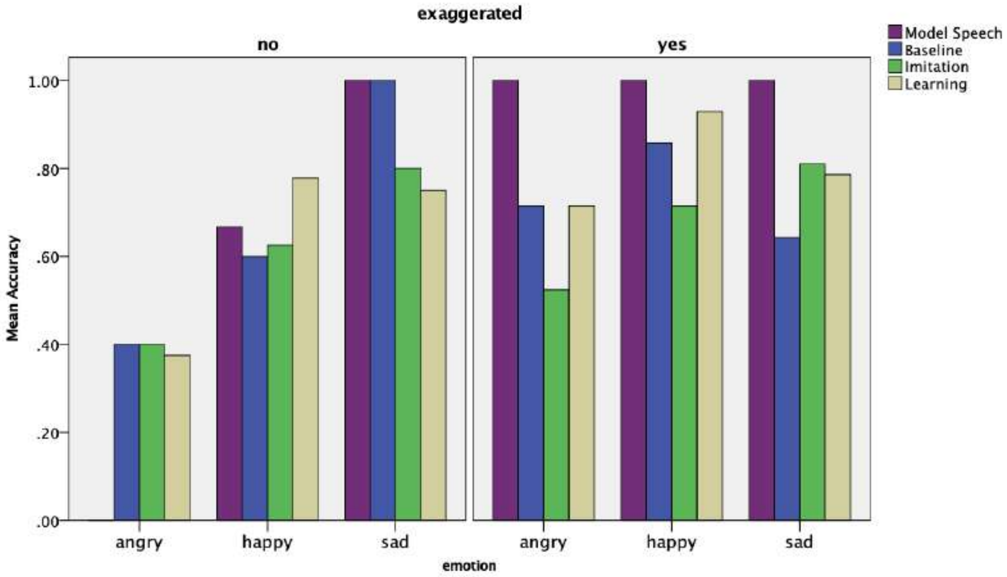


Fig. 3. Mean accuracy for the exaggerated (right) and non-exaggerated (left) conditions, broken down by emotion type. The bars are clustered by experimental phase, with the model speaker's scores also shown here for comparison.

A univariate ANOVA was conducted with accuracy as the dependent variable and condition (exaggerated/non-exaggerated), emotion (happy/angry/sad), and block (model speech/baseline/training/testing) as independent variables. The factors that had a statistically significant effect on accuracy were condition (the accuracy was higher in the exaggerated compared to the non-exaggerated condition, $F(1, 246) = 8.04$, $p < .05$) and emotion ($F(2, 246) = 8.83$, $p < .001$). Post-hoc analyses with the Bonferroni correction revealed that *angry* differed significantly from the other two emotions, but these did not differ from each other. Finally, the interaction between condition x emotion was also significant ($F(2, 246) = 5.02$, $p < .05$). *Angry* and *happy* displayed higher accuracy in the exaggerated condition, but the accuracy was worse for *sad* (Figure 3). We tested all possible interactions but found no other significant ones.

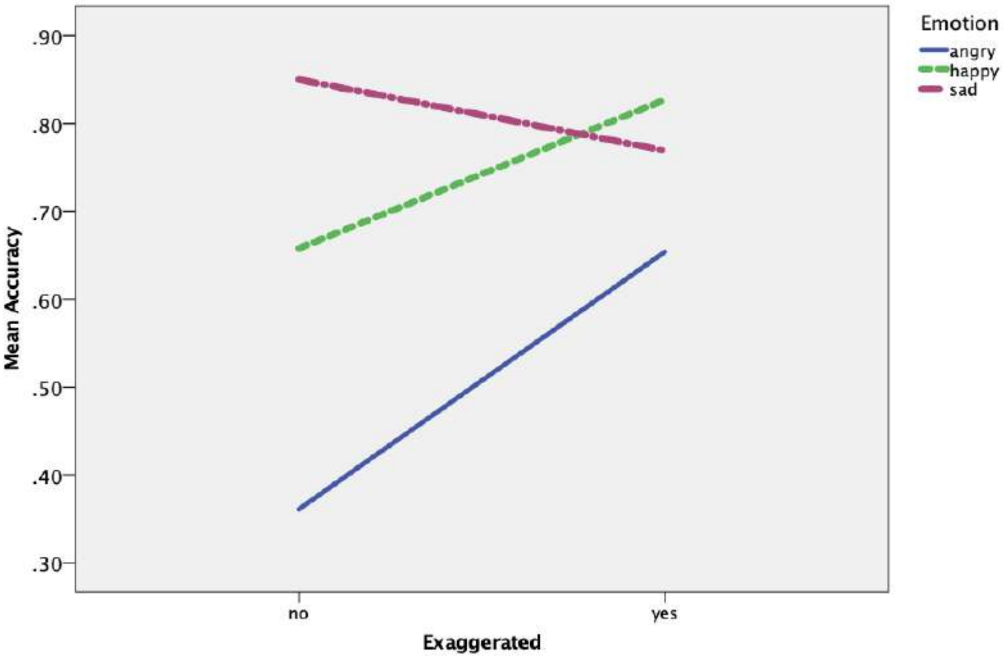


Fig. 4. Mean accuracy for the three emotions tested in the exaggerated (right) and non-exaggerated (left) condition.

4 Discussion and Conclusion

The English language relies heavily on intonation patterns when it comes to portraying emotion. A person may be familiar with the English lexicon while not having a strong grasp of the suprasegmental aspects of the language. These include elements such as tone, stress and rhythm. All spoken languages use pitch and other paralinguistic processes to convey information like emotions [2]. Without sufficient mastery of the suprasegmental aspects of English, non native speakers will be at a communication disadvantage. For speakers of tonal languages, intuitively acquiring English intonation patterns will result in different behaviors due to differences in their native language, which may be particularly detrimental to native speakers of tone languages who are used to employing pitch in very different ways compared to stress language speakers.

Software like Pratt allows instructors and non native speakers the ability to analyze and evaluate visual patterns through intonation contours, providing autonomous feedback that can eliminate the need for native speakers' judgements and aid in adequate training [9]. Developing a training method in order to teach intonation patterns of English to ESL students of tone languages could help improve both their perception of spoken English and the pragmatic uses of the language. and the appropriate use of the language based on the aquire emotion. Because early on in development children benefit greatly from being spoken to in Motherese, we sought to mimic Motherese during the exaggerated condition of our experiment and hypothesized that this would be beneficial to ESL learners as well [8]. Our hypothesis was borne out, as the participants exposed to this condition showed some improvement in the clarity of the emotions they expressed, compared to the control group (exposed to non-exaggerated speech) who did not. We did not see statistical confirmation that the exaggerated method is better for the visual scores but we did see it for the perception (accuracy) scores. We tentatively ascribe this to the difficulty of scoring visual information and also to potential program errors in pitch tracking. Despite its advantages, such as offering a quantitative, unbiased perspective on speech patterns, manual comparison of Praat contours has proven to be more difficult than perception scoring. If humans are unable to perform it successfully, automatic scoring or labeling of intonation is likely even more difficult. Future research efforts should be directed towards producing an algorithm for successfully capturing the degree of similarity between two pitch tracks. It is interesting that even though the visual scores found that the contours for angry statements were well reproduced by the participants in the non exaggerated condition, the accuracy and perception scores were low for the same emotion. One explanation may be that the participants did what they had to do to assimilate into a given environment, resulting in mimicking what they heard well. Unfortunately, the emotion itself was not recognized perceptually, which may be ascribed to cultural influences, with anger being less socially acceptable in US culture and possibly accompanied by acoustic cues that are more subtle or subdued.

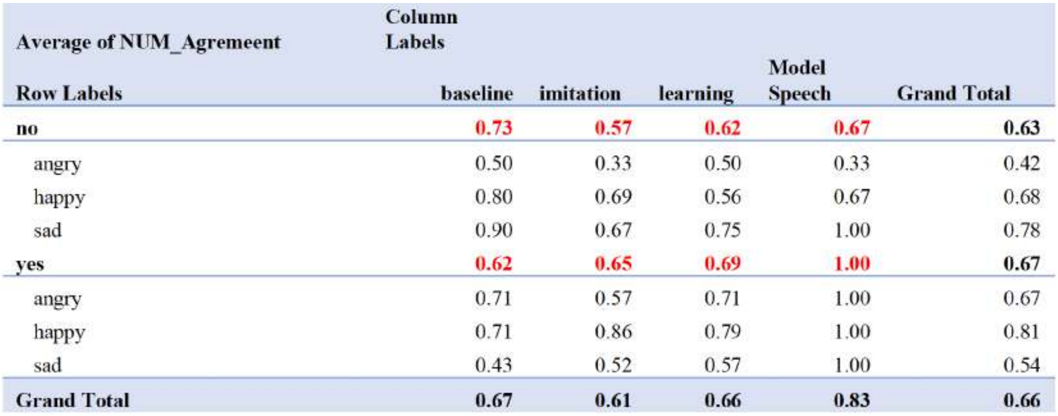
We found certain unexpected patterns during the imitation portion of the training, for example that the model speaker did not obtain 100% inter-rater agreement in the non exaggerated condition. We had expected the native speech to be scored close to 100% far as emotion recognition goes but this was not the case. This raises the question of how recognizable emotions are in our everyday speech and which of the basic emotions pose most challenges to listeners. Notably, the accuracy was 0% for the model speaker in the non-exaggerated condition for angry. As discussed above, this could potentially be due to cultural influences, as it may be less culturally appropriate in North American culture to express anger in spoken interactions [11].

To conclude, our findings are preliminary and more data are needed before drawing a strong conclusion. We have found some evidence that teaching the expression of emotion to non-native ESL learners can be aided by the use of Motherese, imitation, and visual emoticons indicating the emotions being heard. We have not however teased apart the differential contributions of these factors. This will be taken into consideration for future work [7]. A second direction we intend to pursue in the future has to do with comparing the results of tone language speakers to those of non-tone languages to determine whether the former benefit from more improvement.

References

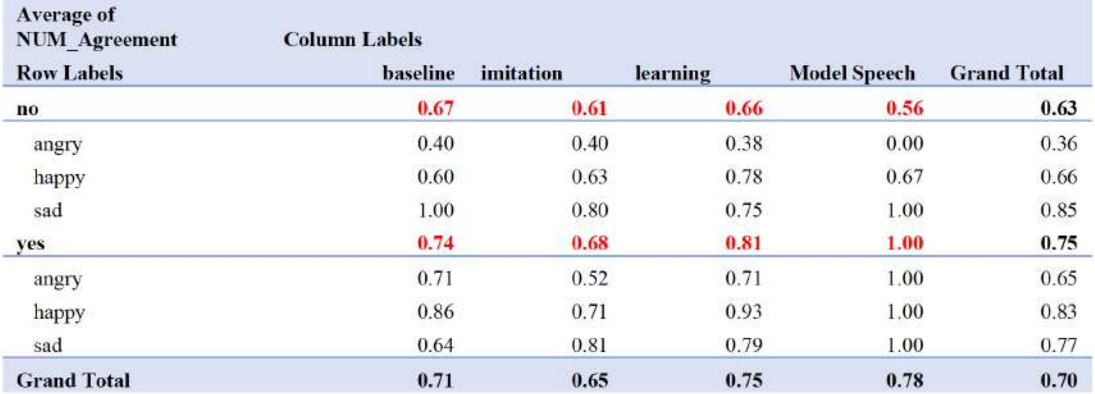
1. Tone language translates to perfect pitch: Mandarin speakers more likely to acquire rare musical ability. (2004, November 15), https://www.sciencedaily.com/releases/2004/11/041114235846.htm: :text=
2. Best, C.T.: The diversity of tone languages and the roles of pitch variation in non-tone languages: Considerations for tone perception research. Frontiers in Psychology **10**, 364 (2019). https://doi.org/10.3389/fpsyg.2019.00364, https://www.frontiersin.org/article/10.3389/fpsyg.2019.00364
3. Bänziger, T., Scherer, K.: The role of intonation in emotional expressions. Speech Communication **46**, 252–267 (07 2005). https://doi.org/10.1016/j.specom.2005.02.016
4. Celik, M.: Teaching english intonation to efl/esl students. The Internet TESL Journal **7(12)**, 261–281 (2001)
5. Chong, C., Kim, J., Davis, C.: Exploring acoustic differences between cantonese (tonal) and english (non-tonal) spoken expressions of emotions. pp. 1522–1526 (Jan 2015), 16th Annual Conference of the International Speech Communication Association, INTERSPEECH 2015 ; Conference date: 06-09-2015
6. Counihan, G.: An activity for teaching intonation awareness to esl/efl students. (Retrieved July 09, 2020), http://iteslj.org/Lessons/Counihan-Activities/Intonation.html
7. Eady, S.J.: Differences in the f0 patterns of speech: Tone language versus stress language. Language and Speech **25**(1), 29–42 (1982). https://doi.org/10.1177/002383098202500103, https://doi.org/10.1177/002383098202500103
8. Fernald, A., Kuhl, P.: Acoustic determinants of infant preference for motherese speech. Infant Behavior and Development **10**(3), 279 – 293 (1987). https://doi.org/https://doi.org/10.1016/0163-6383(87)90017-8, http://www.sciencedirect.com/science/article/pii/0163638387900178
9. Le, H.T., Brook, J.: Using praat to teach intonation to esl students. Hawaii Pacific University TESOL Working Paper Series **9(1), 2** (2011)
10. Li, Z., Lian, A.P., Yodkamlue, B.: Learning english intonation through exposure to resynthesized self-produced stimuli. GEMA Online Journal of Language Studies **20(1)** (2020)
11. Matsumoto, D., Yoo, S., Chung, J.: The Expression of Anger Across Cultures, pp. 125–137. Springer (2010). https://doi.org/10.1007/978-0-387-89676-2_8
12. Miller, E.: How to teach intonation awareness to efl students. ethan miller is a private esl tutor and apart from his passion for teaching. (2017, February 06), https://www.eflmagazine.com/teach-intonation-awareness-efl-students/
13. Valenzuela, M.: The importance of teaching intonation in efl classes; issue 4. (2014, August, Retrieved July 09, 2020), https://old.hltmag.co.uk/aug14/sart02.htm
14. Weiner, Y.: The inspiring backstory of eric s. yuan, founder and ceo of zoom. (2018, July 06, Retrieved July 09, 2020), https://medium.com/thrive-global/the-inspiring-backstory-of-eric-s-yuan-founder-and-ceo-of-zoom-98b7fab8cacc

Appendix A



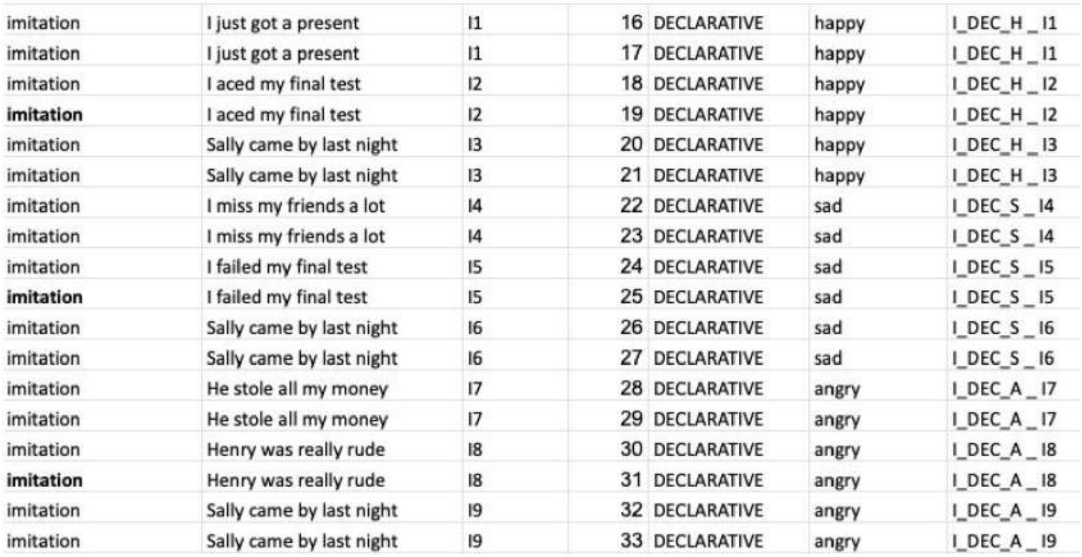
Average of NUM_Agremeent	Column Labels				
Row Labels	baseline	imitation	learning	Model Speech	Grand Total
no	**0.73**	**0.57**	**0.62**	**0.67**	**0.63**
angry	0.50	0.33	0.50	0.33	0.42
happy	0.80	0.69	0.56	0.67	0.68
sad	0.90	0.67	0.75	1.00	0.78
yes	**0.62**	**0.65**	**0.69**	**1.00**	**0.67**
angry	0.71	0.57	0.71	1.00	0.67
happy	0.71	0.86	0.79	1.00	0.81
sad	0.43	0.52	0.57	1.00	0.54
Grand Total	**0.67**	**0.61**	**0.66**	**0.83**	**0.66**

Table 1. % Inter-rater agreement



Average of NUM_Agreement	Column Labels				
Row Labels	baseline	imitation	learning	Model Speech	Grand Total
no	**0.67**	**0.61**	**0.66**	**0.56**	**0.63**
angry	0.40	0.40	0.38	0.00	0.36
happy	0.60	0.63	0.78	0.67	0.66
sad	1.00	0.80	0.75	1.00	0.85
yes	**0.74**	**0.68**	**0.81**	**1.00**	**0.75**
angry	0.71	0.52	0.71	1.00	0.65
happy	0.86	0.71	0.93	1.00	0.83
sad	0.64	0.81	0.79	1.00	0.77
Grand Total	**0.71**	**0.65**	**0.75**	**0.78**	**0.70**

Table 1b.% accuracy (match between intended and perceived emotion by both raters)



imitation	I just got a present	I1	16	DECLARATIVE	happy	I_DEC_H _ I1
imitation	I just got a present	I1	17	DECLARATIVE	happy	I_DEC_H _ I1
imitation	I aced my final test	I2	18	DECLARATIVE	happy	I_DEC_H _ I2
imitation	I aced my final test	I2	19	DECLARATIVE	happy	I_DEC_H _ I2
imitation	Sally came by last night	I3	20	DECLARATIVE	happy	I_DEC_H _ I3
imitation	Sally came by last night	I3	21	DECLARATIVE	happy	I_DEC_H _ I3
imitation	I miss my friends a lot	I4	22	DECLARATIVE	sad	I_DEC_S _ I4
imitation	I miss my friends a lot	I4	23	DECLARATIVE	sad	I_DEC_S _ I4
imitation	I failed my final test	I5	24	DECLARATIVE	sad	I_DEC_S _ I5
imitation	I failed my final test	I5	25	DECLARATIVE	sad	I_DEC_S _ I5
imitation	Sally came by last night	I6	26	DECLARATIVE	sad	I_DEC_S _ I6
imitation	Sally came by last night	I6	27	DECLARATIVE	sad	I_DEC_S _ I6
imitation	He stole all my money	I7	28	DECLARATIVE	angry	I_DEC_A _ I7
imitation	He stole all my money	I7	29	DECLARATIVE	angry	I_DEC_A _ I7
imitation	Henry was really rude	I8	30	DECLARATIVE	angry	I_DEC_A _ I8
imitation	Henry was really rude	I8	31	DECLARATIVE	angry	I_DEC_A _ I8
imitation	Sally came by last night	I9	32	DECLARATIVE	angry	I_DEC_A _ I9
imitation	Sally came by last night	I9	33	DECLARATIVE	angry	I_DEC_A _ I9

Table 2. Each sentence was exactly 6 syllables long used in the exaggerated intonation (similar to Motherese), and once in a very natural manner, without any exaggeration.

A Case Study: Analyzing and Comparing Speech Disfluency Patterns in Non-Native Dialogue

Iuliia Zaitova

Saarland University, 66123 Saarbrücken, Germany
s8iuzait@stud.uni-saarland.de

Abstract. This case study focuses on investigating the speech disfluencies in three English dialogues with three proficient non-native speakers. The aim of the study is to examine the distribution of speech disfluencies in the analyzed dialogues, and discuss the findings relative to previous research on speech fluency and disfluency.
The study participants are independent users of English, who share a similar proficiency level. The analysis is based on four measurable variables that, according to previous research, are related to perceived fluency and disfluency in dialogue (1. rate of speech – number of words per minute of speech, 2. breakdown fluency – number and length of pauses, 3. repair fluency – number of false starts, corrections, and repetitions, 4. correlation of speech disfluencies with the flow of the dialogue – number of disfluencies per one illocutionary act type).
Most disfluencies in the data occur at the beginning of a dialogue act. Major factors that may cause disfluency include difficulties in structuring the phrase, remembering a rarely used word, describing complex entities, and hesitating.
Confirming the previous findings, the analysis showed that more cognitively demanding tasks lead to higher numbers of disruptions. However, some results of the data analysis appear to contradict the previous experiments. As such, the paper found out that the fastest speaker produced the highest number of repair disfluencies, which intersects with the previously established correlation between language proficiency and rate of speech.

Keywords: Speech disfluency · Non-native speech · Dialogue.

1 Introduction

Speech disfluencies have been defined as ”phenomena that interrupt the flow of speech and do not add propositional content to an utterance.” [8, 709] Disfluencies of different kinds (silent pauses, filled pauses, corrections, and repetitions) are commonly seen as a consequence of problems with production [7, 921]. Crucially, they are a normal part of both native and non-native everyday dialogue, that is produced with no plan and prior practice. On average, according to previous research, speakers produce 5.97 disfluencies per every 100 words [3, 135].

Speech disfluencies can be produced by native speakers as slips of the tongue, strong emotions, or tiredness and also result from the limited knowledge of the language to date. Previous research shows that disfluency rates depend on the speaker's proficiency in the foreign language and their exposure to that language [10].

Despite the fact that speech disfluency is currently well understood, most research on it involved monologic speech of native speakers of the language under analysis. There has to date been little investigation of disfluency in dialogue of non-native speakers of English. Given that spontaneous speech occurs most often in the form of dialogue, the investigation of the disfluency phenomena is important in understanding how interlocutors communicate with each other.

In the current article, I present an analysis designed to examine and provide a descriptive account of the distribution and the patterns of speech disfluencies in three non-native English dialogues with three different speakers. The purpose of the study is to discuss the findings relative to previous research on speech fluency and disfluency. I begin by introducing the participants and describing the study design in Section 2. Section 3 gives an account of the obtained results, and Section 4 summarizes what has been accomplished, indicating opportunities for future research. The research described in this paper has added to the understanding of non-native disfluencies in English, and explored different aspects of disfluency phenomena.

1.1 Related Work

Essentially, the previous studies on speech disfluency examine it with regard to fluency. Tavakoli and Skehan [16] distinguish several types of fluency: speed fluency (the rate of speech), breakdown fluency (relating to pausing), and repair fluency (the extent to which speech is repeated, reformulated, or left incomplete). Cucchiarini et al. [6] explores the relationship between objective properties of speech and perceived fluency in read and spontaneous speech with the aim to determine how such quantitative measures can be used. Bosker et al. [4] investigates the contributions of pause, speed and repair parameters to speech fluency.

The research on disfluency/fluency assessment of this kind, besides giving an insight into human speech characteristics, potentially also plays an important role in the development of a qualitative testing instrument, especially when it relates to foreign language learning, teaching, and testing [6].

2 Methodology

2.1 Participants

The current case study uses data that was obtained from speech recordings of conversational dialogues with three non-native speakers of English pursuing their university degree (Master's and Ph.D.). All the three students originally come from different countries and have a different first language (Speaker 1 –

Bengali, Speaker 2 – Polish, and Speaker 3 – Spanish)[1]. All the speakers produced spontaneous speech in English in the form of conversational dialogue with the same person (the Interlocutor), who is also a non-native speaker of English (her first language is Russian).

All the subjects have a similar English proficiency level. All of them, according to their words, starting from at least six months before the experiment, have had daily exposure to communication (both listening and production) in the English language. All the subjects are currently doing their University studies in English as well. S1 started learning English in her kindergarten, and also received her Secondary and Higher Education in English. S2 started learning English when he was 11 years old, and did both his Secondary and Bachelor's studies in his native language. S3 could be considered bilingual since her mother (that she grew up with) is a native English speaker. However, S3 does not consider herself to have a native English proficiency, since she spoke Spanish most of her life, and also did all of her education in this language. The participants have earlier passed internationally recognized English proficiency level tests that can be interpreted using the Common European Framework of Reference for Languages (CEFR) scale[2]. According to the CEFR global scale, S1 and S3 have the C1 or "proficient user" English proficiency level (S1 – IELTS: 8.0/9.0; S3 – Cambridge Advanced level test: grade B). S2 has the B2 or "independent user" proficiency level (First Certificate in English: grade B). The Interlocutor has, according to the CEFR scale, the C1 proficiency level (TOEFL: 113/120). Thus, the sample formed a group representing young (age: S1 – 31, S2 – 26, S3 – 27) educated people with a good proficiency level of English as a foreign language.

Apart from S2 having a slightly lower proficiency level according to the English proficiency scale, there is one more factor that could influence the results. While S2 and S3 did all of their education in their native languages, S1 did her Secondary and Higher Education in English, and thus, had much more exposure to the L2 under analysis. This could suggest that S1 has had more exposure to both listening and production in English from an early age, and that her disfluency rates would be lower than that of S2 and S3.

2.2 Research Design

The analysis is based on the recordings of three dialogues with three non-native speakers of English. The participants knew beforehand that they would be recorded and they had given their consent. However, they did not know which part of the recorded dialogue would be analysed. At some point during the conversation, the subjects were asked the same set of testing questions about their Bachelor's study[3]. None of them was aware of what the questions are prior to

[1] henceforth S1, S2, and S3, respectively

[2] https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

[3] Sample questions are: "How was your major called?", "What was the structure of your curriculum?", "Was it hard to get the best mark?"

the experiment. The Interlocutor tried to keep a friendly atmosphere and did her best to make sure the participants feel comfortable during the dialogue.

All the three conversations were recorded in stereo using PRAAT. To make sure that the participants are assessed equally, on the similar tasks, and complexity, the speech samples chosen for the analysis only covered the utterance turns of the subjects answering to the testing questions and talking about their Bachelor's degree. All speech recordings were transcribed and annotated. Speech disfluencies (defined in Section 2) were transcribed exactly as they were pronounced. Silent pauses and background noises were indicated with square brackets. The final data corpus of all the three conversations with the Interlocutor's speech excluded consists of 1144 words (S1 – 485, S2 – 299, S3 – 360), and approximately 106 illocutionary acts (S1 – 45, S2 – 22, S3 – 39).

2.3 Main Data Points

Previous studies of fluency and disfluency in native and non-native speech have identified a number of measurable quantitative variables that appear to be related to perceived fluency and disfluency[4]. Three of the components that were included in the current analysis of non-native speech in a dialogue are based on the widely used classification of fluency adapted from Tavakoli and Skehan [16]. This classification allows for individual assessment of speaker's fluency/disfluency as distinct from the context:

- rate of speech (the number of words per minute of speech);
- breakdown fluency (the number and length of pauses);
- repair fluency of the speaker (the number of false starts, corrections, and repetitions).

As the fourth component, the correlation of speech disfluencies with the flow of the dialogue is investigated to see how disfluency relates to its position in a dialogue and a dialogue act. In order to examine the position of the disfluencies, the transcribed dialogues have been annotated according to the type of illocutionary acts. To analyze their correspondence with speech disfluency, I follow the approach of Kogure [12], that views dialogue as consisting of minimal units called illocutionary acts introduced by Searle [15]. Since an illocutionary act consists of an illocutionary force and a propositional content, it is independent of the formal structure of language and allows for better assessment when working with several languages or non-native speech [12]. The number of disfluencies was measured per one type of illocutionary acts.

Rate of speech The authors of previous studies suggest different methods of speech rate analysis. In this study, I adopt the most widely used technique (used

[4] Even though the fluency-disfluency dichotomy is controversial [13], the current paper does not account for disfluencies that facilitate communication or make speech more fluent, and thus, regards disfluency as the antipode of fluency.

by [10], [17]) and calculate the relation of the number of words in the utterance per minute of speech including and excluding utterance internal silences. The duration of speech is measured in seconds from the beginning of the first word to the end of the last word for every utterance. Silences present at the beginning and end of every utterance are not considered, but the utterance internal pauses are included in the total time. The total measurements are normalized by 60 seconds.

Breakdown Fluency Breakdown fluency is measured according to the methods introduced by Cucchiarini [6, 2869] using four variables: duration of silent pauses per minute of speech; number of silent pauses per minute of speech; duration of filled pauses per minute of speech; number of filled pauses per minute of speech.

Following previous research [18, 14], pauses shorter than 250 milliseconds are considered to be natural articulatory pauses (e.g. breathing). According to Gotz [10], such pauses are very frequent, and they do not make one's speech less fluent. Thus, they are not regarded as disfluency phenomena.

Repair Fluency Repair fluency is measured as the number of repetitions (exact repetitions of words), corrections (changing a partially/fully uttered incorrect unit by the correct one without repetitions), and restarts (repetitions of initial parts of words) per minute of speech. To gain more insight into the occurrence of these phenomena, not only quantitative but also qualitative properties are taken into account. The distribution of the considered repair phenomena is calculated individually for each speaker.

Position of Disfluency in Dialogue In a dialogue, a speaker expresses their communicative intentions (or performs communicative acts) and a hearer tries to understand them [11]. Dialogue utterances involve various kinds of communicative acts. To analyze their correspondence with speech disfluency, as mentioned above, the current paper views dialogue as consisting of minimal units of human speech called illocutionary acts [15]. The correlation of speech disfluencies with their position in dialogue was measured as the number of disfluencies per one illocutionary act type. The recorded conversations were manually transcribed using PRAAT TextGrid. As the author of the current paper, I also annotated the dialogues myself. To annotate the dialogue utterances under analysis as illocutionary acts I employed the following classification proposed by Searle [15]:

- **assertives** – illocutionary acts that commit a speaker to the truth of the statement;
- **directives** – illocutionary acts that are meant to cause the addresse to do something, e.g. requests, directions, questions and advice;
- **commissives** – illocutionary acts that are meant to reveal the intentions of the speaker, e.g. promises, threats, and oaths;

- **expressives** – illocutionary acts that express on the speaker's attitudes and emotions towards a particular proposition, e.g. congratulations, apologies and thanks;
- **declarations** – illocutionary acts that change the world or the reality, e.g. baptisms and verdicts.

For each speaker, the above-mentioned objective fluency measures were calculated manually over the chosen speech samples set. As a result, one value per objective measure for each speaker and one average value per objective measure for all the three speakers were obtained.

3 Findings

In this section, the results of the analysis would be described. In total, for the first three analysis components 12 variables were obtained. Moreover, 5 variables were considered for the fourth component. The measurements were taken for each speaker individually. In addition, the sum of the three measurements of each speaker was averaged to calculate the average value of the individual results. A summary of the results is given in the two tables below.

Analysis results

Measured variable	**Speaker 1**	**Speaker 2**	**Speaker 3**	**Average**
Speech Rate / min	**175.4**	111.8	147.2	144.8
Total number of disfluencies / min	**15.3**	14.5	13.6	14.5
Duration of all pauses / min	3.5 sec.	**13.2 sec.**	9.6 sec.	8.8 sec.
Number of all pauses / min	8.6	11	**11.2**	10.3
Number of all repair disfluencies	**6.7**	3.5	2.4	4.2
Number of false starts	0	**0.7 (20%)**	0	0.2 (4%)
Number of corrections	**2.9** (43.3%)	2 (**60%**)	2.4 (33.3%)	2.4 (59%)
Number of repetitions	**3.8 (56.7%)**	0.7 (20%)	0	1.5 (37%)

Disfluencies in the dialogue

Illocutionary act type	**Speaker 1**	**Speaker 2**	**Speaker 3**	**Average**
Total number	45 (16)	22 (21)	39 (17)	35.3 (18)
Directives	2 (2)	1 (1)	4 (0)	2.3 (1)
Expressives	0	3 (2)	2 (2)	1.6 (1.3)
Assertives	43 (14)	18 (18)	33 (15)	31.3 (15.6)
Commissives	0	0	0	0
Declarations	0	0	0	0

3.1 Rate of Speech

To compare the subjects' rates of speech to those of native speakers of English, we employed the findings of Tauroza and Allison [17], who investigated the range of speed of everyday conversational English. Their measurements also include articulation rate plus internal pause time.

Range of speech rates in native English speakers' conversation (adapted from Tauroza, S., and Allison, D.)

	Words per minute
Faster than normal	260
Moderately fast	230-260
Average	190-230
Moderately slow	160-190
Slower than normal	160

As can be seen from the table above and the table of Analysis Results, speech rates of all the subjects are lower than those of "Average" native speakers. Only the rate of speech of S1, the highest of the three samples (175.4 words per minute), can be classified as "Moderately slow". Such a result may be connected to the fact (already mentioned in the previous section), that S1 also had more exposure to English since her childhood than the other participants. The rates of both S2 and S3 fall into the category of "Slower than normal". In terms of larger goals of this paper, this confirms the correlation between fluency (on the CEFR scale) and rate of speech, as suggested by Tauroza and Allison [17].

3.2 Breakdown Fluency

The measurements of Breakdown Fluency are summarized in the table below.

Breakdown Fluency per minute of speech

Measured variable	**Speaker 1**	**Speaker 2**	**Speaker 3**	**Average**
Duration of silent pauses / min	2 sec.	7.6 sec.	6.5 sec.	5.4 sec.
Number of silent pauses / min	4.8	4.1	7.2	5.4
Duration of filled pauses / min	1.5 sec.	5.6 sec.	3.1 sec.	3.4 sec.
Number of filled pauses / min	3.8	6.9	4	4.9
Duration of all pauses / min	3.5 sec.	13.2 sec.	9.6 sec.	8.8 sec.
Number of all pauses / min	8.6	11	11.2	10.3

The Breakdown Fluency component was measured by analyzing the data with regard to the length and frequency of both silent and filled pauses. According to [6], that investigates the Pearson Correlation between fluency ratings and primary speech variables on the speech recordings of 60 non-native Dutch-speakers, less fluent speakers, in general, do not make longer pauses than more fluent speakers, but they do pause more often. The findings of the current analysis only partially correlate with the statement, if we consider, following [17], that the speed of speech is the key factor affecting one's speech fluency. Although S1, being the fastest speaker, makes the least number of pauses per minute (8.6 overall), S2 and S3 make almost the same number of pauses per minute (11 and 11.2 overall, respectively). Even though S2 has a lower rate of speech than S3, he does not make more pauses.

Despite this, S2, being the slowest speaker in terms of rate of speech, has the highest duration of an average pause (1.2 seconds). S1, as the fastest speaker, has the lowest duration an average pause (0.4 seconds). One can conclude, that in our analysis, duration of pauses (but not their number) has a strong correlation with rate of speech.

In the current case study, the subjects made both silent and filled pauses mostly when trying to retrieve a less frequently used lexical item from the memory, which can be concluded from the following examples:

(i) Interlocutor: "It was all set." S1: "All set. There was a concept of electives, but the... [**silent pause**] college was like: "We don't have a teacher for that, so you have to take what we give you", so..."

One can notice that S1 paused directly after using a function word (the definite article "the"), which suggests that she already decided to refer to some entity, but could not come up with the right word to describe it. A post-study discussion with S1 confirmed that the pause was not conscious. It is important to note that speaker pauses after a function word and before a content word. [10, 20] cites [2, 152], which calls this phenomenon common and claims that "hesitations in phonemic clauses are most likely to occur after at least a preliminary decision has been made concerning its structure and before the lexical choices have been finally made".

In the analysed dialogues, the speakers also used filled pauses during hesitation:

(ii) Interlocutor: "Yeah, it's the same for us, actually. Was it hard to get the best mark?"
S1: "**Uuhhh...**"

In this example, it is obvious that the speaker hesitates and makes a filled pause ("Uuhhh...") when she is not sure how to answer the question. A filled pause here occurs at the beginning of an utterance. [5, 590] states that hesitation phenomena are generally more likely to occur at the beginning of an utterance, relating that to the greater planning demand at this part of a phrase.

3.3 Repair Fluency

A summary of the repair fluency measurements is presented below. Overall, 4 repair fluency measurements were taken. The means of repair disfluencies are divided as follows: 4% of false starts (repetitions of initial parts of words), 59% of corrections, and 37% of repetitions.

Corrections (changing a partially/fully uttered incorrect unit by the correct one without repetitions) are the first phenomena under scrutiny. Besides being

Repair fluency per minute of speech

Repair disfluency variable	**Speaker 1**	**Speaker 2**	**Speaker 3**	**Average**
Number of all repair disfluencies	6.7	3.5	2.4	4.2
Number of false starts	0	0.7 (20%)	0	0.2 (4%)
Number of corrections	2.9 (43.3%)	2 (60%)	2.4 (33.3%)	2.4 (59%)
Number of repetitions	3.8 (56.7%)	0.7 (20%)	0	1.5 (37%)

the most frequent repair disfluency, they are also the only type observed in the dialogue with S3. Mostly the subjects correct themselves at the beginning of an utterance (on average 70% of all corrections), such as in the following example:

(iii) Interlocutor: "Uhg-um... But what about, like, small classes, where you have to do, like, lab work? "
S1: Uhm... **when– uh... in the lab classes**, we would only, like, given a laboratory."

In this case, the subject changed the entire structure of the utterance by correcting the conjunction "when" with a prepositional phrase.

Repetitions (exact repetitions of words), the next type of Repair Disfluencies, are used by the speakers to win some extra time for formulating a sentence [6]. In our data repetitions seem to have the same function:

(iv) Interlocutor: "So, like, mine was Linguistics. Even though it's also called differently in my language. In our system."
S1: "**Ok. Ok**. So that was the stream – Computer Science and Engineering."

In this case, S1 repeats the word "Ok" at the beginning of the phrase, which is likely to be related to the increase cognitive demand needed for formulating the next utterance.

False starts (repetitions of initial parts of words) were only observed once, in the dialogue with S2:

(v) Interlocutor: "Like, your degree–"
S2: "**Techn– Yeah, technical** physics."

Here, according to a post-discussion with S2 himself, the speaker was not sure about what he already started to say, and took some time to assure himself that what he was saying is right.

As mentioned in [4, 165], that investigates the contribution of different fluency parameters to perceived fluency, repair fluency variables have a direct relationship with speech fluency. The higher a value, the less fluent (and more disfluent) the fragment.

The results showed that S1 produced the highest number of repair disfluencies per minute of speech, despite her having the highest rate of speech. One of the reasons for that could be that the number of repair disfluencies could have increased in proportion to the total increase in the number of words per minute of speech. However, though repetitions comprised more than half of S1's total disfluencies, S2 and S3 used either very few or none of them.

3.4 Correlation of the disfluencies with the flow of the dialogue

The numbers of occurrences of each type of illocutionary acts present in the data set are given below. The number of total disfluencies corresponding to the illocutionary act type is given in brackets.

Disfluencies in the dialogue

Illocutionary act type	**Speaker 1**	**Speaker 2**	**Speaker 3**	**Average**
Total number	45 (16)	22 (21)	39 (17)	35.3 (18)
Directives	2 (2)	1 (1)	4 (0)	2.3 (1)
Expressives	0	3 (2)	2 (2)	1.6 (1.3)
Assertives	43 (14)	18 (18)	33 (15)	31.3 (15.6)
Commissives	0	0	0	0
Declarations	0	0	0	0

Since all the speakers talked and answered questions about their Bachelor's degree, the analyzed parts of dialogue consisted mostly of illocutionary acts of the type Assertives. They comprised on average 31.3 of 35.3 illocutionary acts (88.6%). Assertives in the dialogues under study refer to explanation, clarification and description of some entities or events. Two examples of assertives in the data are provided below:

(vi) Interlocutor: "And what was the structure of your curriculum? Like, how many subjects did you take per semester?"
S1: "Uuhhh... In total you had to take, like, 42 theoretical subjects."

The speaker makes a statement, to which truth she commits.

In the next example, S2 provides an explanation as a response to the Interlocutor's questions:

(vii) Interlocutor: "Why?"
S2: "We had -eh... some opportunities. Like, we could choose between two subjects for, like, next semester.

Most of the disfluency phenomena that occurred in the dialogues (on average 15.6 of 18, or 86.6%) are also observed in Assertives. One reason for assertives to be more susceptible to disfluencies could be that assertives normally require more planning and have more content.

In the whole conversation, the speakers on average used only 2.3 directives (in on average 35.3 speech acts), mostly to clarify if they understood the question correctly and to make sure that the Interlocutor understands what they are saying, such as in these two examples:

(viii) Interlocutor: "I see. Did you have that thing, like, Bachelor's with honors?
S3: "What's Bachelor's with honors?"

We can see that here S3 uses the question ”What’s Bachelor’s with honors?” to get information from the Interlocutor.

Overall, 5.5% of all disfluencies occured in Directives. Directives in the analyzed dialogues might have caused less disfluencies due to the fact they usually referred to the previously set context (which is, required less processing).

The expressive illocutionary acts were also used quite rarely (on average only 1.6 of overall 35.3 illocutionary acts), and only by S2 and S3, mostly to express their attitude or assessment:

(ix) S2: ”Uhh... It’s uh... [silent pause] I think it would be betterrr to have, like, eh, core subjects.”
Interlocutor: ”Uh hum...”

Here S2 expresses his positive assessment of an event with the use of Subjunctive Mood.

Only 7.2% of all disfluency phenomena occured in Expressives.

As can be seen from the table, Commissives and Declarations were not used in the analyzed parts of dialogues at all.

Talking about the position of a disfluency in the dialogue act itself, it can be concluded that the disfluencies mostly occur at the beginning of it. This is, again, most likely connected to the fact that at the beginning of an utterance speakers usually experience a higher cognitive load that is caused by the need to plan a phrase. An example is given below:

(x) S3: ”That scale... [laughing] Would you call it scale? Yeah, I have courses, that I have, like, best grades, or very good grades. And other courses that I barely passed.”
Interlocutor: ”Uh-hm.”
S3: ”So... [**silent pause**] but then, in average it was good.”

Here the speaker makes a silent pause before making an assessment.

4 Conclusion and Further Research

In this paper, the analysis on disfluencies in three non-native dialogues with three different subjects has been presented. The results of the case study correspond to its major goal, which is to examine and describe the distribution of speech disfluencies in the analyzed dialogues, and compare the analysis to the previous research on speech disfluencies.

In connection to future investigation, a more qualitative and possibly more extensive analysis of disfluency phenomena is required to understand the characteristics of speech disfluency in non-native dialogue. It would be interesting to explore how disfluencies of non-native speakers differ at various stages of language learning. Apart from that, it is important to establish what specific measures should be employed in the assessment of one’s speech disfluency to develop more objects instruments of fluency and proficiency testing.

References

1. Arnold, J. E., and Tanenhaus, M. E., and Hudson K. , Carla L.: If You Say Thee uh You Are Describing Something Hard: The On-Line Attribution of Disfluency During Reference Comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition **33**(5), 914–930 (2007)
2. Boomer, D. S.: Hesitation and Grammatical Encoding. National Institute of Mental Health, Bethesda **8**(3), 148–158 (1965)
3. Bortfeld, H., and Leon, S. D., and Bloom, J. E., and Schober, M. F., and Brennan, S. E.: Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. Language and Speech **44**(2), 123–147 (2001)
4. Bosker, H. R., and Pinget, A., and Quené, H., and de Jong, N. H.: What makes speech sound fluent? The contributions of pauses, speed and repairs. Language Testing **30**(2), 159–175 (2012)
5. Corley, M., and Stewart, O. W.: Hesitation Disfluencies in Spontaneous Speech: The Meaning of um. Language and Linguistics **2/4**, 589–602 (2008)
6. Cucchiarini, C., and Boves, L., and Strik, H.: Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. The Journal of the Acoustical Society of America **44**(6), on Proceedings 2862–2873 (2002)
7. Gürbüz, N.: Disfluency in dialogue: an intentional signal from the speaker? Psychon Bull Rev **19**, 921–928 (2012)
8. Fox Tree, J. E.: The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. Journal of Memory and Language **34**, 709–738 (1995)
9. Grosjean, F.: Spoken word recognition processes and the gating paradigm. Perception Psychophysics volume **28**, 267–283 (1980)
10. Götz, S.: Fluency in Native and Nonnative English Speech. Planning and task performance in a second language. Amsterdam: John Benjamins 238 (2013)
11. Iida, H., and Kogure, K., and Yoshimoto, K., and Aizawa, T.: An experimental spoken natural dialogue translation system using a lexicon-driven grammar. In The Proceedings of the European Conference on Speech Communication and Technology (1989)
12. Kogure, K., and Kume, M., and Iida, H.: Illocutionary Act Based Translation of Dialogues. COLING (1990)
13. Lintunen, P., Mutta, M., Peltonen, P.: Fluency in L2 Learning and Use. Multilingual Matters. 216 (2019)
14. Towell, R., and Hawkins, R., and Bazergui, N.: Native speakers' perceptions of fluency and accent in L2 speech. Language Testing **31**(3), 349–365 (2014)
15. Searle, J. R.: A Classification of Illocutionary Acts. Language in Society. Cambridge University Press **5**(1), 1–23 (1976)
16. Tavakoli, P., and Skehan, P.: Strategic planning, task structure, and performance testing. Amsterdam: John Benjamins 239–273 (2001)
17. Tauroza, S., and Allison, D.: Speech Rates in British English. Applied Linguistics **11**(1), 90–105 (1990)
18. Towell, R., and Hawkins, R., and Bazergui, N.: The Development of Fluency in Advanced Learners of French. Applied Linguistics **17**(1), 84–119 (1996)

A Cross-Linguistic Examination of Geminate Consonant Attrition

Michelle Ciccone[1], Rawan Hanini[1], and Mariagrazia Sciannantena[2]

[1]City University of New York - Brooklyn College, Brooklyn, NY
rawan.hanini26@bcmail.cuny.edu
michelle.ciccone86@bcmail.cuny.edu
[2]City University of New York - Kingsborough Community
College, Brooklyn, NY
Mariagrazia.Sciannantena51@students.kbcc.cuny.edu

Abstract: This study explores the phenomenon of language attrition (Schmid & Kooke, 2007, Celata & Cancilla, 2010; Chang, 2012; De Leeuw, Tusha & Schmid, 2017). Specifically, we investigate the acoustic properties of consonant gemination across three groups of Italian and Palestinian Arabic speakers: (1) monolinguals i.e. native speakers born and raised in either Italy or Palestine and who have lived abroad their entire life, (2) late bilinguals or first-generation immigrantsi.e. speakers who emigrated to the US during their teens, and (3) heritage speakers or second generationthe speakers was approximately 25 years old. The participants were tested using a delayed word repetition task, following Alkhudidi et al (2018). The stimuli comprised 60 bi-syllabic minimal and near-minimal pairs in either Arabic or Italian including long and short stops (e.g. for Italian, /fato/ 'fate' vs. /fatto/ 'done', for Arabic, /sadaq/ 'he said the truth' vs. /sad:aq/ 'he approved'). We controlled for stress and syllabic position. Distractors were also included. The analysis consisted of manually aligning the target consonants using the Praat software (Boersma & Weenink, 2012). We extracted the mean consonant duration, and compared it statistically across the different groups using univariate ANOVAs. Our findings show significant main effects of group (monolingual/late bilingual/heritage speaker), Voicing (voiced/voiceless), and Consonant Type (singleton/geminate) on duration in both languages. We also note the existence of universal tendencies in language attrition regardless of language or cultural background, i.e. voiced and velar consonants appear more prone to attrition. Overall, our study adds to the body of work on phonological attrition by examining ongoing change in two bilingual communities living in the United States. Our findings are similar to those of similar studies conducted in Canada (Alkhudidi et al. 2018, Rafat et al 2017).

Keywords: Bilingualism, language attrition, language contact; Arabic-English; Italian-English; phonetic; phonology; geminates

1 Introduction

First Language Attrition and Change in Bilinguals

First language attrition and change in bilinguals can be defined as a loss or gradual decline in the proficiency of a first language (L1), mainly caused by interference from a second language (L2; Schmid & Köpke, 2007). Language attrition can be observed when speakers of L1 immigrate to a different country or region where another language is predominantly spoken (L2) in which these individuals have to function. The process of those speaking L2 more than L1 in their lives will cause L1 attrition (Celata & Cancilla, 2010; Chang, 2012; De Leeuw, Tusha & Schmid, 2017). There are different aspects of the linguistic system which can undergo attrition. One of these is the lexicon, also known as vocabulary. Within this particular component, individuals may forget or have slower access to their lexicon items (Kopke & Schmid, 2004). Another component is represented by the grammatical aspects of a language. In grammar, particular structures may be used less and less, and sometimes not used at all anymore in one's production (Kopke & Schmid, 2004). The last component is the phonology, the attrition of which is the focal point of the current paper. While most research in the past has focused on the lexicon and the grammatical aspects of the L1 (Kopke & Schmid, 2004), fewer studies have addressed the phonology of a language in the process of attrition (Celata & Cancila, 2010; Rafat, Mohaghegh & Stevenson, 2017; de Leeuw; Tusha &; Schmid, 2018; Alkhudidi et.al, 2018; & Hanini et.al, 2019) or the topic of phonetic attrition (Flege, 1987; Major,1992; Guion, 2003; Mayr, Price & Mennen, 2012).

This study has two aims. First, we compare the acoustic properties of the length contrast between singleton and geminate consonants across speakers from two languages pertaining to different families, specifically Italian and Arabic. We examine three groups of speakers of each language to determine if phonological attrition is underway. These groups include monolinguals, and late and simultaneous bilinguals living in New York City, the United States of America. This is relevant to the critical period hypothesis (Lenneberg, 1969), stating that the grammar of native speakers becomes set after their childhood. To the best of our knowledge, research studies have only rarely addressed the phonological attrition of geminates. One such study focused on Italian-American bilingual immigrants (Celata and Cancilla, 2010). Another study examined Farsi-English bilingual immigrants (Rafat et al., 2017). Two additional studies investigated the attrition of Arabic in Arabic-English bilinguals, specifically Syrian Arabic-English bilingual immigrants (Alkhudidi et.al., 2018), and Palestinan Arabic-English bilingual immigrants (Hanini et.al, 2019). The second aim of this study is to uncover universal phonetic tendencies in the process of attrition. We seek to determine whether place of articulation and voicing have an effect on language/geminate attrition. Specifically, we want to know whether voiced or voiceless phonemes that have different places of articulation will behave differently in terms of their attrition properties in Italian and Arabic.

2 Background

2.1 Phonological attrition and Phonetic Shift

There are numerous attested cases of phonological attrition and phonetic shift within the speech of first and second generation immigrants. The literature on phonological attrition examined different elements of the L1. For example, De Leeuw, Mennen & Scobbie (2012) focused on the production of the German lateral phoneme /l/ of late German-English bilingual speakers. In this study, differences among the first and second formants (F1 and F2) of the laterals were found. The German lateral /l/ of the bilinguals diverged from the production of native German speakers, showing a gradual drift towards the L2 English values.

Regarding the specific topic of geminate change in bilingual speakers, only a handful of studies have explored this concept. Celata and Cancila (2010) examined the singleton-geminate contrast in bilingual English-Lucchese immigrants in the United States and monolingual Lucchese speakers. The authors concluded that reduced perceptual discrimination of the second-generation group was caused by the influence of the American English phonological system, in which consonant length is not phonologically contrastive. The authors thus demonstrated the effect of generation as a predictor of L1 attrition.

Third, Flege (1987) conducted a bilingual study which found proof of a shift in the voice onset time (VOT) values in the L1 for English-French and French-English adult bilinguals resulting from exposure to the L2.

Fourth, Rafat et al. (2017) explored whether the singleton-geminate consonant length contrast declines across three generations of Farsi-English bilinguals. It was found that across successive generations, gemination slowly undergoes attrition. With English contact, there was singleton-geminate loss, with the durations of both members of the pair becoming more similar to each other. The authors also addressed the question whether universal phonetic factors, such as manner of articulation and voicing, impose constraints on geminate production. The findings did not suggest that these elements constrained to the degree of geminate attrition across generations.

Fifth, Alkhudidi et al. (2018) and Hanini et al. (2019) found that the transition in geminate attrition is gradual, with the first generation of immigrants not differing significantly from either the monolinguals or the heritage speakers, while significant differences were found between monolinguals and heritage speakers.

Lastly, Hrycyna, Lapinskaya, Kochetov, and Nagy, (2011), examined the VOT values in voiceless stops /p, t, k/ in the L1 of successive generations (first, second, and third) of Italian-, Russian-, and Ukrainian-English bilingual communities in a sociolinguistic approach. It was found that for all language groups there was a shift in the VOT values towards English. The differences between language groups were attributed to social factors such as the cohesiveness and size of a community as well as participants' attitudes towards their L1.

The present study follows Hanini et al. (2019), Alkhudidi et al. (2018), and Rafat et al. (2017) by examining geminate attrition in two different groups of participants, namely Italian-English bilinguals and Arabic-English bilinguals living in the United States, where a different pattern of geminate attrition may result because gemination is lexically more frequent in Arabic than Italian.

2.2 What are Geminates?

Geminates are phonetically long sounds (Homma, 1980: Lahiri & Hankamer, 1998; Esposito & Bendetto, 1999; Ohala, 2007; Kraechenmann & Lahiri, 2008). As discussed by Alkhudidi et al. (2018), among others, consonant length can be contrastive in a language. A length contrast between singleton (short) and geminate (long) consonant can be observed in 3.3% of the world's languages (Maddieson, 1984). This suggests that gemination is a marked phonological phenomenon (Payne, 2005; Celata and Canila, 2010). Languages such as Italian (Payne, 2005; Celata and Canila, 2010), Arabic (Hassan & Payne, 2008; Khattab & Altamimi, 2014), and Farsi (e.g. Hansen, 2004; Rafat, 2008, 2010) display these binary length contrasts.

In English, we encounter gemination as well but it is phonetic rather than phonemic. This occurs when two consonants of the same kind are placed adjacent to each other across word boundaries. For example, when saying the phrases /night time/ or /makes sense/, the two adjacent identical consonants result in a longer /t/ or a longer /s/, but even if they were shorter, they would not change the meaning of these phrases and hence are not considered phonemic.

2.3 Factors Affecting L2 Geminate Production

In L2 production, geminate consonant realization can be affected by phonetic universals, and implicational principles. The production of Italian geminates by German, Spanish-and Chinese-speaking learners of Italian was investigated by Sorianello (2014). Place of articulation, voicing, and stress all affected L2 geminate consonant production, albeit place of articulation and voicing were better predictors of gemination than stress. It was concluded that gemination was more likely to happen with voiceless stops, and degemination with sonorant consonants.

3 Hypotheses

The hypotheses for the present study are as follows:

H1. Geminate durational properties will change in Italian-English and Arabic-English bilinguals living in the United States. Specifically, these sounds will become shorter across generations.
H2. Universal phonetic factors will determine the degree of geminate change across generations, (but the extent to which this happens may depend on the specific language).

4 Methodology

4.1 Participants

There were 13 Italian participants in the study: 5 first generation late bilingual Italian-English speakers, 4 second generation heritage Italian-English speakers and 4 monolinguals that were native Italian speakers, who do not speak any English. The late bilinguals are speakers born in Italy who emigrated to the United States during their teens. The heritage speakers were born in the United States and speak both English and Italian in their daily lives. The native speakers were born in Italy and have lived there their entire life. The age range was from 19 to 42 and the mean age was 24 years old. In total there were 8 females and 5 males.

There were 16 participants in the Arabic-English study (Hanini et al, 2019): 5 monolinguals native speakers of Arabic, who did not speak any English. 4 late adult immigrant Arabic-English bilinguals, and 7 early adult Arabic-English bilinguals who are heritage speakers of Arabic. The late bilinguals are those who came to the United States after puberty and the adult heritage speakers are the children of Arabic immigrants who were born in the United States. All together there were 7 males and 9 females with various ages of arrival. The participants' ages ranged from 20 to 42 years old with a mean age of 25.

4.2 Stimuli

For the Italian-English study, the stimuli consisted of 60 bi-syllabic Italian minimal and near minimal pairs. There were 34 minimal pairs (e.g., /ziti/ 'pasta' vs. /zitti/ 'quiet'), and 26 near minimal pairs (e.g., /baba/ 'rum (used in sweets)' vs. /babbo/ 'father'). Both geminates and singletons included voiced and voiceless stop segments from three places of articulation: labial (b, p), dental (d,t), and velar (k,g). We controlled the stress and syllabic position for the stimuli. A total of 10 distractors were included (e.g. /congratulazioni/ 'congratulations' and /marrone/ 'brown') The stimuli were recorded by a native speaker of Italian who read them off of a PowerPoint presentation.

For the Arabic-English portion of this study, the stimuli consisted of 60 bi- syllabic and tri- syllabic minimal and near minimal pairs (e.g. /sadaq/ "he said the truth" vs. /sad:aq/ "he approved"). Both geminate and singleton words included one class of sounds: stops (/b, d, t, k, t, q/). The stimuli were controlled for position (intervocalic position) and voicing. A total of 15 distractors were used (e.g., /shujerat/ 'small trees' and /shihada/ 'diploma'). The stimuli were recorded by a native speaker of Arabic who read them from a PowerPoint presentation.

4.3 Procedure

For both the Italian and Arabic portions of the study, the participants first answered a detailed questionnaire that included general questions about language background (e.g. the participants first language, parents' language background) and specific questions about their fluency and their experience with Italian/Arabic and English (e.g. when they started acquiring each language, when they became fluent in reading, writing and speaking the language). The info provided in this questionnaire ensured that the group assignments were accurate.

A delayed word repetition task was administered to each participant individually. The participants were seated in a quiet room and were presented with the auditory and written form of the target words in Italian on a laptop screen via PowerPoint. Each target word was embedded into a carrier phrase. "Quando sono andata ad Est, __ é quello che ho detto."[1] [When I go East, __ is what I say]. The target words were immediately preceded and followed by stop consonants to facilitate splicing for acoustic analysis. The participants listened to and saw each phrase on the screen. The phrase then disappeared from the screen and a reverse countdown of 7 seconds showed up, after which the participants were asked to repeat the phrase. The backward 7-second countdown was used to minimize traces of phonological input from memory (Bassetti, 2017). Each session took about 20-25 minutes to complete. Due to social distancing requirements imposed by the COVID-19 pandemic, the last 6 participants were recorded using Zoom Video Communications software (Yuan, 2011). Care was taken to ensure these participants were located in a quiet room away from distractions and they were asked to speak in an adequately loud voice after they saw and heard the carrier phrases spontaneously on a screen shared from one of the

[1] To increase the naturalness of the speakers' production, they were provided with a context for the carrier phase, specifically that the speaker is a wizard whose mission is to utter a specific 'magic' word whenever traveling in the direction of certain cardinal points.

authors' laptops. The sessions were recorded via Zoom software (Yuan, 2011) and Praat software (Boersma & Weenink, 2012).

For the Arabic-English study, a delayed word repetition task was given to each participant. The participants were seated in a quiet room and presented with the auditory and written form of the target words in Arabic on a laptop screen via PowerPoint. Each target word was placed in a carrier phrase /ana aqul … alan/ 'I say ... now'. Just like for the Italian portion of the study, the participants listened to and saw each phrase on the screen in Arabic script. The phrase then disappeared from the screen and a countdown of seven seconds was initiated. The participants were asked to repeat the phrase after the countdown disappeared. Participants' production was recorded using an iPhone 7 plus microphone. All participants were recorded individually in a face-to-face session which lasted approximately 20-25 minutes.

5 Data analysis and results

A total of 780 Italian tokens were aligned using Praat (Boersma & Weenink, 2012) by two of the authors that have native/near-native fluency in both Italian and English and an author with native/near-fluency in Arabic and English. Also, a total of 960 Arabic tokens were aligned using Praat (Boersma & Weenink, 2012) by one of the authors who is a native speaker of Arabic and has near-native fluency in English. The closure duration was measured for each geminate and singleton stop. The authors manually aligned the sounds for both languages, following which the segment durations were extracted with the help of a Praat script. For each consonant, mean duration was obtained and for each pair the ratio of geminate: singleton duration was calculated.

Figures 1, 2 and 3 illustrate the decrease in duration observed in the spectrograms for the word /cadde/ 's/he fell' across each of the three Italian-English groups:

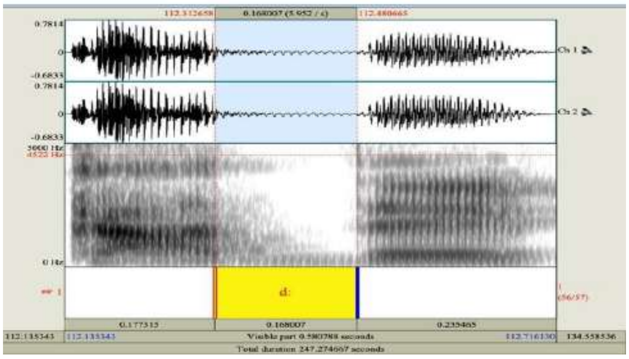


Figure 1. *Sample waveforms and spectrograms of the geminate consonant /d:/ by a monolingual native speaker of Italian (168 ms)*

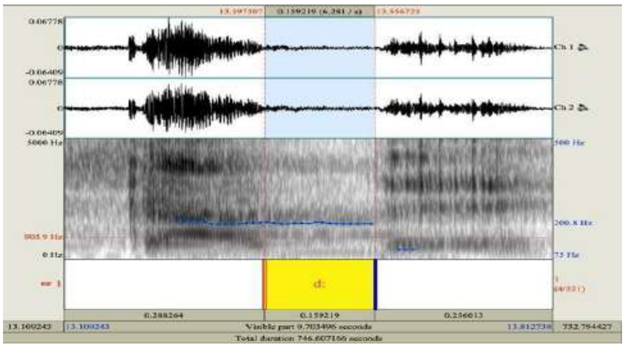


Figure 2. *Sample waveforms and spectrograms of the geminate consonant /d:/ by a late*

bilingual speaker of Italian (159 ms)

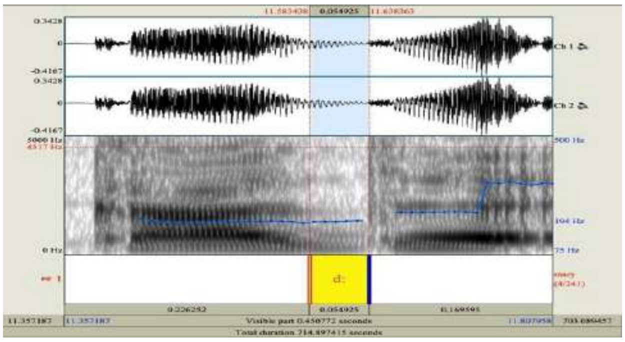


Figure 3. *Sample waveforms and spectrograms of the geminate consonant /d:/ by a heritage speaker of Italian (55ms)*

Figure 4, 5 and 6 illustrate the decrease in duration observed in the spectrograms for the word /sat:ar/ 'he covered' across each of the three Palestinian Arabic-English groups:

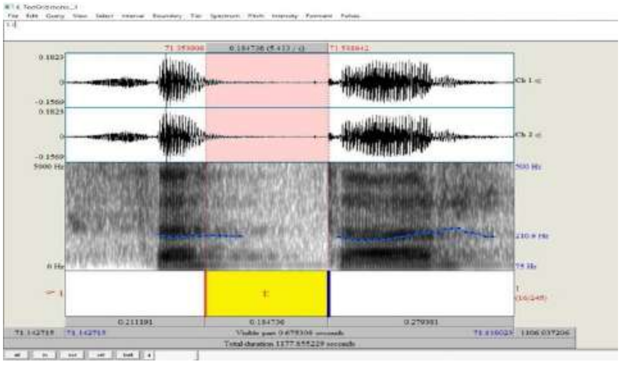


Figure 4. *Sample waveforms and spectrograms of the geminate consonant /t:/ by a monolingual native speaker of Arabic (184ms)*

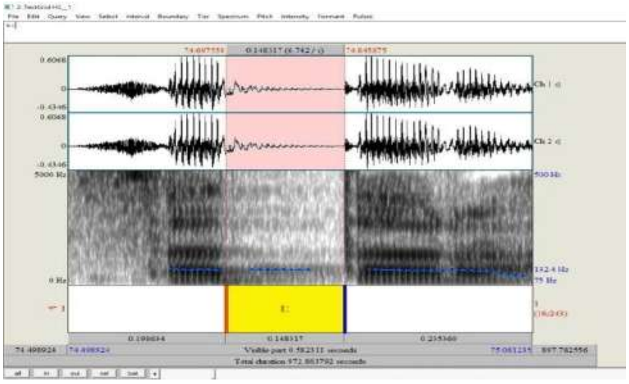


Figure 5. *Sample waveforms and spectrograms of the geminate consonant /t:/ by a late bilingual speaker of Arabic(148ms)*

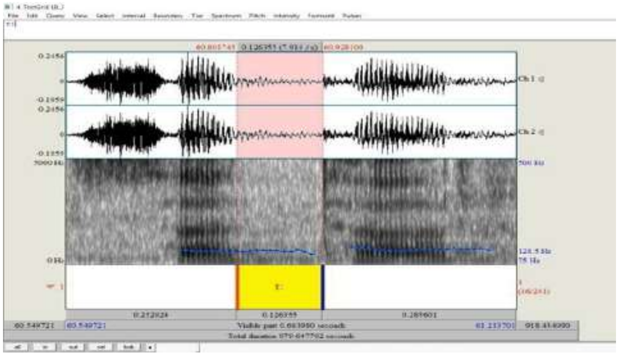

File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses
0.2456
0
Ch 1
Ch 2
5000 Hz
0 Hz
t:
0.212024
0.126355
0.289601
60.549721
61.213701

Figure 6. *Sample waveforms and spectrograms of the geminate consonant /t:/ by a heritage speaker of Arabic (126ms)*

5.1 Overall findings

5.1.1 Italian

A univariate ANOVA with duration as the dependent variable and speaker type (Monolingual, Late Bilingual, Heritage), consonant type (singleton/geminate), voicing (voiced/voiceless), and place of articulation (labial, coronal, dorsal) as independent variables revealed significant main effects of speaker type, consonant type, and voicing (but not place of articulation, F(2, 1194) = .533, p = .587) on Duration. For Speaker Type, F(2, 1194) = 49.36, p < .001, for Consonant Type, F(1, 1194) = 123.78, p < .001, and for Voicing, F(1, 1194) = 16.43, p < .001. The interaction between Speaker Type x Consonant Type was also significant, F(2, 1194) = 4.38, p = .013. No other two-way significant interactions were observed.

5.1.2 Arabic

Just like for the Italian data, a univariate ANOVA with duration as the dependent variable and speaker type (Monolingual, Late Bilingual, Heritage), consonant type (singleton/geminate), voicing (voiced/voiceless), and place of articulation (labial, coronal, dorsal) as independent variables revealed significant main effects of all of the independent categories on Duration. For speaker type, F(2, 1760) = 23.34, p < .001, for consonant type, F(1, 1760) = 3398.33, p < .001, for voicing, F(1, 1760) = 61.52, p < .001, and for place of articulation, F(2, 1760) = 10.34, p < .001). The interaction between speaker type x consonant type was also significant, F(2, 1760) = 10.92, p < .001, as was the interaction between voicing x consonant type, F(1, 1760) = 4.85, p = .02. We also investigated the interaction between speaker type x voicing but it was not significant.

5.2 The impact of early vs. late bilingualism

Post-hoc analyses with the Bonferroni correction revealed that for Arabic speakers, geminates were significantly shorter for heritage and late bilingual speakers compared to monolinguals. The same pattern was observed with Italian speakers, but the difference from monolinguals was more pronounced.

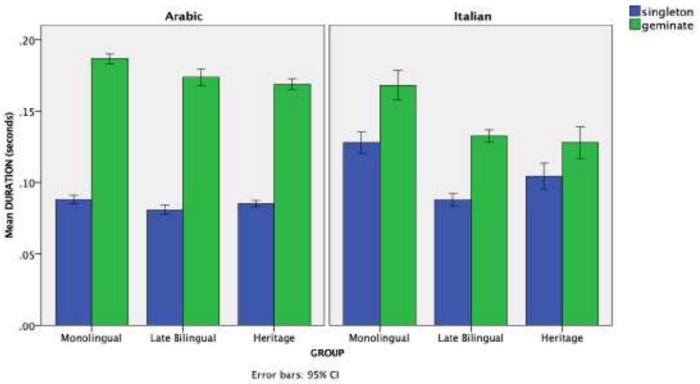


Figure 7. Mean consonant duration for Arabic and Italian singletons and geminates across each group of speakers.

5.3 The impact of place of articulation

Figure 8 shows the duration properties of singleton and geminate consonants at different places of articulation across the three different groups of speakers for the two languages. For Arabic, labial and coronal singleton consonants did not vary in duration across the three groups, but velar singletons were longer in monolinguals compared with the other two groups. For geminates, we note a gradual decrease in duration for all places of articulation. This decrease is most noticeable with dorsals and labials.

For Italian speakers, singletons at all places of articulation are generally comparable in duration between monolinguals and heritage speakers, and shorter for the late bilinguals. Geminates are longer for monolinguals compared to the other two groups, except for coronals where heritage speakers do not differ statistically from monolinguals.

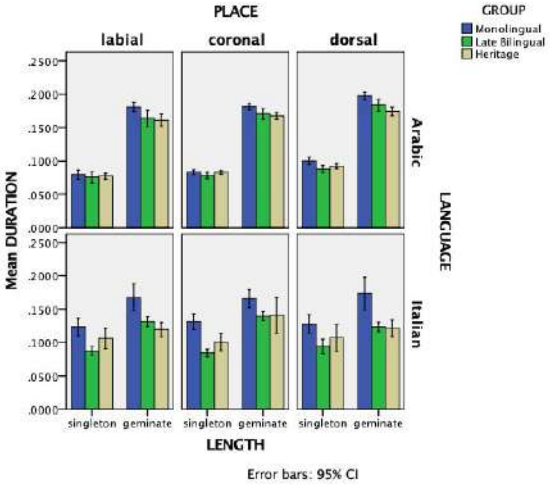


Figure 8. The impact of class of sounds on geminate attrition in late Arabic-English and late Italian-English bilinguals and heritage speakers.

5.4 The impact of voicing

As far as voicing is concerned (Figure 9), we note that voiced consonants tend to be shorter across the board. Arabic speakers produced singletons of similar durations across groups, regardless of whether these were voiced or voiceless. Geminates, however, were significantly longer for monolinguals compared to the other two groups when voiced. When voiceless, the heritage speakers produced shorter geminates compared to the monolinguals, but not the late bilinguals. Italian late bilinguals and heritage speakers produced shorter consonants compared to monolinguals almost across the board. The reduction in duration across groups is more noticeable compared to Arabic speakers.

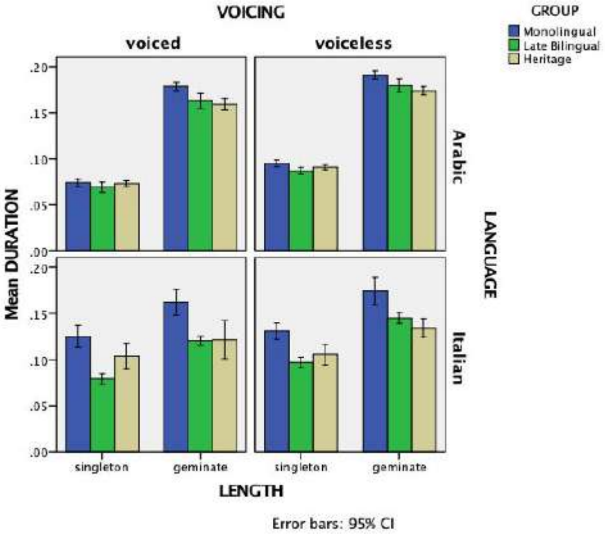


Figure 9. *The impact of voicing on geminate change in late Arabic-English and late Italian-English bilinguals and heritage speakers.*

6 Discussion

The first aim of this study was to investigate whether the geminate consonant production in Arabic-English bilinguals and Italian-English bilinguals living in the United States would undergo attrition. We hypothesized that this would be the case, following existing work on the topic (Rafat et al., 2017, Alkhudidi et al. 2018). Our study focused on geminate attrition in two groups of speakers that speak two different languages: late bilinguals and heritage speakers of Arabic-English and Italian English. We compared their production to that of monolingual native speakers of both languages, who served as a control group. While singleton and geminate consonants differed significantly in duration for all three groups of speakers within each language, the decrease in the duration of geminate consonants in both generations of immigrants is consistent with linguistic attrition, and may signal a process of incipient neutralization. This tendency was more prominent in Italian speakers compared to Arabic speakers.

The second aim of this study was to investigate whether universal phonetic factors such as place of articulation and voicing have an effect on geminate attrition. It was hypothesized that this would be the case, but no specific predictions were formulated. However, given general markedness tendencies, voiced segments (compared to voiceless), and labial/velar segments (compared to coronal) could be expected to display a higher degree of attrition than voiceless consonants across groups in both Arabic and Italian. This prediction for voicing was borne out in the speech of both late bilingual and heritage speakers of both languages, but it was more pronounced in the late bilinguals. Our findings support Sorianello (2014). As far as place of articulation is concerned, the decrease in duration across the different groups of speakers was more pronounced in labials and compared to coronal geminates, particularly in the case of Italian speakers, similarly to the findings of Rafat et al. (2017) and Alkhudidi (2018). Our second hypothesis thus received partial support.

Results from an earlier study on Arabic-English consonant attrition confirmed this first hypothesis as well (Hanini, et.al., 2019), with voiced segments exhibiting higher attrition rates. A possible explanation for this phenomenon is that sustaining voicing through a stop closure is generally difficult and more marked in the world's languages. It is also possible that the first generation is more affected by the other language (English) as they were exposed to it later and may have found it more difficult to reconcile the phonetic differences (as compared to heritage speakers who are exposed to both languages from birth). However, the present study found more similarities between heritage (second generation) and monolingual speakers of Italian, whereas Hanini et al. found that Arabic monolingual speakers were more similar to late bilinguals. There was a more noticeable gradual decline across different generations of Arabic-English speakers. This finding is consistent with Rafat et al. (2017) who reported evidence of geminate attrition in the speech of three successive generations of Farsi-English bilinguals living in Canada, and Alkhudidi (2018) who reported evidence of geminate attrition in the speech of three successive generations of Arabic-English bilinguals living in Canada. The present study is also consistent with previous literature findings (Celata & Cancila, 2010; Rafat et al., 2017)

In sum, our findings reveal the existence of universal tendencies in language attrition regardless of language or cultural background. At the same time, they raise questions regarding the role played by sociolinguistic factors in the maintenance of a linguistic contrast. The fact that there were more similarities between heritage (second generation) and monolingual speakers in Italian, whereas in Hanini et al. (2019) Arabic monolingual speakers were more similar to late bilinguals (that is, first generation immigrants) appears to suggest that the two cultures may be characterized by different patterns of intergenerational interaction. This question is going to be addressed in future studies by administering a more comprehensive linguistic background questionnaire asking the participants about the amount and type of exposure to each of their languages in more detail. We will also determine whether participants' feelings of group identification (i.e. how much they identify with the culture represented by each of their languages) play a part in the degree of attrition present.

7 Conclusions

Our study provided a cross linguistic perspective on Arabic and Italian geminate consonants in three groups of speakers: monolinguals, late bilinguals (first generation speakers), and early bilinguals (second generation or heritage speakers). Just like previous studies with different languages, we have found evidence of geminate attrition across both groups of bilinguals in both languages, Arabic and Italian. In the Arabic-English case, heritage speakers showed shorter geminates than late bilinguals and so the attrition was gradual across the three groups. However, in the Italian-English case, both heritage speakers and late bilinguals showed similar results in shorter geminates, with late bilinguals being more similar to monolinguals in some cases. Overall, the decrease in mean durations between monolinguals and the other two groups was more pronounced in Italian speakers compared to Arabic speakers. We have also found effects of universal phonetic factors on attrition, specifically voicing (with voiced stops displaying higher attrition) and place of articulation (with velars displaying higher attrition compared to coronals) for both languages.

In conclusion, our study adds to the body of work on phonological attrition by examining ongoing change in bilingual communities living in the United States. Our findings are similar to those of similar studies conducted in Canada (Alkhudidi, et al. 2018, Rafat, et al 2017).

References

Alkhudidi, A., Rafat, Y., & Stevenson, R., (2018). Geminate Attrition in the Speech of Arabic-English Bilinguals Living in Canada. Heritage Languages, 17(1), 1-37.

Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(11), 1835-1842. doi:10.1037/xlm0000417.

Boersma, P. & Weenink, D., (2012). Praat: doing phonetics by computer [Computer program]. Version 6.1.15, retrieved 20 May 2020 from http://www.praat.org/

Celata, C., & Cancila, J. (2010). Phonological attrition and the perception of geminate consonants in the Lucchese community of San Francisco (CA). International Journal of Bilingualism, 14, 1–25.

De Leeuw (2017) Language and cognition: Individual phonological attrition in Albanian-English late bilinguals

De Leeuw E., Mennen I., Scobbie J. (2012). Singing a different tune in your native language: L1 attrition of prosody. International Journal of Bilingualism, 16, 101116

De Leeuw, E., Mennen, I., & Scobbie, J. M. (2013). Dynamic systems, maturational Constraints and L1 phonetic attrition. International Journal of Bilingualism, 17,683–700.

Esposito Anna, & Maria G. Di Benedetto. 1999. Acoustical and perceptual study of gemination in Italian stops. Journal of the Acoustical Society of America 106. 20512062.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. Journal of Phonetics, 15,4765.

Flege, J.E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (ed.), Speech perception and linguistic experience: Issues in cross linguistic research, pp. 233–277. Timonium, MD: York Press.

Guion, S. G. (2003). he vowel systems of Quichua–Spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second languages. Phonetics, 60, 98128.

Hanini, R., Alkhudidi, A., Rafat, Y., & Spinu, L. (2019, 12). Geminate Attrition in the Speech of Arabic-English Bilinguals Living in the United States. Paper presented at the 178th meeting of Acoustical Society of American, San Diego, CA

Hansen, B. (2004). Persian geminate stops: Effects of varying speaking rate. In Agwuele, Augustine,Warren, Willis, & Park, Sang-Hoon (eds.), Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception (pp. 86–95). Somerville, MA: Cascadilla Proceedings Project.

Heliger, Herbert L.; Vago, Robert M. (1991). First language attrition. Cambridge: Cambridge University Press. p. 4.

Homma, Y. 1981. Durational relations between Japanese stops and vowels. Journal of Phonetics 9. 273–281.

Khattab, G., & Al-Tamimi, J. (2014). Geminate timing in Lebanese Arabic: the relationship between phonetic timing and phonological structure. Laboratory Phonology, 5(2), 231-269.

Köpke, Barbara., Schmid, Monika (2007) Rijksuniversiteit Groningen, The Netherlands/Laboratoire de Neuro Psycholinguistique, Université de Toulouse, Le Mirail, France &quot;Bilingualism and Attrition "

Kraehenmann, Astrid; and Lahiri Aditi. 2008. Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. Journal of the Acoustical Society of America 123. 4446–4455.

Kopke, B. & Schmid M. S. (2004). Language attrition: the next phase. First Language Attrition: Interdisciplinary perspectives on methodological issues.

Lahiri, Adithi, & Jorge Hankamer. 1988. The timing of geminate consonants. Journal of Phonetics 16. 327–338.

Lenneberg, E. (1969). On Explaining Language. Science,164(3880), 635-643. Retrieved from http://www.jstor.org/stable/1725957

Maddieson, I. (1984). Patterns of sounds. Cambridge: Cambridge University Press.

Major, R. C.(1992). Losing English as a first language. The Modern Language Journal, 76(2), 190-208.

Mayr, R., Price, S., & Mennen, I. (2012). First language attrition in the speech of Dutch–English bilinguals: the case of monozygotic twin sisters. Bilingualism: Language and Cognition, 15, 687–700.

Ohala, Manjari. 2007. Experimental method in the study of Hindi geminate consonants. Experimental Approaches to Phonology 351–368.

Payne, E. (2005). Phonetic variation in Italian consonant gemination. Journal of the International: Phonetic Association, 35(2), 153-181.

Rafat, Y., (2010). A socio-phonetic investigation of rhotics in Persian. Iranian Studies, 43, 667 682.

Rafat, Y. (2011). Orthography-induced transfer in the production of novice adult English-speaking learners of Spanish (Doctoral dissertation, University of Toronto).

Rafat, Y., Mohaghegh, M., & Stevenson, R.A. (2017). Geminate attrition across three generations of Farsi-English bilinguals living in Canada: An acoustic study.

Yuan, E., (2011). Zoom Video Communications. https://zoom.us/