Bringing Together Computational and Linguistic Models of Implicit Role Interpretation

by

Timothy J. O'Gorman

B.A., University of Illinois Champaign-Urbana, 2009

M.S., University of Colorado-Boulder, 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

2019

This thesis entitled:
Bringing Together Computational and Linguistic Models of Implicit Role Interpretation
written by Timothy J. O'Gorman
has been approved for the Department of Linguistics

Prof. Martha Palmer

Prof. Laura Michaelis

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

O'Gorman, Timothy J. (Ph.D., Linguistics)

Bringing Together Computational and Linguistic Models of Implicit Role Interpretation

Thesis directed by Prof. Martha Palmer and Prof. Laura Michaelis

This dissertation studies *implicit semantic roles* – instances where a participant is not explicitly stated in the text, such as the arguments in "$\emptyset_{you}$ eat $\emptyset_{food}$ yet?". It focuses upon the task of resolution of these implicit roles – determining what these unstated arguments refer to.

This thesis proposes a typology of different kinds of these implicit roles, distinguished not by their syntactic behavior (which is very language-specific), but by their referential behavior. Implicit roles in some contexts act like pronouns, looking to recently mentioned referents; in other contexts, implicit roles can refer generically to "people in general", or to a speaker or addressee. The first contribution of the thesis is to outline the range of these various interpretations seen for these implicit roles across different languages so that we might make apples-to-apples comparisons from language to language.

The second part of this thesis presents new corpora of English implicit semantic roles, and presents computational models trained upon those corpora to do implicit role resolution. This provides data to do the full task of resolving all unstated semantic roles in a document. On those new corpora, a set of implicit role resolution models are trained, showing that while this data is difficult, one can build wide-coverage systems which predict implicit semantic roles using the PropBank semantic role inventory.

These implicit role resolution models are used to illuminate characteristics currently being learned by implicit role resolution models, and to highlight issues that are still poorly represented. It is hoped that this thesis lays the groundwork for future work in implicit role resolution, and for addressing related linguistic questions which those models might enable.

Dedication

To my parents,

Tom and Martha

Acknowledgements

I owe my thanks to so many people over the years. First of all, I'd like to thank my advisors, Martha Palmer and Laura Michaelis, for all of their support and advice over these years. Martha has been a rock of support and advice through my entire time in Boulder, and I'd like to thank her for introducing me to the world of computational semantics, and providing me the projects, funding and prodding that got me to where I am today. I'd like to thank Laura Michaelis as well, for the exposure to construction-grammatical ideas, and who I could always trust to have insightful, probing questions about any topic in linguistics.

I'd like to thank all the rest of my committee members – Jim Martin, Barbara Fox, Mans Hulden and Ben Van Durme – for all your comments and suggestions, as well as advice during classes, CompSem meetings and conferences. I'd also like to thank Cynthia Clark and Paula Dufour, for all their administrative help over the years.

I'm indebted to Julia Bonn, Michael Regan, Sam Beer, and Kristin Wright-Bettner for annotation and collaboration that helped make this work possible, and Aous Mansouri for advise and expertise regarding the Arabic data. I'd also like to thank the many folks from the CompSem lab, my doctoral colleages in the linguistics department, and my masters cohort, for filling my time at CU with friendship and interesting conversation. A special thanks to my parents, to Kelly and Jay, and to my Chicago/UIUC friends.

Above all, I want to thank my fiancé Lauren for her endless support, love and encouragement.

Contents

Chapter

Tables

Table

Figures

Figure

Preface

This dissertation studies *implicit semantic roles*, instances where a linguistically mentioned event can be inferred to have a particular participant, but where there is no explicit syntactic encoding expressing that relationship to the participant. Such *implicit roles* can refer to many different kinds of missing participants – from vague and nonspecific referents, as in the missing food or meal in “He already ate”, to anaphoric references to previous mentioned referents, as in conversational statements such as “Told us to leave.”.

There are diverging, but related, goals for implicit roles in the linguistics community and computational linguistics research community. We might broadly assert that both would benefit from more fine-grained characterizations of how implicit roles are interpreted – the kinds of factors which we use to determine the correct referent in context, and how the grammar of a particular implicit role, as used in a particular context, shapes those factors. This thesis therefore explores the consequences of a more fine-grained characterization of these roles, across both theoretical and computational approaches. It is therefore in four core parts – characterizing a range of different cross-linguistically valid types of implicit roles, discussing how we might define language-specific constructions in relation to these types, presenting new annotated datasets in English using these types, and finally, by training computational models for English to find and resolve the referents of implicit roles.

Motivation behind Fine-grained Implicit Role Representation

If someone in English exclaims "We won!", most listeners will be able to identity *what was won* – generally, it would be a pragmatically odd thing to say if one had no information about the game that was won. Similarly, if one says "I've already eaten", most would assume that some kind of food was eaten. Such "implicit" participants[1] are a pervasive phenomenon in language, and the languages of the world show wide variation in what can be left implicit and what participants need to be explicitly stated in context.

The starting point is the generally accepted idea that implicit roles have different *interpretation types*. If we assume that there are implicitly a set of "referential instructions" for figuring out what the unstated arguments are, those instructions might refer to things identifiable because they are unique within a situation or genre (as in example 1), or because of general knowledge of other known facts in the common ground (as in discerning what is secret in example 2), or because, as with a pronoun use, the referent is recently mentioned, as in example 3.

(1) He shoots $\emptyset_{the\ ball}$ $\emptyset_{at\ the\ goal}$ and misses $\emptyset_{the\ goal}$

(2) She found out $\emptyset$

(3) *Questa mattina, Gianni ha visitato la mostra. Più tardi* $\emptyset$ *ha*
this morning Gianni has visit the exhibition. More late $\emptyset$ has visited
visitato l'università
the-university

"This morning, Gianni visited the exhibition. Later **(he)** visited the university (Samek-Lodovici 1996)"

[1] Implicit roles — or particular subtypes of implicit roles — are also referred to with a wide variety of other names, such as zero anaphora, omission, null complement anaphora or "dropped" roles; this dizzying array of related terms is reviewed in the next chapter.

Other implicit roles do not have such clear-cut referents, but instead can refer generically to all people (as in example 4) or refer to nonspecific, conventional referent that might normally fill that role, as in example 5. Modeling such non-referential implicit roles have their own challenges, as it can be unclear when to treat such nonspecific roles as a separate participant at all, or simply as a characteristic of event understanding.

(4) *Täällä ei saa uida*

here not-3S may-PRS swim

(One) must not swim here (Marati; Holmberg et al. 2008:5)

(5) I've already eaten $\emptyset_{a\ meal}$

Chapter 1 will propose a cross-lingusitically robust inventory of these different types of interpretation, which we propose can be a useful starting point in characterizing and modeling these implicit role uses.

The Definition and Scope of Implicit Roles

Implicit roles have been approached in many ways, and definitions for what counts as an implicit role can vary wildly. This thesis will assume the approach to implicit roles associated with cognitive and construction-grammatical analyses (Fillmore 1986; Goldberg 2001; Ruppenhofer & Michaelis 2014), in which these implicit roles are a semantic, pragmatic or psycholinguistic phenomenon. Such construction-grammatical approaches focus upon arguments that are pragmatically implied, regardless of whether they are syntactically missing — so that an eating event is probably construed as involving an *eater* and a *meal*, or a selling event may imply the existence of a *buyer*, a *seller*, and a *thing being sold*, even when the event itself is not a verb, such as with the nominal predicate "the sale". These approaches differ from the treatment of implicit roles in the generative traditions, which focus upon absent arguments that are considered to be syntactically "required" – treating those implicit roles as either phonologically deleted during production (Perlmutter 1971;

Mittwoch 1971; Perez-Leroux **et al.** 2017) or as phonologically empty forms (Taraldsen 1978). Those generative traditions therefore also only look at verbs, lacking rigid syntactic argument expectations for nominal predicates.

In order to maintain a tractable scope, this thesis is constrained to "core" arguments of explicitly mentioned events and situations, avoiding the complexities of looking at event locations (one can generally presume an inferrable location of most events (Recanati 2007)), avoiding times and tenses (which can be treated as implicit roles (Partee 1984)), and avoiding the issues posed by very peripheral roles of events, such as their manners, purposes, or frequencies.

This thesis also avoids "distantly instantiated" implicit roles – situations in which the referent, not mentioned in the same clause as the predicate, can be deterministically recovered from the syntax, as in control constructions. While there is no explicit subject of "climb" within "He wanted to attempt to **climb** the ladder.", determining that the agent of climb refers to "he" is a simple matter of control, thus out of the scope of the study, as are any other syntactically deterministic distant instantiation phenomena such as left-dislocation or wh-movement (Fillmore & Kay 1995).

Implicit Role Interpretation in Contrast with Licensing

Most existing work on implicit roles in linguistics frames conclusions regarding the *licensing* of those implicit roles – the grammatical rules which allow one to leave something implicit. This might be framed, alternatively, as arguments about the factors leading to the *referential choice* of using an implicit role rather than a pronoun or full noun phrase. We instead focus upon factors that lead to a particular *interpretation* of an implicit role – information about what the implicit role refers to.

There is a tension between discussing licensing and interpretation, because they are fundamentally related. Givon (2018) states this as the generalization that implicit role licensing (the situations where implicit roles are appropriate) are some combination of predictability (it is easy to find the right referent using the implicit role) and the importance of the mention (whether one is using an explicit form to emphasize the participant, or omitting it to de-emphasize the participant).

Many prior works on licensing of implicit roles have proposed "single-factor" approaches to how implicit roles are licensed, and therefore ascribe implicit role usage to some version of predictability or importance. Resnik(1993, 1996) showed that implicit object role constructions are correlated with how predictable the implicit object's type would be, while others have linked licensing to ways of deprofiling (de-emphasizing the importance of) the object (Goldberg 2001) or emphasizing the event itself (Rice 1988). Others have focused upon whether the argument is necessary to an internal logical structure of an event (Rappaport Hovav & Levin 1998).

This thesis, in line with the more construction-grammatical approach to implicit roles (Fillmore 1986; Goldberg 2001; Ruppenhofer & Michaelis 2014), assumes that there is no simple equation that governs all implicit roles, but rather that each language has its own toolkit of constructions for instantiating arguments of a predicate, including "implicit role constructions", and each implicit role construction defines the rules for what that role refers to and the pragmatic consequences of expressing that referent implicitly. This approach means that each choice between implicit and explicit roles may have referential consequences (what the referent is, and how hard it is to determine that) and over a range of other aspectual, information-structural, and discourse structural features (Olsen & Resnik 1997; Iida 1996; Tao 1996). We assume, however, that we cannot truly discuss these factors other than interpretation if we don't first model the interpretive behavior of implicit roles. This thesis focuses upon how we define these "interpretation" biases of implicit roles, leaving the other factors which might effect implicit role licensing for future work; while we discuss formal models of implicit role licensing briefly in Chapter 2, these do not focus upon any of the non-referential constraints which one might expect to apply to implicit role usage.

Structure of This Thesis

The following work consists of six chapters. Chapter 1 overviews a range of approaches to implicit roles, across linguistics and related fields. Chapter 2 proposes an inventory of different types of implicit role interpretations, so that one could describe the landscape of implicit role constructions in a language by using a simple set of labels which might enable crosslinguistic comparison. Chapter

3 provides three studies of using that inventory in defining language-specific constructions and in cross-linguistic comparison, by outlining a mapping of Spoken Arabic implicit roles into these types, outlining formal models for implicit role construction licensing using these types, and by presenting basic models for comparing implicit role constructions across languages using semantic map approaches. Chapter 4 outlines the landscape of data annotated with implicit role types, and presents new corpora which increase the amount of annotated data available for English implicit roles. Chapter 5 then presents a general model of implicit role resolution (following a range of existing work in English (Palmer **et al.** 1986; Gerber & Chai 2010; Tonelli & Delmonte 2011; Wang **et al.** 2018). Finally, Chapter 6 will overview the conclusions of the current work and attempt to define a roadmap of next steps to be made regarding implicit roles, both theoretically and computationally.

Chapter 1

Related Work on Implicit Role Phenomena

Implicit roles have been approached in many ways – across psycholinguistics, pragmatics, philosophy of language, and a range of different linguistic schools of thought. These approaches, although unified in examining unstated content, vary both in their definition of implicit content, and also in the questions that they ask.

This thesis focuses upon better characterizing *interpretation* behaviors of implicit roles – what an implicit role refers to, and what biases are provided by the grammar about its interpretation. However, this topic of interpretation has not been the primary topic of prior research regarding implicit roles. Alongside interpretations, one might characterize many of these studies of implicit roles as studying three kinds of questions – the question of *licensing* implicit roles (why use an implicit role rather than an explicit form), the question of the *extent of implicitness* of an implicit role (how strong is the implication of that implicit role), and finally what one call the *mechanism* of implication – whether the implicit role is something emergent out of the syntactic system, semantic inference, or pragmatic interpretation.

This thesis will largely follow the assumptions of construction-grammatical approaches mentioned below (Fillmore 1986; Goldberg 2001; Ruppenhofer & Michaelis 2014) when necessary, but generally will attempt to remain neutral in such debates – instead positing inventories of interpretation types which might make all such directions of research more precise and typologically informed.

1.1 Prior Analysis of Implicit Roles

1.1.1 Analyses of Implicit Roles as Missing or Deleted Syntactic Positions in Formal Linguistics

Generative models of syntax view implicit roles as phonologically null (or deleted) positions within a syntactic tree – treating them as a subset of the larger set of "empty categories" (Chomsky 1981), which includes the implicit roles discussed here as well as non-referential empty categories and traces of "moved" or "displaced" constituents. As these represent null elements within a syntactic tree, they are referred to through those grammatical roles, such as "null subjects", "deleted subjects" or "deleted objects" (or as a group, with "zero anaphora" or "zero pronouns" (Huang 1984)). This formulation means that implicit roles in Chomksian approaches do not refer to all inferrable semantic arguments, or to a mismatch between the predicator's array of semantic roles and its syntactic valence, but to the specific subset of inferrable content where the syntactic model can be postulated to always have a constituent at a particular position – such as subjects and occasionally objects of transitive verbs. In that manner, the null grammatical roles in Chomskian grammatical traditions are firmly tied to grammar-wide assertions about whether those roles must exist, and whether such an argument must be explicit (such as mechanisms such as Extended Projection Principle or Null Subject Parameter for defining how a language expects subjects to be syntactically instantiated). How "implicit roles" work in generative traditions is therefore often quite dependent upon the syntactic position – such that languages with morphological agreement have been modeled as being "null subject" phenomena influenced by a null subject parameter (Rizzi 1982; Taraldsen 1978), implicit subjects of languages without morphological subject agreement have been modeled as a different phenomena entirely as a kind of "topic drop" (Huang 1984), and being sometimes a separate phenomena of "argument ellipsis" (Oku 1998), with implicit complements being another issue of "null complement anaphora" or "VP-ellipsis" (discussed below). This description of how implicit roles work in generative models is necessarily an approximation, as well, as the relation may differ in more recent analyses within Minimalist Program syntax (Barbosa 2011). However, both

Minimalism and its generative forebears share the more general commitment to a syntactocentric representation (Jackendoff 1999), wherein syntax makes no direct links to semantic categories, so that if one is to model semantic roles as being implied in a grammar, they must be captured through actual positions within syntactic structure. Other approaches (such as the construction-grammatical approach assumed here) have no need for such elements, as the semantic roles may be directly connected to the grammar (for example, as feature values within an AVM), and indexation may be independent of the syntactic positions.

1.1.2 Literature on Complement Anaphora

A large set of studies have looked at a range of phenomena in which an unstated event in complement position can be inferred from the prior text.

(6) I asked Bill to leave, but he refused $\emptyset_{to \quad leave}$

(7) Sue was attempting to kiss a gorilla, and Harry didn't approve $\emptyset_{of \quad it}$ (Kasher 1998)

The major question that is focused upon in this literature is whether these implicit roles are a part of the syntax, or are implicit because of a pragmatic mechanism. Sag & Hankamer (1976) introduced the notion of modeling some of these as syntactic phenomena ("surface anaphora") when there is a clear linguistic antecedent, as in example 8 below, and contrasted that with null complements referring to "pragmatic" co-present events that are not linguistically anaphoric (first noted by Shopen (1972)), as in example 9. This can be viewed as a testable distinction relevant to the grammar of English, as one can use both NCA and VP-deletion constructions for surface anaphora (examples 8a and 8b, respectively), but where the pragmatically controlled anaphora disallows VP-deletion constructions, as in example 9b.

(8) Sag: Why don't you stuff that ball through that hoop?
   a. Hankamer: I'm trying.
   b. Hankamer: I'm trying to. (Sag & Hankamer 1976)

(9) (Observing Hankamer attempting to stuff a 2“ ball through 6“ hoop)

a. Sag: I don’t see why you even try.

b. *I don’t see why you even try to (Sag & Hankamer 1976)

Tests such as this – and related questions regarding the ability to use “sloppy anaphora” constructions – are used in this literature primarily to test whether a given construction is syntactic or pragmatic in nature.

This vein of literature often also looks at implicit *predicates*, but such topics (and related issues in VP ellipsis) are outside the scope of the current work. While the current work considers predicative mentions when they are the semantic arguments of other predicates (as in the examples above), we do not address the interaction between unstated predicates and their arguments, as in the study of gapping or stripping constructions in English, or ellipsis of duplicated predicates during comparative constructions, such as seen in examples 10 and 11.

(10) Alan likes to play volleyball, but not Sally (Hankamer and Sag) (Sag & Hankamer 1976)

(11) Hankamer: Ivan is now going to peel an apple. Sag: And Jorge, an orange. (Sag & Hankamer 1976)

1.1.3 Studies of Licensing of Implicit Roles

There are many studies which may treat implicit roles as omission of particular syntactic positions, but which do not focus upon their status within a syntactic tree, instead looking at how implicit roles are *licensed* – what factors determine whether one will leave an argument implicit.

One set of these studies looks at the influence of aspect or predictability upon implicit role licensing. Mittwoch noted how some implicit object constructions such as “he ate” having aspectual implications (Mittwoch 1971; Mittwoch 1982), and later works such as Resnik (1993) and Olsen & Resnik (1997) have looked at “object omission” in order to discuss interactions between the likelihood of implicit objects and the predictability of the object role for those predicates, and how that relates to issues like aspect.

Others have looked at interaction between the use of implicit roles and some expression of whether the argument (or the event) should be given importance or unimportance. Rice (1988) treated implicit object constructions as acting to give emphasis to the occurrence of the event. Goldberg, Goldberg (2001, 2005) represented implicit object constructions as enacting the "de-profiling" of the object. Similarly, Givón (2017) characterized implicit role licensing as generally being an interaction between expression of the importance of an argument and expression of the predictability of that argument.

Finally, a range of works in typology, discourse-functional syntax or pragmatics have viewed the licensing of implicit roles in terms of the impact of implicit role use upon discourse structure or discourse coherence. Tao (1993), following approaches connecting reference form to discourse structure in Fox (1984), connected implicit role usage to the assertion of coherence with prior text. Related work over the years on implicit causality and discourse relations (Ueno & Kehler 2010), or upon implicit roles in Centering Theroy(Walker **et al.** 1994) have studied similar questions.

1.1.4 General Semantic and Pragmatic Treatments of Implicit Roles

More generally, a range of cognitive and construction-grammatical analyses (Fillmore 1986; Goldberg 2001; Ruppenhofer & Michaelis 2014) have looked at implicit roles. Such construction-grammatical approaches focus characterize implicit roles as semantically and pragmatically implied, regardless of whether they are syntactically missing — so that an eating event is probably construed as involving an *eater* and a *meal*, or a selling event may imply the existence of a *buyer*, a *seller*, and a *thing being sold*, even when the event itself is not a verb, such as with the nominal predicate "the sale". These approaches refer to implicit roles as "null instantiation", following the generalization that these arguments differ from normal arguments only in the *manner of how they are instantiated* – splitting up the ways that one might instantiate an argument into *local instantiation* (arguments directly within the scope of a predicate), long-distance but syntactically deterministic relations (*distant instantiation* for things like filler-gap structures and *coinstantiation* for control) and then implicit roles as *null instantiation*. These null instantiation characterizations actually do discuss the *interpretation* of implicit roles – splitting up instantiation into definite, indefinite and constructional null instantiations – as discussed in the next chapter.

Other approaches outside of construction-grammar have also focused upon semantic or pragmatic characterizations of implicit roles, instead of syntactic representations. Ono & Thompson (2000) also presented arguments against syntactically-defined implicit roles, noting that "referents, like much else in linguistic communication, would be inferred from the entire range of semantic and pragmatic factors which are present in the actual interactions in which speakers engage in everyday life" (Ono & Thompson 2000:489). A number of approaches in computational linguistics have also looked at implicit roles not as syntactic representations, but in terms of semantically essential roles according to a particular lexical representation of a predicate. Palmer **et al.** (1986) modeled implicit roles in reference to a set of "essential roles" expected for each predicate, using that approach. Tetreault (2002) also focused upon implicit role interpretation using such a semantic definition, stating that "Verbs have certain required roles, which refer to discourse entities, that are necessary for comprehending the verb phrase".

1.1.5 Implicit Roles as in Pragmatics and Philosophy of Language

There are also many studies in pragmatics and philosophy of language regarding "unarticulated constituents", which can be traced back to Sellars **et al.** (1956). These "unarticulated constituents" have been a major source of examples in larger discussions regarding the boundary between pragmatic implicature and more rigid representations of the direct semantic content of a sentence. This topic therefore tends to encompass not only the kind of implicit roles discussed in this thesis, but also a range of other unarticulated content, such as the implicit times and locations of physical events (Perry & Blackburn 1986; Partee 1984), the implicit referential matter required to make definite referents uniquely identifiable (as in examples 17 and 18 below), and implicit modifiers, manners, and domains of comparison. Many of these questions concern two related questions – one being what we referred to above as the *extent of implicitness* for these more peripheral implicit arguments, and the related question regarding these are "implicit" because of a syntactic mechanism, or through pragmatic implicature.

This first question regarding the *extent of implicitness* looks, in its simplest form, at distinguishing the implicit content which should be viewed as being part of the intended, expressed content of an utterance – Grice's notion of "what is said" – from the range of weaker implicatures that might be derived from an utterance. Such distinctions can be made more nuanced, beyond that, depending upon the granularity of one's typology of different kinds of implied content. Examples 12 – 13 illustrate the kind of content which are often treated as being part of the logical, expressed content:

(12) John is ready (for the interview)

(13) John hasn't completed (his PhD thesis)

Much of the meat of discussions in this vein have focused upon examples such as examples 14 – 18, where one can argue about whether various examples pass various tests – such as whether interpretation of a given argument is necessary for evaluating the truth-conditional content of the utterance (Perry & Blackburn 1986; Bach 1994), with others appealing to more natural tests, such as claiming that material is implicit if a layperson would find it inferrable (Recanati 2002).

(14) John is tall (for an average adult male)

(15) It is raining (outside / in Seattle)

(16) It's summer (in Australia) (in springtime) (at nine o'clock)

(17) You should clean the table (that is in our kitchen)

(18) Everyone (in the tour group) was sick (Huang 2018)

Others have discussed ways of characterizing the other end of this scale – with Searle (1980) discussing "background assumptions" , or otherwise discussing meanings that are "meant but are not conveyed" (Carston 2002). Such examples include weakly implied assumptions which might be provided by world knowledge, such as guesses regarding the manner, duration and instruments of an event. Sternau **et al.** (2015) present one appealing recent way of approaching such questions – characterizing meaning with a scale of four levels, ranging from *linguistic meaning*, to *explicature*, to *strong implicature*, and finally to *weak implicature*.

Such characterizations are hard to view as separate from the mechanistic question of how "unarticulated constituents" are inferred. Some characterize these as being necessarily present and deleted during the syntactic process (Stanley & Gendler Szabó 2000; Stanley 2002), in a manner similar to generative linguistic models mentioned above (motivated by the "argument from binding" (Elbourne 2008) which suggests that co-instantiation between implicit arguments proves the syntactic nature of these implicit roles). Many other approaches treat these unarticulated constituents as being emergent out of pragmatic inference processes, either Relevance-theoretic (Sperber & Wilson 1986), or Gricean/neo-Gricean inference (Huang 2018).

1.1.6 Experimental Study of Implicit Roles

Garrod & Terras (2000) discussed implicit roles (referred to as *discourse roles*) in situations such as example 19, where one might view *drive* as introducing the *car* participant.

(19) Keith drove $\emptyset_i$ to London yesterday. The car$_i$ kept overheating.

This illustrates the kind of focus taken in psycholinguistics, of studying implicit roles often in the context of studying the general timecourse of reference and lexical activation. The study of Garrod & Terras (2000) studied that topic with eye-tracking measures – modulating situations where a given lexical item alone might introduce implicit arguments (such as "write" introducing the instrument of a "pen") along with situations wherein those lexical assumptions might be modified (e.g. where "pen" may be replaced by "chalk" in a context such as "the teacher wrote the exercise on the blackboard"). Such a study (and similar works such as Cook & Myers (2004)) can provide insights into the time-course of activation of such an implicit referent, but it can be hard to separate those conclusions from broader question regarding how words are primed and accessed. Such studies often must use indefinite implicit roles (where a referent is introduced, rather than anaphoric) so that one can measure the "implicit role" simply by measuring reactions to additional anaphoric reference to that participant. For example, Mauner & Tanenhaus (1995) studied implicit agents (followed by Mauner **et al.** (2002) with implicit instruments and locations) by showing that while passive voice verbs tended to allow later reference to the agent (using purpose infinitives, such as "this ship was sunk to hide the treasure"), middle voice with *have* tended to remove that implicit agent ("this ship has sunk to hide the treasure").

Tao & Healy (2005) looked at implicit roles (zero anaphora, in their terminology) with a different methodology, studying the comprehension of anaphoric implicit roles by having participants read a passage in which arguments were artificially removed, and studying whether those participants could answer comprehension questions about the implicit content. This study continued a set of studies (Tao & Healy 1996; Tao & Healy 1998) regarding whether speakers of languages with heavy use of implicit role constructions (Chinese and Japanese) could be claimed to transfer any amount of those skills over to English, when contrasted with speakers of languages with fewer implicit roles (such as English or Dutch) – studies finding that the Chinese and Japanese participants were generally able to have better comprehension of even English sentences which were transformed to have Chinese-like implicit core arguments.

All such studies have generally provided a small set of experimental conclusions regarding implicit roles, but remain limited in their conclusions – serving primarily as clues to the larger puzzles of lexical access and reference.

1.1.7 Licensing Studies in Language Acquisition and Sociolinguistics

Finally, a number of studies characterizing implicit roles have focused upon modeling the *licensing* question regarding implicit roles, not within particular linguistic theories, but by building more statistically oriented regression models using hand-annotation features such as referential distance, givenness and grammatical role status. A number of language-specific studies of referential choice have been applied within sociolinguistics, particularly in the study of implicit subject constructions in dialects of Spanish, finding variation in both the frequency of implicit subject use and in the exact features which license implicit subjects (Alfaraz 2015; Flores-Ferrán 2004; Otheguy **et al.** 2007; Erker & Guy 2012).

Other studies have sought to characterize implicit role usages in child language acquisition, an ongoing topic due to the fact that children are known to use implicit roles (sometimes referred to as "argument omission" in that literature) both more frequently and in different contexts than adults of the same language(Allen **et al.** 2008; Allen **et al.** 2015). Notably, recent work in implicit role usage (including studies of acquisition of Inuktitut) has found that what was once thought to be the aberrant use of implicit roles in discourse-new contexts may instead reflect a tendency in children to use implicit roles in contexts of *joint attention* – where referents may be lingusitically new to the discourse, but are highly salient and cognitively activated (Skarabela & Allen 2010).

1.1.8 Overview

Such an overview illustrates the breadth of different approaches to this single topic of implicit roles. One of the ongoing issues in looking at these different studies, however, is that many such studies necessarily focus upon particular phenomena, rather than the full range of implicit content. We suggest that having a richer understanding of the different kinds of implicit roles is necessary in order to synthesize all of these research works into larger generalizations about how implicit roles are used in language. Therefore, while the following chapter embraces the assumptions of construction-grammatical approaches discussed above – treating implicit roles as emerging from complex interactions of syntax, semantics and pragmatics, and as being licensed by language-specific constructions – the discussion will generally attempt to propose a set of interpretation types which might be compatible with a wide range of different ways of studying implicit roles, so that there might be greater syntehsis across these very different approaches to the topic.

1.2 Scope of the Current Thesis

In order to maintain a tractable scope, this thesis is constrained to "core" arguments of explicitly mentioned events and situations, avoiding the complexities of looking at event locations or times(Recanati 2007; Partee 1984)), and avoiding the issues posed by very peripheral roles of events, such as their manners, purposes, or frequencies. This avoids many of the weak implicatures discussed in some of the pragmatics literature.

This thesis also avoids "distantly instantiated" implicit roles – situations in which the referent, not mentioned in the same clause as the predicate, can be deterministically recovered from the syntax, as in control constructions. While there is no explicit subject of "climb" within "He wanted to attempt to **climb** the ladder.", determining that the agent of climb refers to "he" is a simple matter of control, thus out of the scope of the study, as are any other syntactically deterministic distant instantiation phenomena such as left-dislocation or wh-movement (Fillmore & Kay 1995). Doing so allows one to avoid a focus upon describing well-studied phenomena which would be more dependent upon a particulation theory of syntax, and allows a focus upon a more general categorization of these interpretation types.

Chapter 2

Categorization of Implicit Role Interpretations

This chapter proposes a categorization of different kinds of these implicit roles – different ways that something can be left implicit. This is not intended as a perfect inventory, but rather as a set of useful comparative concepts for describing implicit role interpretations which might allow one to outline the coarse-grained implicit role constructions of a language. This is an inventory of *interpretation behavior* for implicit roles, also known as *referent types* (Becker 2018). These characterize the kinds of referential situations that might be characterized by a particular implicit role, or similarly, characterize the kinds of explicit referential forms which that implicit role might characterize – such as the difference between reference to a cat using “it” or “the cat” or “a cat” or “that cat”.

We characterize eleven interpretation types of implicit roles, but can generally present them in three groups – those which are clearly definite (where the referent is known to the addressee), those which are clearly indefinite (the referent is clearly not known), and a number of kinds of implicit roles which are on the border between the two. While the inventory proposed here is a contribution of this thesis itself, and not proposed in other works, most of the actual distinctions between particular categories have been proposed prior work, such that the actual contribution is the synthesis of different analytic claims into a single set of types.

2.0.1 Prior typologies of implicit role interpretation and reference types

Most approaches in which implicit roles are categorized into different interpretation types have focused upon making distinctions between "definite" and "indefinite" implicit roles (Fillmore 1986), or making similar distinctions, such as between anaphoric and existential (Condoravdi & Gawron 1996), specific vs nonspecific (Allerton 1975), or between recoverable and non-recoverable (Moor **et al.** 2013b). Fillmore's test of this distinction for implicit roles was a *profession of ignorance* test, wherein "indefinite" implicit roles generally allow a speaker to admit a lack of knowledge of the implicit referent, as in "He ate already; I wonder what he ate", but wherein one cannot say the same for definite implicit roles, as in "He found out ∅! I wonder what he found out". Additional categories have been sometimes proposed – such as generic implicit roles (Lyngfelt 2012) and "Free" null instantiation (Lambrecht & Lemoine 2005), or a set of four types from Zifoun (1997; via Ruppenhofer 2005), who proposed four types: *Situational* (recoverable from speech context, including deixis and genre); *Empratical* ("predicates that are missing can be inferred from joint activity they are engaged in"); *Phatic Ellipsis* (ellipsis when a speaker gives up on production); and *Structural* (omissions and ellipsis licensed in specific text types due to considerations of economy and condensation).

Other typologies, while often defining distinctions used throughout this chapter, defined implicit roles in language-specific inventories of implicit role construction types. Ruppenhofer (2004:368) and Lambrecht and Lemoine (2005) also both proposed inventories of implicit role licensing, providing inventories of implicit role constructions in English and French. These therefore conflated comparable categorizations of implicit role interpretation with language-specific grounding of those kinds of implicit role constructions in the language in question – as in the English "Diary style omission" or "labelese" genre-based omissions in Ruppenhofer 2005, or the specific types for of implicit roles in French noted in Lambrecht and Lemoine, such as "recipient of communication event" or "experiencer".

A more relevant comparison, however, may be to inventories of explicit referential markers

Dryer (2014)	anaphoric definite	non-anaphoric definite	pragmatically specific indefinite	semantically specific indefinite	semantically nonspecific indefinite
Givon (1978)	anaphoric definite	non-anaphoric definite, generic	referential indefinite	referential nondefinite	nonreferential object, predicate nominal
Becker (2018)	anaphoric, deictic	specific, generic, bridging	establishing, abs-unique, sit.unique	nonspecific	
	definite	←		→	*indefinite*

Table 2.1: Inventories of article types from Dryer (2014), Givon (1978) and referent types of Becker (2018). All are roughly oriented from definite to indefinite, but strict hierarchies not maintained.

such as articles, demonstratives or pronouns – which, like implicit roles, are sometimes lumped into "definite" and "indefinite", but which also tend to have more cross-linguistic variation than such a split might imply. The "reference hierarchy" of Dryer (2014) , shown in Table 2.1, provides one inventory of such types, organized so that a given article in a language will generally be used for a contiguous part of the hierarchy, and is partially based upon Givon's earlier "wheel of reference" (1978), whose types are also illustrated. A more recent inventory of types (Becker 2018) presents a larger inventory of such types.

Other approaches have proposed single scales from maximally reduced to unreduced forms, which also can be viewed as inventories of different "referent types" (and which will be discussed further in this Chapter and the next), such as Ariel's accessibility hiearchy (Ariel 2013; Ariel 1988; Ariel 2004), the Givenness hierarchy (Gundel **et al.** 1993), topic continuity (Givón 1983), or assumed familiarity (Prince 1981). Over all of these ways of splitting up the space of "how to refer to a referent", most of the distinctions which are relevant to implicit roles have been proposed at one time or another, but no single system seems to cleanly capture the entire space of implicit role referent types.

2.1 Distinctions of implicit role types – Definite Implicit Role Types

One might start with the difference between four kinds of definite – and often anaphoric – implicit role constructions. The SCRIPT-INFERRABLE type of implicit role characterizes the kind of definite implicit object constructions seen in English, as illustrated in examples 20 – 21. The referent of such implicit roles is determined by general understanding of a scene and facts in the common ground; such that the theme of "found out" involves looking to prior context for the fact or thing being kept secret, or thing signed in example 21 requires consulting prior contexts for things that were bid, offered or considered:

(20) She found out $\emptyset$ (Fillmore 1986)

(21) *Elle a signé*

she PERF signed

"She signed $\emptyset_{\text{the contract/deal}}$"

In this contrast, we can note languages where implicit object constructions are much more constrained to refer to recently mentioned, prominent referents (what would be "highly accessible" in the Ariel accessibility hierarchy or "in focus" in the Givenness Hierarchy (Gundel **et al.** 1993). Such SALIENT/RECENT antecedents are illustrated in examples 22–23 from Japanese and French, and might often be replaced with a pronominal form:

(22) *Avant, j' avais mon dossier á Jester, mais j' ai enlevá* $\emptyset_{\textit{the file}}$.

before i have my file LOC Jester but I PERF take.away $\emptyset_{\text{the file}}$

"Before I had my file at Jester but I took (it) away. (French; Lambrecht and Lemoine 2005)"

(23) *mukasi Bill*$_j$*-o osieta sensei-mo* $\emptyset_j$ *homete iru*

years.ago Bill-acc taught teacher-also $\emptyset_j$ is praising

/ "[The teacher who taught Bill$_j$ years ago] too is praising (him)." (Japanese; Hoji 1998)

Another kind of definite interpretation is when the referent is DEICTIC, referring to a co-present referent (often a speaker or addressee), as in example 24 from Japanese:

(24) $\emptyset_I$ *hon-wo* *yon-da*

$\emptyset_{\mathrm{I}}$ book-obj read-PAST

"(I) read a book" (Nakaiwa and Shirai 1996)

A final class of clearly definite implicit roles are REMEMBERED ROLE implicit roles – instances where the semantic role (as a link between the event and the participant) has already been asserted, because the event has already been discussed. These therefore require neither the pragmatic rules of SCRIPT-INFERRABLE implicit roles nor local syntactic context, but simply require one to link the current predicate to a prior mention, as in the arguments of the second mention of "losses" in example 25, which can be recovered by linking it to the prior event of "losses":

(25) **The network** had been expected to have **losses of as much as $20 million on baseball this year**. It isn't clear how much those **losses** $\emptyset_{of\ the\ network}$ $\emptyset_{of\ that\ amount}$ $\emptyset_{on\ baseball}$ may widen because of the short Series. (Laparra & Rigau 2013)

2.1.1 Details of Script-inferrable, pragmatic implicit roles

The SCRIPT-INFERRABLE roles mentioned above do not have any constraints, nor must they even be linguistically anaphoric. For example, English implicit object constructions discussed in Fillmore (1986) may refer to definite referents not linguistically mentioned in the recent context. To illustrate, one might walk into a room and utter "we won!" or "Bill resigned!", referring entirely to facts in the common ground shared between the speaker and addressees. These may therefore be best exemplified by definite implicit object constructions in English (Fillmore 1986; Ruppenhofer & Michaelis 2014; Glass 2014), and most implicit object constructions in French (Lambrecht & Lemoine 2005). Lambrecht and Lemoine (2005) call some of the cases they studied "frame-induced" in French, noting that although one can leave an object implicit when it fits into a clear frame in context (as in example 26), leaving objects implicit is less viable when the event does not cleanly link to a single scenario or frame, as in example 27.

(26) *Elle* *a* *signé*

she PERF signed

"She signed $\emptyset_{\text{the contract/deal}}$"

(27) *ils* *ont* *battu*

they PERF beat

? They beat (them)

Such implicit role instances also dominate the implicit role interpretation of less "core" roles, such as the arguments of eventive nouns or adjectives – a very common issue for computational models of implicit role resolution.

2.1.2 Salient/recent implicit role referents

The class of pronoun-like SALIENT/RECENT implicit role interpretations is most commonly seen in prototypical "null subject" constructions, where a subject is left implicit (often combined with morphological indexing of its gender or number) and refers to a recently mentioned referent. Examples 28 – 29 illustrate examples of such implicit subject constructions; while they differ slightly in the exact details from language to language (as discussed in Chapter 2), we suggest that all such instances share the same rough behavior – one would want to put them into the same general comparative class:

(28) *Questa mattina, Gianni ha visitato la mostra. Più tardi ∅ ha visitato*
this morning Gianni has visit the exhibition. More late ∅ has visited
l'università
the-university
"This morning, Gianni visited the exhibition. Later he visited the university (Samek-Lodovici 1996)"

(29) *Juan$_i$ llegaba a casa. ∅$_i$ Tenía las llaves.*
John arrive.3s.imperf to house. ∅$_i$ have-1/3sg.imperf the keys
"John was arriving home. **He** had the keys (Cole 2010:280)"

As noted above, we can also see this for implicit object constructions in some languages, such as Hebrew (example 30, Landau 2018), Latin (example 31; Luraghi (1997)), Finnish Sign Language (example 32; Jantunen (2013)), Japanese (Kayama 2003), and Ancient Greek and Sanksrit (Keydana & Luraghi 2012). These do not pattern like the more script-inferrable mentions used in English, but refer to very prominent and recently mentioned prior referents (sometimes limited to previously mentioned objects):

(30) a. *adayin eyn li manxe la-doktorat*
still no to.me advisor to.the-doctorate
"A: I still don't have a PhD advisor" "

b. *lifney ŝe-ata moce ∅, ata carix nose*
before that-you find ∅ you need a
' B: "Before you find (one), you need a topic (Landau 2018)"

(31) *quo cum Catilina*$_i$ *venisset, quis eum*$_i$ *senator appellavit? qui* $\emptyset_i$ *salutavit?*

"although Catilina was there, who among senators called him? Who greeted **(him)**? (Luraghi 1997:4; Cic. Cat. 2.12)"

(32) MAN$_i$ GO-IN / WOMAN LOOK-AT *emptyset*$_i$'

The man goes in and the woman looks at (the man/him).'(Finnish Sign Language; Jantunen 2013:317)

One extreme subtype of these SALIENT/RECENT instances are instances where the implicit role is nearly deterministic in referring to a prior grammatical position, such as referring to the last mentioned subject, last mentioned object, or last mentioned topic. These have recieved a variety of analyses which actually formalize them into grammatical constructions (e.g. modeling implicit objects as "left node raising" (Yatabe 2001) or otherwise as a grammatically defined (Luraghi 1997)), or led to discussions of the effect of lingering gramamtical roles across sentence boundaries(Auer 2014;534); this also can be relevant to long "null topic" chains seen in languages such as Japanese (Yatabe 2001):

(33) a. *"harii pottaa" no eega-ga*$_i$ *ninki-da*

Harry Potter of movie-nom popular-pres

The movie "Harry Potter" is popular

b. *takusan no hito-ga* $\emptyset_i$ *mita soo da*

many of person-NOM $\emptyset_i$ watched hear-say

"I heard many people watched (it) "

c. *tomodachi mo* $\emptyset_i$ *omoshirokatta to itte-ita*

friends too $\emptyset_i$ interesting-past comp say-pst.prg

My friends also said that (it) was interesting.

d. *watashi mo raishuu* $\emptyset_i$ *miru tsumori-da*

i also next.week $\emptyset_i$ watch intend-to

I am going to watch (it) next week too

We refer to such types with the general label of SALIENT/RECENT in part to acknowledge the ambiguities involving this type. Various approaches might view the exact dimension which is used for these implicit roles as being the topicality of the referent, its familiarity, the cognitive "activation" or "accessibility" of the referent, or even whether the current sentence expresses discourse continuity, or continuity of topic, with the preceding text (Givón 1983; Ariel 1988; Prince 1992; Gundel **et al.** 1993; Ariel 2004). Kaiser & Trueswell (2008) note that "It is generally agreed that there exists a correlation between the type of referential form used to refer to an entity and the level of salience/prominence of the entity". However, not only is it unclear which characterization is correct, but it is very likely that languages vary in which of these factors is most important in implicit role resolution – sometimes referred to as the "form-specific multiple constraints" idea of Kaiser & Trueswell (2008). Because of this, we suggest that the best way to discuss implicit roles with such SALIENT/RECENT behavior is to first determine which implicit role constructions have this kind of interpretation, and only later to compare different kinds of such constructions for more precise details of their interpretation, such as the relative importance of grammatical parallelism (Carminati 2002), topicality or discourse coherence.

2.1.3 Deictic Implicit Roles

Deictic implicit roles in general – and reference to the speaker and addressee in particular – are another extremely common kind of definite implicit role. Although in many languages these are commonly encoded by morphological person agreement or specific constructions (as in the English imperative), there are many languages where reference to the speaker or addressee must be inferred purely from context (as in examples 34 or example 35.

(34) $\emptyset_I$ *hon-wo* *yon-da*

$\emptyset_{\mathrm{I}}$ book-obj read-PAST

"(I) read a book" (Nakaiwa & Shirai 1996)

(35) $\emptyset_{we}$ Took im fer a ride on that'n Bill said thet he wz et least goin eighty miles'n hour (Oh 2006:823)

This is commonly the case for arguments of predicative adjectives, or other situations in which a deictic experiencer might be expressed (although, as noted below, it can be ambiguous with generic interpretations):

(36) Most interesting $\emptyset_{to \quad me}$!

2.1.4 Remembered Roles of Event Reference

A final category for clearly knowable implicit roles occurs with REMEMBERED ROLES, where the predicate has already been previously mentioned, and the relation to the current participant has already been mentioned, so that the semantic role could be simply "remembered" rather than directly inferred. This is common for English for mentions such as definite eventive nouns, where one can clearly identify that the referent should be found in prior context (as in the theme of "sale" in 37 (Gerber & Chai 2010)), or in conversation, as with question-answer sequences where an event is under discussion, as in example 38 from Japanese.

(37) The real estate and mortgage banking concern had hoped to use proceeds from the sale $\emptyset_{theme}$ to reduce its debt .

(38) a. *joshidaisee ni wa mieru?*

college.girl like TOP look

"Do (I) look like a college girl, then?"

b. *hyottoshitara mieru tomo ieru kana=*

maybe look TOP can.say FP

(one) may possibly say that (you) look (like a college girl).

As it is common to reframe events or refer to a slightly different facet of the same event or a subevent of that event, this type of implicit role can have some ambiguities with the more general SCRIPT-INFERRABLE implicit role type above, as many events are related not through strict identity but a kind of bridging or quasi-identity relationships (Poesio **et al.** 1997; Hovy **et al.** 2013). Thus while one might quibble in example 39 regarding whether "One morning he spent in town" and "this one excursion" are exactly coreferential, they are clearly related as different parts of the same roughly connected set of events for the sake of inferring who was travelling and where they went:

(39) "One morning he$_i$ spent in town$_j$, and I learned from a casual reference that he had visited the British Museum. "

"...Save for this one excursion $\emptyset_{by\ him}$ $\emptyset_{into\ town}$, he spent his days in long and often solitary walks"

2.2 Edge Cases: Interpretations Between Definite and Indefinite

There are a variety of ways in which an implicit role may be "known" – in the sense that something about the referent's type or identity is known – but where it does not link to an actual previously mentioned referent. We focus on four types – sloppy anaphora/bridging instances (where it refers to a prior antecedent, but introduces a new referent analogous to that antecedent) as in example 40a; situationally unique GENRE-BASED DEFAULTS where a referent is unique in a context (as in implicit balls in a sports context, as in example 41), and TYPE-IDENTIFIABLE implicit role interpretations, where the referent is nonspecific but its type is "known", most prototypically with very conventionalized referents as in example 42, and finally, implicit roles used in generalizations or characterizing statements where the implicit role refers to all mentions of a type – most commonly referring to people in general, as in example 43:

(40) a. *John-wa* *zibun-no* *tegami-o* *suteta*

John-TOP self-gen letter-ACC discarded

"John$_i$ threw out his$_i$ own letters"

b. *Mary-mo* $\emptyset$ *suteta*

Mary-also $\emptyset$ discarded

"Mary also threw out John's letters / Mary also threw out her letters (Takahashi 2011)"

(41) Juice Williams keeps $\emptyset_{the\ ball}$. ((Ruppenhofer & Michaelis 2014), p58)

(42) Bill started drinking $\emptyset_{\text{alcoholic beverages}}$ again

(43) *Gianni é sempre pronto ad accontentare* $\emptyset_{lagente}$

John is always ready to please (people)

John is always ready to please people (Rizzi 1986:1)

2.2.0.1 Implicit "sloppy anaphora" and Bridging implicit roles

So-called "sloppy anaphora" implicit roles are instances where the predicate of a current implicit role is not exactly coreferent with a prior predicate, but rather refers to a modified copy of a prior event. This is commonly seen with "sloppy" verb phrase ellipsis situations, as shown in example 44a for Japanese (also noted for Gurani (Tonhauser 2017), and for Bangala, Hindi and Malayam (Simpson **et al.** 2013)):

(44) a. *John-wa* *zibun-no* *tegami-o* *suteta*

John-TOP self-gen letter-ACC discarded

"John$_i$ threw out his$_i$ own letters"

b. *Mary-mo* ∅ *suteta*

Mary-also ∅ discarded

"Mary also threw out John's letters / Mary also threw out her letters (Takahashi 2011)"

Example 45 shows an example from Bangala, in which the implicit role (the thing sent) in the second sentence is a newly introduced referent, Arun's servant:

(45) a. *Abhik nijer chakar-ke Dakghor-e pathalo*

Abhik his servant-ACC post.office-to send-PST.3

Abhik$_i$ sent his$_i$ servant to the post office

b. *Arun-o* ∅ *Dakghor-e pathalo*

Arun-also ∅ post.office-to send-PST.3

Arun$_j$ also sent (his$_j$ servant) to the post office (Bangla; Simpson, Choudhury and Menon 2013:111)

Hoji (1998) notes one extreme example in Japanese, in which a reading has multiple such "sloppy" readings possible:

(46) a. *John-ga John-no gakusei-o suisensita*

John-NOM John-GEN student-ACC recommended

"John recommended John's student"

b. *Mary-wa Bill-mo suisensita to omotteita*

Mary-NOM Bill-ALSO recommended that thought

"Mary thought that Bill also recommender (John's student / Bill's student / Mary's student)"

This roughly matches a class of implicit roles that was proposed in Kay (2004) for elided nominal heads, such as "I had three eggs and Bill had four $\emptyset_{eggs}$" – referred to by Kay as IDENTITY of Sense null instantiations (following Bresnan 1971). We might also group into this type other kinds of "bridging" where the implicit role has a clear antecedent upon which the new referent is based (Clark 1977; Poesio **et al.** 1997).

2.2.1 Genre-based Default Implicit Roles

There are other kinds of arguments which are often not "anaphoric", but express a situationally unique "default" argument, made unique by the particular genre or context. For example, reference to the ball in match reports (as in example 47), or the engineer in repair reports (as in example 48) can be generally left implicit – not because all kicking events entail balls, but because the context means that only one referent can fill that role, and defines what that referent is.

(47) Juice Williams keeps $\emptyset_{the\ ball}$. (Ruppenhofer & Michaelis 2014:58), p58

(48) $\emptyset_{the\ field\ engineer}$ Thinks problem is in the head select area.(Dahl 1986)

Such instances have some similarity to "situationally unique" definite descriptions (Hawkins 1978; Becker 2018), such as "He went **to the hospital**" or "I talked **to the principal**", which similarly fail to point to a universally unique item, but define a referent that is unique within a particular situation. However, lacking the semantic content such as "hospital" or "principal" seen in nominal phrases, these "default" arguments only occur when the genre helps to define the referent as a default.

These genre-based default implicit role interpretations tend to correspond with a number of specific constructions unique to particular genres, well explored in Ruppenhofer & Michaelis (2010). We suggest that these are different constructions than the more general implicit object constructions seen in English. One piece of supporting evidence for the separate licensing of special "genre-based default" constructions is that for some situations in English where a referent is not lexically licensed (such as the object of "vacate"), some examples have been found where implicit object use can be observed if you find a context where that referent is also the clear default referent of the genre (Glass 2014). Example 49 illustrates that for "vacate", wherein one cannot normally omit the thing vacated, but in a discussion forum about owning real estate, the houses being rented are the primary topic under discussion, and we see such implicit arguments:

(49) Divorce raised its (not so) ugly head and they vacated $\emptyset_{the\ house}$ at the end of the lease (Glass 2014)

We can also note the observation from Ruppenhofer & Michaelis (2010), which notes that for both English "labelese" and match report implicit roles – which one might characterize as this kind of GENRE-BASED DEFAULT argument – they "require predicates to denote actions or properties that are canonical in the genre" (Ruppenhofer & Michaelis 2010:169). For example they note that one could not omit the ball in non-match events such as superstitious gestures in example 50, nor can the contents of a box of quinoa be referred to outside of the labelese genre of describing its uses and properties (example 51):

(50) Before he took that free kick, he kissed (*∅/the ball) for luck. (Ruppenhofer **et al.** 2010b:169)

(51) * ∅ has flourished in cultivation for over 5,000 years.

One set of implicit role constructions which might fix into this category of GENRE-BASED DEFAULTS is that of *labelese* and *instructional imperative* implicit role constructions in English. These seem to be licensed by derivational constructions (Ruppenhofer & Michaelis 2010; Ruppenhofer & Michaelis 2014) rather than the more consistent lexicalized behavior noted for other kinds of script-inferrable mentions in English (Fillmore 1986). The *instructional imperative* examples illustrate that otherwise-required objects in English may be omitted within imperatives when they are the discourse topic of an instructional context and in an imperative:

(52) Store ∅ away from direct sunlight. (Ruppenhofer & Michaelis 2014)

(53) Fry $\emptyset$, stirring $\emptyset$ frequently, for five minutes, until the paste is soft, fragrant and reducedin volume. (Ruda 2014:ex. B-35)

In English, we also see implicit subjects for such default arguments, seen either in labelese (examples 54 – 55) or in the omission of default arguments in repair reports in example 56.

(54) $\emptyset$ packaged in a facility that also processes nuts.

(55) $\emptyset_{it}$ Serves four as a side dish (B-81, recipe paper)

(56) $\emptyset_{the\ field\ engineer}$ Thinks problem is in the head select area. (Dahl 1986; Palmer **et al.** 1986)

The instructional imperatives and omissions in recipe contexts have been extensively studied (Ruda 2014; Ruppenhofer & Michaelis 2010; Haegeman 1997; Massam & Roberge 1989). While these particular constructions likely possess unique details and constraints, they illustrate the proposed class of GENRE-BASED DEFAULT roles in terms of being situations where a particular role in the scenario (in these cases, the object being used or created) can be consistently omitted in a range of different contexts.

(57) Keep it on a nice low heat, whisk $\emptyset$ constantly.(www.youtube.com/watch?v=x6Qcurdgu5A) (Ruda 2014)

(58) Check [motor protection filter]$_j$ every time you change the paper filter bag. Replace $\emptyset_j$ by a new one if it is very dirty.(Ruppenhofer **et al.** 2010b:159)

(59) Chill dough, then roll $\emptyset$ to $\frac{1}{4}$"-thick and spread $\emptyset$ with date filling and turn $\emptyset$ over on itself, making a jelly roll. (Ruppenhofer & Michaelis 2010:167)

(60) Slice the mushrooms finely, and put $\emptyset$ in a large bowl with the oil [. . .] (Ruda 2014:341)

(61) Take a crepe$_j$. Cover one half with the jam. Fold $\emptyset_j$ over onto itself and sprinkle with sugar.(Massam & Roberge 1989:137)

The last of these, example 61, illustrates that such implicit roles in these contexts can be "controllers" in a variety of constructions. There are also noted exceptions, in which such labelese constructions do not work, as in example 62–63.

(62) ?Lift [the chicken pieces]$_j$ out of the wine, preserving the mixture in which you have marinated $\emptyset_j$

(63) ?Boil eggs for the salad while you roast $\emptyset$.

2.2.2 Type-identifiable Implicit Roles

There are other kinds of implicit roles in which the referent does not anaphorically refer to a prior mention, and is not necessarily known to the hearer or even the speaker, but where the type of the referent is very predictable, such that it may sometimes be thought of as a "known" referent – David (2016) makes a distinction between *referentiality* and *type-specificity* based upon such referents. The most extreme versions of these are the "subtype" examples in English, wherein the implicit role referent has an even more specific subtype than what is seen for the predicate in general (Mittwoch 1971; Fillmore 1986), as in examples 64 and 65 from Fillmore (1986).

(64) Bill started drinking $\emptyset_{\text{alcoholic beverages}}$ again

(65) Bill is trying to stop smoking $\emptyset_{\text{cigarettes}}$

Lambrecht et al. (2005) note the ability to omit "type-identifiable" referents for some predicates as well. They contrast the examples in 66 with other predicates such as *réparer*(repair) or *brosser*(brush) which have no clear conventionalized default and therefore cannot leave their objects implicit.

(66) *Maman est occupeé; elle could / repasse / lit / peint*

Mom is busy; she sew / iron / read / paint

"Mom is busy, she is sewing / ironing/ reading / painting" (Lambrecht and Lemoine 2005)

Similarly, Ruppenhofer & Michaelis (2014) note a number of FrameNet frames that all have the same indefinite role, which can be left implicit; for example, the "Weapon" role of verbs in the Bearing Arms frame can be left implicit, referring to a nonspecific, unstated gun or guns.

(67) Careful, that guy carries $\emptyset_{a\_gun}$. (Ruppenhofer et al. 2014:8)

(68) Watch out, I think he's packing $\emptyset_{a\_gun}$.

The most extreme forms of such implicit roles can essentially encode a default referent, and may be so type-identifiable as to make it very rare to have an explicit form . This could be seen with denominals in English such as "to butter", "to box", or "to crate", and for verbs in languages with extremely specific verbal forms, such as Tzeltal "heavy" verbs *boj* "to cut (with a machete or knife)", *jatz* "to rip cloth or paper", *kók* "break off or pluck [a fruit from a stem]"(Brown 2008). As noted in Michaelis and Ruppenhoffer (2001) for German, many denominal or incorporation instances do not actually bar the use of an explicit form, such as "butter the bread with apple butter", but merely provide a stereotypical assumption that one might use if no explicit form is used. Nevertheless, it is sensible that many models view many of the most extreme examples of type-identifiable implicit roles as not having a separate "participant" at all, but keeping such a mention as part of the listener's inherent understanding of the event itself.

2.2.3 Generic ("People in General") implicit roles

Lambrecht et al. (2005) discuss a "folks in general" or "everyone" reading for French, in which the person bothered in example 69, and the people invited to stay in example 70, can be construed in a broad sense as "people in general". This kind of role has been referred to as Generic

Null Instantiation (Lyngfelt 2012), or *emploi generique* (Larjavaara 2000), and has been observed for other languages, such as example 71 from Italian (Rizzi 1986).

(69) *ça* *gêne* ∅

that bothers ∅

That bothers (people) (Lambrecht and Lemoine 2005:21)

(70) *Le* *beau* *temps* *invitait* ∅ *á* *rester*

The nice weather invite.imperf ∅ to stay

The nice weather invited (us/them/people in general) to stay

(71) *Gianni* *é* *sempre* *pronto* *ad* *accontentare* $\emptyset_{lagente}$

John is always ready to please (people)

John is always ready to please people (Rizzi 1986:1)

For some languages, this GENERIC construal is the only available construal for an implicit subject as well. This is noted for the Finnish implicit subject construction (Hakulinen & Karttunen 1973; Holmberg **et al.** 2009), as shown in example 72, and Holmberg et al. (2009) noted this across Finnish, Brazilian Portuguese, Finnish and Marathi.

(72) *Shelliasemalla* *voi* *pestä* *autonsa.*

Shell-station-ADE can-3SG wash car-POSS.RFL

"(You) can wash (your) car at the Shell Station (Holmberg 2005)"

The difference between such a GENERIC implicit subject construction and a more common SALIENT/RECENT implicit role construction can be most clearly illustrated by the contrast between the Brazilian Portuguese and European Portuguese interpretations of the same sentence in example 73; in Brazilian Portuguese, the unstated subject of *faz* must be generic, and in European Portuguese it must refer to a recently mentioned referent.

(73) *é assim que faz o doce*

is thus that makes the sweet

"This is how one makes the dessert (Brazilian Portuguese)"

"This is how he makes the dessert (European Portuguese)"

(Rodgrigues 2004:72)

We can refer to these implicit roles, when they refer to broad "people in general" readings (similar to the English *one* or French *on*), as being Generic implicit roles. Han (2006) notes the same issue and refers to them as "generic zero" pronouns. These have also been noted to be used in certain implicit subject constructions in Korean and Russian, as in examples 74:

(74) $\emptyset$ *holangi-lul cap-ulye-myen* $\emptyset$ *san-ey ka-ya-hanta*

$\emptyset_{one}$ tiger-acc catch-intend-if $\emptyset_{one}$ mountain-Des go-must-PresDec

"If one wishes to catch a tiger, (one/he) must go to the mountains" (Han 20006: 56).

(75) *Na galerke* $\emptyset$ *zataili dyhanie.*

On gallery $\emptyset$ held-3pl breath

"In the gallery, (they) held their breath. (Malamud 2004, via Han 2006)

Lambrecht & Lemoine (2005) also discuss "habitual" cases in French, wherein "the predicate is understood as denoting a habitual activity or state of the subject, or the negation of such an activity or state", as in:

(76) *Mon chien ne mord pas*

my dog neg bite neg

"My dog doesn't bite $\emptyset$"

Generic or "People in general" readings also often emerge for the unstated agents of impersonal passive constructions in various languages (Keenan & Dryer 1981); e.g. Nakipoğlu-Demiralp (2001) notes general usage of the Turkish impersonal passive 78, as is also noted in Icelandic impersonal passve (Sigurðsson & Egerland 2009):

(77) *burada iyi kosĉ-ul-ur*

here well jog-pass-AOR-3PER

"It is jogged well here " (Nakipoğlu-Demiralp 2001:136)

(78) *Bu goẍl-de bogûl-un-ur*

this lake-LOC drown-PASS-AOR-3PER

"It is drowned in this lake (People drown in this lake / one might drown in this lake) "(Nakipoğlu-Demiralp 2001:139)

(79) *Fyrst er* ∅ *beygttil haegri*

first is.3sg $\emptyset_{one}$ turned to

"First, one turns to the right." (Sigurðsson & Egerland 2009:160)

Such GENERIC implicit roles also seem to consistently show ambiguities with DEICTIC implicit role interpretations. Lambrecht et al. (2005) note examples where the grammar does not disambiguate between the two, as in examples 80 and 81, where one might view the grammar as encoding something generic, but where the referent could be construed as a specific person (noting that it could even be "a third person whose point of view is being expressed by the sentence"(Lambrecht et al. 2005:35)). Ono & Thompson(2000) noted similar ambiguities in Japanese where there was an ambiguity between a "we" reading and a "one" or generic "you" reading, as in example 82, and noting that it is not clear that a listener is even required to disambiguate between the two readings:

(80) *Attention, ça va faire mal*

watch demonst fut do.inf bad

"Watch out, it's gonna hurt $\emptyset_{you}$!"

(81) *Arrête d' embêter!*

stop of bother.inf

"Stop bothering $\emptyset_{me}$!"

(82) *ana hayaku i ano are shi-na -kya ikenai n desu ka?*

uh quickly uh that do-not if bad NOM COP Q

"is (it) bad if (we) don't do that quickly?" / "Is not doing that quickly bad (Ono et al. 1997:8)"

Put broadly, many languages have ways of referring to events in generalizations wherein the target of generalization is no more specific than "one" or "people", and where the language simply leaves that referent unsaid. In that sense, these would often not be separate "implicit role" constructions, but rather they illustrate interpretations of implicit roles that would be enacted by a generalization construction.

2.3 Clearly Indefinite Arguments – Cataphoric, Iterated/Slot-identifiable, and Arbitrary

2.3.1 Cataphoric implicit roles

Within many typologies of reference marking such as Dyer (2014), Givon (1979) or Becker (2018), there is a category for "cataphoric" (or "establishing") referent types – which Dyer also calls "pragmatically specific indefinite". Some languages have a special treatment for noun phrases which are not known to the addressee, but which refer to referents which are likely to be referred to again in the discourse – as with the cataphoric usages of the English "this".

While most usages of implicit roles in a language such as English are generally either clearly anaphoric or clearly nonspecific, in some languages it seems quite possible to have the kind of clearly referential chains of normal mentions (the kind discussed as "Salient/Recent" definite implicit roles) initiated not by a full noun phrase, but simply by an implicit role – used in this same kind cataphoric, establishing usage. Such "cataphoric" reference was emphasized by Li (2004) for Chinese, claiming that not only may an implicit role occur without prior mention, but that such an implicit role can start a long chain of mentions – citing one story having 18 clauses in a chain before an actual explicit mention of that referent occurs. Stoll & Bikel (2009), in discussing the extreme amount of implicit role reference in seen in Belhare, provide an example of this as well, showing in example 83 how even at the start of a narrative, the agent of **syau phighe** "pick apples" is not even mentioned, and remains unmentioned when a new participant is introduced – the third-person agreement marking alone is sufficient to establish the referent in context:

(83) a.

ʌbo	*pʌila*	*syau*	*phige*
now	first	apple	pick.3s.pst

"Now, first ∅ was picking apples."

b.

syau	*phighe*	*kina*	*dhakie*	*andhe*
apple	pick.3s.pst	and	basket.loc	fill.in.3s>3sPST

"∅ picked apples and filled them into a basket."

c.

ʌni	*ʌni*	*meri*	*sassa*	*tahe,*	*ibaŋ,*
then	then	goat	pull.CVB	come.3spST	one.human

"Then someone came along pulling a goat."

For a language such as English, this kind of cataphoric situation is rare, but can emerge out of arguments of nominal mentions – such as example 84 – wherein events can be asserted without otherwise-necessary roles, and which we might categorize as implying their quick resolution. However, for English at least, this is not characterized by a particular construction type; nominal arguments may simply vary in the perceived level of importance, and may imply to the listener that they would be rapidly resolved:

(84) It was suggested , but never proved , that the deceased gentleman may have had valuables in the house , and that their **abstraction**$_{Theft}$ $\emptyset_{perpetrator}$ was the motive of the crime .

(85) Through the fogged glass I dimly saw a man spring up from a chair beside the fire , and heard a sharp **cry** $\emptyset_{source of sound}$ from within the room .

These all might be judged as introducing a referent, in that mentions of "the thief" or "the source" or "the crier" would all be clear-cut in the larger discourse. These might therefore be "specific known" or "specific unknown" in Haspelmath's map of indefinite pronouns – analogous to "something" in English.

2.3.2 Low-information Arbitrary roles

The opposite end of the spectrum from cataphoric implicit roles are those in which there is no real information provided to the referent, nor any real implication about the referent– even about whether it is a prototypical version of the referent. These therefore do not contain the information one might expect from explicitly indefinite referential forms, such as "someone", in that it is unclear whether a referent is a known, definite referent or something unknown:

(86) The books were delivered (by $\emptyset$) on time. (Lyngfelt 2012)

(87) He burglarized $\emptyset$, but she murdered $\emptyset$! (Goldberg 2001)

(88) Pat gave $\emptyset$ and gave $\emptyset$, but Chris just took $\emptyset$ and took $\emptyset$. (Goldberg 2001:507)

(89) The lion has killed.

These implicit roles tend to be expressed in constructions which could be said to *deprofile* their referent (Goldberg 2001). A range of studies have even questioned the sheer semantic existence of some implicit roles (Koenig & Mauner 1999), or explored the contexts in which these might not be mentally present (Mauner **et al.** 2002; Mauner & Tanenhaus 1995; Mauner & Koenig 2000). For the purpose of this specific categorization into types of referents, one could view this is as a kind of implicit role where there is no information about how to interpret the referent at all.

2.3.3 Iterated Events and Indefinite Slot-identifiable Implicit Roles

In contrast to genuinely arbitrary implicit roles, many indefinite implicit roles are simply non-specific referents – lacking the clear implication of a specific type seen with TYPE-IDENTIFIABLE implicit roles, but nevertheless proposing vague referents which are interpretable in context. This is most notable for the referents of iterated events, as discussed with Goldberg (2001, 2005); such referents often define a heterogeneous set of entities. In these contexts, the referent is not simply unstated, but refers to a "set" of referents only really defined by their role in that sentence. For example, the referent of the patient of "kick" in example 90 is not necessarily a single variable or even a known group of people, but simply the set of people that were kicked. These are therefore not example nonspecific (the speaker, and even the hearer, may know the identity of the referents), nor is it a reference to a natural kind, but the set described has no meaning beyond that event:

(90) Pat kicked his way out of the operating room.(Goldberg 2001)

(91)

Le	*bourgeois*	*ne*	*produit*	*pas:*	*il*	*dirige,*	*administre,*	*répartit,*	*achéte*	*et*	*vend*
The	bourgeois	neg	produce	neg:	3.msc	direct	manage	distribute	buy	and	sell

"The bourgeois does not produce: he directs, manages, distributes, buys and sells"(Sarte; via Lambrecht and Lemoine 2005:23)

(92) She picked up her carving knife and began to chop.

As Goldberg notes, examples such as these generally do not introduce a referent into the discourse, and cannot be easily referred to anaphorically in later context (as shown in example 93). This therefore differs from cataphoric implicit roles in which the unstated argument may be referred to.

(93) The chef chopped and diced all day. *It was put into a large bowl (Goldberg 2005)

2.4 Ambiguities between Interpretation Types

Throughout these specific interpretations, contexts have been noted in which a particular implicit role may receive multiple possible interpretations, which might be disambiguated by the pragmatic realities of the context.

(94) However, it's also easy $\emptyset_{for\_you}$ / $\emptyset_{for\_one}$ to get to Guangzhou by train or ferry(Lyngfelt 2012)

(95) *arrête* *d'embêter* ∅

stop of-annoy.inf ∅

"Stop annoying $\emptyset_{everyone}$ / $\emptyset_{me}$ (Lambrecht and Lemoine 2005:35)"

(96) *hon tog pjašen ur bokhyllan och började läsa*

she took play-DEF from bookshelf-DEF and started read

"She took the play$_i$ from the bookshelf and started reading $\emptyset_i$"

We can note that there are many different ways of handling such ambiguities. Lambrecht defines them as "Free Null Instantiation", and for the distinction between PEOPLE-IN-GENERAL readings and INTERLOCUTOR readings, both the Lambrecht & Lemoine (2005) and Ono & Thompson (2000) suggest that such ambiguity is often left ambiguous by the grammar. Ono et al. (1997)

maintain that even the speakers themselves might not know which referent is exactly intended, and note that "we may want our model of language to reflect the fact that there are many contexts in which the 'referent' is intended to be left 'open'."

We emphasize that regardless of one's theoretical conclusion about such ambiguities, they are not random ambiguities between any implicit roles, but generally are only two specific kinds of edge cases: distinctions where construction may be encoding a GENERIC reading but it can be inferred to be an INTERLOCUTOR, and instances where the construction may be encoded as a TYPE-IDENTIFIABLE reading and where it can be construed in context as SALIENT/RECENT. We generally assume that some kind of pragmatic inference or enrichment allows one to construe such more referential readings from these contexts, but note that there simply is not enough data about how these occur or their linguistic behavior to be more specific than that. Ruppenhofer & Michaelis (2014) drew links from such ambiguities to the kind of "upward entailment" suggested by Gundel **et al.** (1993), where listeners can construe a more anaphoric reading from indefinite reference forms; we will loosely embrace this approach.

2.5 Summary

These implicit role interpretation types are not necessarily an exhaustive nor perfect inventory for expressing implicit role interpretations, but we suggest that this set of relatively simple types is a good starting point for characterizing implicit role constructions, allowing one to separate the syntactic and grammatical effects of a particular construction from the pragmatic details of how these implicit roles are interpreted. Table 2.2 illustrate a set of features which can help characterize these implicit roles, to illustrate which properties are shared between these types.

		Refer to Recent Mention	Known Type/Role	Introduces new Referent	Specific (hearer-known)	Discourse-topical
Definite	Salient/Recent	yes			yes	yes
	Script-inferrable		yes		yes	possible
	Deictic		yes		yes	possible
	Remembered Events				yes	possible
Marginally Definite	Genre defaults		yes		yes	yes
	Generic		yes		yes	
	Identity-of-Sense Anaphora	yes		yes	yes	
	Type-identifiable		yes			
Indefinite	Cataphoric			yes		sometimes
	Repeated/Iterated Roles					
	Arbitrary/Nonspecific					

Table 2.2: An approximate set of features generalizing over these types.

Chapter 3

Using Implicit Role Interpretation Types for Describing Constructions Within and Across Languages

This chapter explores the use of this inventory of implicit role interpretation types for describing the implicit roles constructions within a language. The first such exploration is a case study of the Spoken Arabic implicit role system; it provides an overview of Arabic implicit role phenomena and maps these roles onto the current inventory. The second exploration outlines how to incorporate these types into formal models of implicit role constructions, building on existing work in SBCG (Ruppenhofer & Michaelis 2010; Ruppenhofer & Michaelis 2014). Finally, a third exploration discusses going beyond those interpretation types for the sake of typological comparison, discussing how one might represent the more subtle distinctions that could differentiate particular constructions within the same general type.

These explorations are not intended solely to evaluate the particular inventory of interpretation types proposed here, but rather to use this inventory as a starting point for looking at how we might model different interpretations of implicit roles. The chapter will procede as follows: overviewing the intepretation types proposed in Chapter 2, followed by a review of Spoken Arabic implicit role constructions, followed by a section about formalizing these roles into SBCG, and finishing with a section discussing the representation and comparision of how implicit role constructions vary along continuous scales.

3.1 A review of the proposed categories

To summarize, we present below a set of definitions for each of the implicit role types, with a prototypical example for each.

SALIENT/RECENT: Instances in which an implicit role referent is salient or recently mentioned, and might be realized as a pronoun if explicit.

(97)

Juan$_i$	*llegaba*	*a*	*casa.*	$\emptyset_i$	*Tenía*	*las*	*llaves.*
John	arrive.3s.imperf	to	house.	$\emptyset_i$	have-1/3sg.imperf	the	keys

"John was arriving home. **(He)** had the keys (Spanish; Cole 2010:280)"

FRAME-RECOVERABLE: Implicit roles which are pragmatically identifiable, not because of local coherence or activation, but because of general pragmatic knowledge and inference.

(98) *They accepted.* (Fillmore 1986:99)

REMEMBERED: A semantic role of an event that has already been mentioned in the discourse, and where that prior event predication has already introduced this semantic role; it is *implied* only in the sense of being remembered.

(99) *The company was acquired by Google in 2009. The acquisition* $\emptyset_{of\ the\ company}$ $\emptyset_{by\ Google}$ *cost a total of 500 million.*

IDENTITY-OF-SENSE: The implicit role introduces a *new* referent, which roughly copies a recently mentioned implicit role in the discourse.

(100) a.

John-wa	*zibun-no*	*tegami-o*	*suteta*
John-TOP	self-gen	letter-ACC	discarded

"John$_i$ threw out his$_i$ own letters"

b.

Mary-mo	$\emptyset$	*suteta*
Mary-also	$\emptyset$	discarded

"Mary also threw out John's letters / Mary also threw out her letters (Japanese; (Şener & Takahashi 2010))"

GENRE-BASED DEFAULT: Reference to a (potentially unstated) referent that is uniquely identifiable in that situation or genre

(101) *He smashed* $\emptyset_{the\ ball}$ *into the net when a close call went against him* (Ruppenhofer & Michaelis 2010:p.173)

TYPE-IDENTIFIABLE: An indefinite implicit role that is nonspecific and existentially qaunti-fied, but where the type is discernable, being a prototypical filler of that particular role

(102) *Watch out, I think he's packing* $\emptyset_{a\_gun}$.

PEOPLE IN GENERAL: reference to a generic, "people in general" class which could often be rephrased with "one"

(103)

Shelliasemalla	*voi*	*pestä*	*autonsa.*
Shell-station-ADE	can-3SG	wash	car-POSS.RFL

"(You) can wash (your) car at the Shell Station (Finnish; Holmberg(2005)"

DEICTIC: Deictic implicit role constructions could be locuphoric (first or second person) implicit roles, or could generally refer to anything co-present with the interlocutors.

(104)

$\emptyset_I$	*hon-wo*	*yon-da*
$\emptyset_{\mathrm{I}}$	book-obj	read-PAST

"(I) read a book" (Japanese; Nakaiwa & Shirai 1996)

CATAPHORIC: Mentions of a referent that may be tracked in the discourse, even though not yet explicitly mentioned.

(105) *It was suggested , but never proved , that the deceased gentleman may have had valuables in the house , and that their* ***abstraction***$_{Theft}$ $\emptyset_{perpetrator}$ *was the motive of the crime .*

Iterated/Set: These refer to implicit roles which identify a general set of unspecified referents.

(106) *Tigers only kill ∅ at night.* (Goldberg 2001)

Arbitrary: Implicit roles genuinely underspecified as to the referent

(107) *The books were delivered (by ∅) on time. (Lyngfelt 2012)*

3.2 Arabic Case Study

This section overviews the inventory of implicit role constructions in Spoken Arabic[1] . Although it could be characterized as a "canonical pro-drop" language, we can see that there is a wider range of phenomena to discuss, summarized in Table 3.1:

Construction	Grammatical Slot	Intepretation type
Third-person anaphoric subjects	subject	Salient/Recent
Nominal/adjectival predication	subject	Salient/Recent Deictic
First/second person agreement	subject	Deictic People-in-General
Impersonal passive	subject	People-in-General
Omissible essential arguments of APs	object/oblique	Script-inferrable
Lexicalized Object/Oblique omission	object/oblique	Script-inferrable Type-identifiable

Table 3.1: Overview of Arabic implicit role types

3.2.1 Third-person Anaphoric Subjects

Spoken Arabic generally has an "implicit subject" construction, following the general behavior of a salient/recent implicit role interpretation. Such implicit subjects are primarily constrained to refer to very recently mentioned referents, prototypically referring to the last explicitly mentioned

[1] This study covers a range of colloquial Arabic varieties, primarily using Egyptian Arabic, Tunisian Arabic and Gulf Arabic. While it is likely that dialectal differences in implicit role behavior do exist between these varieties, there is not yet clear-cut information regarding such differences. All examples were converted to coarse-grained IPA by the author from a mix of orthographies and transliterations, so any mispronunciations are those of the author.

subject. Example 108 shows one such example for Tunisian (Gundel **et al.** 2010), which was used to illustrate “in-focus” status within the Givenness Hierarchy (Gundel **et al.** 1993). The prior subject (l-druʒ, “the stairs”), although not a prototypically agentive subject, is the most given and syntactically prominent referent, and thus receives implicit-subject encoding.

(108) a.

l-druʒ	maʔnaha	ma-j-ban-ʃ	l-barra
the-stairs	that.is	NEG-3ms-appear-neg	the-outside

“That is, the stairs don’t show on the outside.”

b.

∅	ma-jaʔmil-l-ikʃ	muʃkla	mʔa	ʒarik	aðaka
∅	neg-3m-do-to-2s-neg	problem	with	neighbor-poss-2s	nonprex2.3ms

“They won’t cause you a problem with your neighbor”(Tunisian Arabic; (Gundel **et al.** 2010) (Gundel **et al.** 2010))

Specifically, there is a question of whether third-person implicit subjects in Arabic reflect a specific subtype of SALIIENT/RECENT interpretation – those of a *same subject* construction. For many dialects, there has been the claim that pronouns tend to encode a salient reference to the last mentioned subject (what could be a “continuation” in Centering Theory (Grosz **et al.** 1994)), and that pronominal subjects strongly imply referent to a recent, salient mention other than the last-mentioned subject. This pattern has been noted across Arabic dialects, first for Egyptian Arabic (Eid 1983), and later for Peninsular Arabic (Emirati, Kuwaiti and Hijazi; (Owens **et al.** 2009)), and Tunisian (Gundel **et al.** 2010). It is also a pattern that Givón has linked to implicit subject constructions in general (Givón 2017), noting the similarity to switch-reference morphological systems.

Example 109 from Eid(1983) illustrates the most canonical example of this “same subject” tendency. While gender agreement allows both referents to be referred to unambiguously, referring to the prior subject (Ali) must use an implicit subject construction, and reference to the prior object (Nadia) requires the use of an explicit pronoun *hiyya* (example 109c).

(109) a. ʕali ʃaːf ŋadia?
Ali see.3ms Nadia

"Did Ali see Nadia?"

b. ah wi ∅ *kallim-ha* *kamaan*
yes and ∅ talk.to-ObjClit.3fs too

"Yes, and he talked to her too"

c. ah wi *hiyya/*∅* *kallim-it-u* *kamaan*
yes and she talk.to-3fs-ObjClit.3ms too

"Yes, and she talked to him too"

However, an examination of a corpus of Egyptian conversational data (Song **et al.** 2014) reveals a number of examples in which one may use an implicit subject even when the antecedent was not the last explicit subject. These may illustrate specific pragmatic edge cases, however. Examples 110 and 111 illustrate one such class of examples, wherein the last subject is an implicit first- or second-person reference and there is a distinct third-person reference in either subject or object position, so that there are no other third-person referents in competition with the referent.

(110) a. *kuntu* ləsha bɪ-ʔa-kləm *maam$_j$-i*
was-prfv just prog-1sg-talk mom-1sg

"I was just talking to mum"

b. ∅$_j$ suʔal-tə-ni ʕləi-k
∅$_j$ question.perf-3f-1sg on-2sg

"She asked me about you"(BOLT SMS - CHT-ARZ-20130504.0001, lines 1104-5)

(111) *wa* *6* *minu:t*ˤ *latr* ʔi-tsˤal ∅ *maa* ∅ ja-rad-ʃ
and 6 minutes later 1sg-call.imperf ∅ neg ∅ 3ms-reply.imperf-neg

"And 6 minutes later, I call (him), and (he) does not answer (BOLT SMS-CHT)"

Similarly, the use of causal connectives can lead to *implicit causality* effects which may reverse the preference for last implicit subject as antecedent. Examples 112 and 112b(Eid 1983)(Najhi 2004:74) illustrate that while coordination with "*wa*" (and) leads to the expected interpretation, coordination with the more causal "fa" (*so/then*) can lead to a different interpretation of the implicit subject, as the need for a plausible cause-and-effect sequence seems to overwhelm this same-subject bias.[2] .

(112) a. ʕali dˁarab *samiir wi* ∅ ʔaam ʃatam-u
Ali hit.3ms Samir and ∅ got_up.3ms insulted.3ms-him

"Ali hit Samir and he (Ali) got up and insulted him" (Eid 1983)

b. ʕali dˁarab *samiir fa* ∅ ʔaam ʃatam-u
Ali hit.3ms Samir and ∅ got_up.3ms insulted.3ms-him

"Ali hit Samir so he (Samir) got up and insulted him" (Najhi 2004)

3.2.1.1 Implicit First and Second Subjects – Deictic and Generic Readings

As expected from the taxonomy described in Chapter 2, first- and second-person implicit subjects do not behave like the anaphoric implicit subject construction discussed above – there are no constraints upon when the "antecedent" has been mentioned, since speaker and addressee are always accessible. Indeed, implicit realization seems to be the default for such first and second-person subjects; Parkinson (1987) finds that 65.4% of first person are implicit, as are 76.6% of second person subjects. As seen with explicit *you* in English, the Arabic second person implicit subject is also very commonly used to express a "generic" or "people in general" reading. This is noted across a range of dialects, as in examples 113 and 114

(113)
bass, ∅ *itmill-iin* arbaʕa wʒ-iʃriin saaʕa gaaʕda *mʤaabl* *kutub*
but, $\emptyset_{one}$ bore-2sg four and-twenty hour sit near book.pl

"but you get bored sitting in front of books for twenty-four hours (Peninsular Arabic; Owens **et al.** 2009; 46)"

(114) ʔin *ta-drus* ta-nʤaɦ
if 2sg-study 2sg-pass

"If you study you will pass (Classical Arabic; Suleiman 259)"

[2] It should be noted that consultations with native speakers on this example finds wide variation in the default interpretation

3.2.1.2 Generic or Indefinite Agents of Impersonal Passive

The other implicit role construction for expressing generic arguments is the "impersonal" usage of the passive morphological inflection (Fehri 2009). As with other impersonal passives, this can be used to demote a subject even of an intransitive verb, causing the predication to the read as a generalization, as in example 115 below. Consultation with native speakers suggests this impersonal reading is largely constrained to intransitive verbs. This therefore results in the implicit agent with a generic referent(people in general).

(115)	∅	y-u-sbaħ-u	*hunaa*	*bi-duuni*	*muqaabil-in*
	$\emptyset_{one}$	3-pass-swim	here	bi-duuni	counterpart-gen

"One swims here without paying (Fehri 2009: 8)"

3.2.1.3 Implicit Object Constructions – Lexicalized Implicit Objects

Arabic objects may also be left implicit in specific contexts and with specific predicates. These objects seem to be lexically licensed in contexts similar to the English lexically licensed implicit objects noted in Fillmore (Fillmore 1986). As in English, these are mostly examples of either type-identifiable or script-inferrable implicit roles. Examples 116 and 117 from the Egyptian text chat data illustrate such type-identifiable and script-inferrable examples, respectively:

(116)	ɦa-na-ruːg	bɪʔaːkl	∅
	fut-1pl-go	with-eat.nmlz	∅

"We 're going to eat (something / a meal)"(BOLT CHT 20130805.0005 line 1624)"

(117)	məʃ	faːhm-a	∅
	neg	understand.AP-fem	∅

"I don't understand (the question)"(BOLT CHT 20130504.0001 line 284)"

Examples that illustrate anaphoric implicit objects, also have similar interpretations, akin to those seen in English, showing no clear requirements for recency of mention, but purely being inferred by one's understanding of the events and scenarios in the discourse.

This similarity with English provides supporting evidence for the idea that such "lexicalized" behavior may be emergent from general understanding of scenarios and frames(similar to the proposal of Ruppenhofer & Michaelis (2014), in a manner independent of individual languages, rather than purely being lexical idiosyncrasies.

3.2.1.4 Implicit "Objects" of Argument Nominalizations

Spoken Arabic also uses active-participle noun phrases, which can function both as nominal arguments and as reports of ongoing activities (Mansouri 2016). These act like verbal predications, and the agent or "subject" may generally be left implicit, following the noted SALIENT/RECENT implicit subject construction. However, these can also act like nominalizations, and as in English, nominals generally lack the valence requirements seen in verbs. In such contexts, otherwise "essential" arguments of predicates can still be implicit, as seen in example 118 below, where the complement of wanting can be left unexpressed, because the predicate ʕaːiza is a nominalization (wanter).

(118) a. anta sˤɑħəb-i*?*
2sg boyfriend-1sgposs

"(Are) you my boyfriend??"

b. aiwa məʃ ʕaːiza ∅ jə-ʕəni*?*
yes neg want.AP ∅ 3sg-mean

"Yes, don't (you) want (that)? (BOLT CHT-ARZ-20121203.0002 lines 595-6)"

The same pattern can be seen in Example 119, wherein the proposition of faːkɪra, *to know / knower* can be omitted. This refers to a recent topic of discussion (the secret under discussion).

(119) a. kuntu ∅ faːkɪra ∅ ʕaːrɪf ∅ witˤaliʕa məʃ ∅ ʕaːrɪf ∅?
be.3sg.pref $\emptyset_{she}$ think.AP-fem $\emptyset_{he}$ know.AP $\emptyset_{it}$ and-follow NEG $\emptyset_{he}$ know.AP $\emptyset_{it}$

"(she) thought (he) was aware (of it), but it turned out that (he) wasn't "(BOLT CHT-ARZ-20121203.0002 lines 595-6)

3.2.1.5 Probing the possibility of Salient Implicit Object Constructions

Hebrew has been observed to have what we would categorize as a SALIENT/RECENT implicit object construction, as in example 120 (Landau 2018). These do not seem to have the same kind of lexically licensed object omission behaviors seen for Arabic, but rather allow implicit realization of objects in SALIENT/RECENT contexts (also noted in French (Lambrecht & Lemoine 2005)).

(120) a. *adayin eyn li manxe la-doktorat*
still no to.me advisor to.the-doktorat

"I still don't have a PhD advisor"

b. *lifney še-ata moce ∅, ata carix nose*
before that-you find ∅, you need topic

"Before you find one, you need a topic"(Landau 2018:16)

We sought to test for the existence of such structures in Arabic. Not only were no examples found in all data examined, but native speakers were prompted with contexts which would be plausible contexts for such an implicit object construction to occur. Example 121 below illustrates one situation where similar implicit object behavior might occur. However, native speakers reported that the anaphoric reference to the money (the *-ha* object clitic) below cannot be left implicit, supporting the idea that Spoken Arabic shows no such construction.

(121) *ana kunt nawi lau* tˤalaba *minii flu:s*
1sg cop.perfv-1sg intend.AP counterfact request.prfctv.3ms from-1sg money
ʔa-diː-***haː*** ləhu
1sg-give-**3fem** to-3ms

"I was intending, if he asked me for money, to give **it** to him"

3.2.2 Conclusions from Arabic Case Study – Comparing Types

This review of Arabic implicit role behaviors follows the reviews done for French (Lambrecht & Lemoine 2005) and English (Fillmore 1986; Ruppenhofer & Michaelis 2010; Ruppenhofer & Michaelis 2014). It differs in that these interpretation types here provide a structure for comparison against other kinds of implicit role constructions in other languages, so that constructions may be more easily compared across different language varieties.

One way to compare constructions in different languages has been through the use of “classical semantic maps” (Croft **et al.** 1987; Anderson 1982; Haspelmath 1997; Van Der Auwera & Plungian 1998). Such semantic maps[3] are organized to express co-expression patterns between different functions or meanings (in this case, the referential interpretation types). For discussing implicit roles, pronouns and other referring expressions with classical semantic maps, each point in the map would represent a referential interpretation – such as the interpretation types used here – and constructions or lexical items defined as regions over those interpretations. This essentially expressed reference forms as being polysemous, where the different senses are taken from a set of shared comparative concepts.

Figure 3.1 presents a simple representation of these using the basic implicit role interpretation types, and depicting the Arabic implicit role constructions. One can see from the figure how Spoken Arabic has a variety of constructions for expressing Generic interpretations. One may also see how nuanced distinctions – such as the simple difference expressed by the use of implicit subjects for “same subject” situations and use of pronouns for “different subject” situations – are not expressed in such a framework.

3.3 Formalization of Language-Specific Implicit Role Contructions

The second exploration of ways to model fine-grained representations of implicit roles is through formal grammatical representations, with a focus upon representing these within Sign-

[3] Semantic Maps are also referred to a construction-schema maps (Traugott 2016) or as a subtype of *conceptual spaces* (Cristofaro 2010; Gärdenfors 2014; Zwarts 2015) in which similarity is defined by similarity of typological co-instantiation.

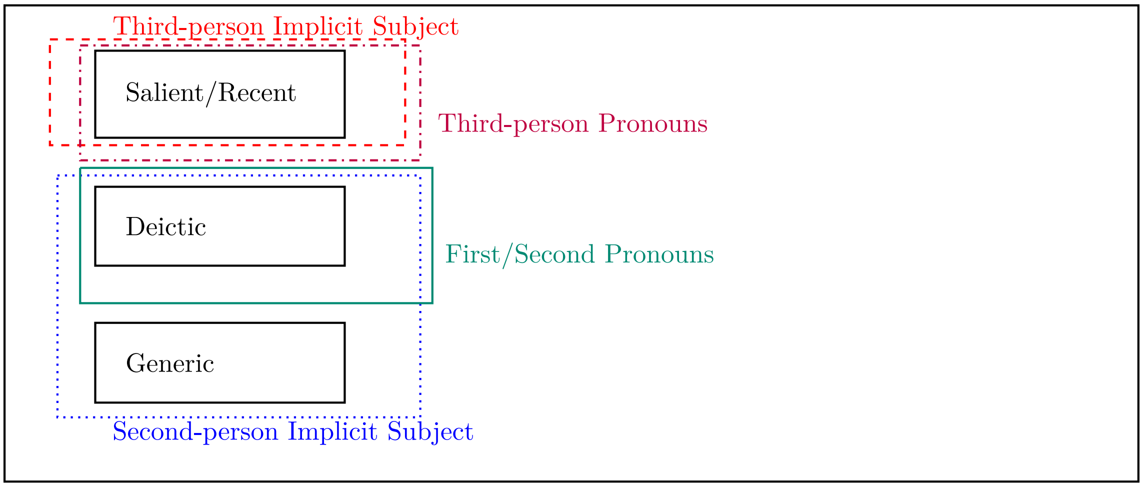

Figure 3.1: Classical Semantic Map of Arabic implicit roles (top), augmented with explicit and English(bottom) referential forms

based Construction Grammar (Boas & Sag 2012; Michaelis 2009). We follow the tradition of construction-grammatical treatments of implicit roles, in which the goal is to present the lexical and grammatical licensors of implicit roles and their interpretations. This section illustrates that these more fine-grained interpretation types can be modeled through a series of analyses using Sign-Based Construction Grammar (Boas & Sag 2012; Michaelis 2009), a formalism developed by the HSPG and Construction Grammar communities. [4]

3.3.1 Essential semantic roles, argument structure and valency

When dealing with implicit instantiation of grammatically core arguments, it is often possible to model those implicit roles as part of a formal representation of a grammar, usually through a representation of which roles are semantically "essential" even if they were left unsaid. One might link this notion of "essential" roles to a range of prior ways of modeling this, from the simple

[4] We omit a specific exploration of how these types map onto predicted sets of behavior in Mainstream Generative Grammar, but it can generally be viewed as having a set of parametric syntactic parameters (parameters for "null subject", for morphological agreement with subjects, or for whether subjects are syntactically necessary) and analytic traditions for how those parameters resolve into classes of languages, such as "canonical pro-drop" languages (those with an implicit subject construction with a SALIENT/RECENT interpretation and morphological agreement), "radical pro-drop" (those with a SALIENT/RECENT implicit object construction and/or no morphological agreement), "partial pro-drop" (those with a GENERIC implicit subject construction) and "expletive NSL" (those which allow omission of subjects due to impersonal passives but which lack a SALIENT/RECENT implicit subject construction).

assumption that objects of transitive verbs are expected, to prior computational ideas for essential roles (such as the essential semantic roles of Palmer **et al.** 1986), or decompositional representations of event meaning which define which arguments are logically necessary (Rappaport Hovav & Levin 1998).

Approaches in the HPSG and SBCG lineage (Pollard & Sag 1994; Sag 2012) follow the conclusions of Manning & Sag (1999) regarding this issue, which handles this issue through the use of two separate lists of "expected arguments": a valence list (VAL) of the syntactically expected arguments, and an argument structure list (ARG-ST), capturing grammatically relevant arguments even when they are not explicitly realized.

This approach allows one to model implicit roles through the reduction of that "valence" list. Doing so allows one to still treat these implicit roles as grammatically relevant, which is important for modeling grammatical phenomena such as binding implicit roles to reflexives (example 122), or implicit roles governing depictive secondary predication, as in example 123 (Haegeman 1997; Ruppenhofer & Michaelis 2010):

(122) Yesterday finally $\emptyset_j$ managed to get **myself** back on the end of a paint brush.

(123) Avoid chewing $\emptyset$ or swallowing $\emptyset_j$ **whole**(Ruppenhofer & Michaelis 2010:172)

We follow the approach of Ruppenhofer and Michaelis (2010;2014) in using this approach to model non-lexicalized implicit roles, but in using a different approach of *optionally covert types* (Kay 2004) for lexicalized implicit roles (discussed below). For both approaches, we suggest that the same set of formal approaches in SBCG may be extended using the new fine-grained referential types proposed herein, simply by adding more pragmatic constraints.

3.3.2 Basic Implicit Role Construction Representations

We utilize SBCG combinatory construction representations, as seen in Figure 3.2. Such typed feature-structure representations have a single parent MTR node, and a list of one or more child

DTR nodes, and can be thought of as descriptions of how the DTR nodes combine to form the MTR node. However, these are not descriptions of a top-down or bottom-up procedural grammar, but rather constraints regarding how different forms can be combined, and only need to describe the feature-structures being constrained.

Figure 3.2 illustrates such an SBCG construction, in a possible model for the English imperative construction. This provides a simple illustration of this approach, wherein lexical rules can introduce "implicit roles" through reduction of the VAL (valence) list. This can be viewed as a derivational construction outlining the mapping between a lexical representation of the verb (at bottom) to a verb representation with valence of a imperative construction (top). The single constituent in DTR represents a normal verb in English, with the VAL and ARG-ST lists both containing a link to the argument that would fill the "subject" position. The construction applies a lexical transformation so that in the resulting MTR node, that argument (referred to with the index *i*) is removed from the VAL list, but remains in the ARG-ST representation – and in this case, that referent index is also added to the CONTEXT|C-INDS feature structure, which is the SBCG way of representing a speaker or addressee. This is essentially just a formal way of removing a verbal expectation that it will see a subject, and linking that subject to the addressee. Similar constructions could be applied for other implicit role constructions – keeping the core approach (reduction of the VAL requirements) but with different pragmatic constraints.

3.3.3 Implicit Role Constructions for Salient and Discourse Contexts

A similar approach has been proposed directly for some HSPG treatments of implicit subjects. Melnik (2007) presented a model of Hebrew implicit subjects for HPSG which represented the implicit subject construction by using the same method of Manning & Sag (1999) of reducing the valence list. However, such an approach without any additional pragmatic constraints fails to capture the underlying meaning of such implicit subject constructions, not even requiring an anaphoric referent at all, much less the SALIENT/RECENT referents we could expect from an implicit Hebrew subject. The SBCG formalism uses CONTEXT features (Green 1996) for capturing such pragmatic

$$\begin{bmatrix} \textit{implicit-subject} & \\ \text{MTR} & \begin{bmatrix} \textsc{Syn} & \begin{bmatrix} \text{VAL} & \left\langle ... \right\rangle \\ \text{ARG-ST} & \left\langle \text{NP} \left[\textsc{inst}\ \text{i}\right], ... \right\rangle \\ \textsc{Mode}\ \textit{imperative} & \end{bmatrix} \\ \textsc{Cntxt} & \begin{bmatrix} \text{C-INDS} & \left\langle \text{ADDR i} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \textsc{Syn} & \begin{bmatrix} \text{VAL} & \left\langle \text{NP} \begin{bmatrix} \textit{overt} \\ \textsc{inst}\ \text{i} \end{bmatrix}, ... \right\rangle \\ \textsc{Arg-st} & \left\langle \text{NP} \left[\textsc{inst}\ \text{i}\right], ... \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 3.2: Simple representation of English imperative

and information-structural content, and prior work exists in capturing such constraints (Song & Bender 2012; Ruppenhofer & Michaelis 2014; Ruppenhofer & Michaelis 2010). Ruppenhofer and Michaelis (2014) model English implicit object constructions in genre-based omission (DEFAULT REFERENT, in the current inventory) using a CONTEXT constraint in which the referent is constrained to be a TOPIC in relation to a particular GENRE node. Song & Bender (2012), for HSPG, also model information-structural phenomena, but do so in more local topic and focus constructions – so that an antecedent is a TOPIC in relation to a particular clause.

We follow these approaches in assuming that the antecedent has an information-structural status defined in relation to a particular clause, but do not label that status as TOPIC. Since such an approach requires that one convert this general complex bundle of features into a series of discrete types, we suggest that one could use the discrete categories proposed by the Givenness Hierarchy (Gundel **et al.** 1993), where the referent with the highest activation would generally be labeled as IN-FOCUS. We therefore use that IN-FOCUS label to model a possible SBCG implicit subject construction in Figure 3.3. This argument must be *in focus* in relation to a particular predicate, and the *ltop* label in SBCG models provides a semantic index for the event itself.

$$
\begin{bmatrix}
\textit{implicit-salient-subject} & \\
\text{MTR} & \begin{bmatrix}
\text{ARG-ST} & \left\langle \text{NP} \begin{bmatrix} \textsc{inst}\ \text{i} \end{bmatrix}, \ldots \right\rangle \\
\textsc{Syn}|\text{VAL} & \left\langle \ldots \right\rangle \\
\textsc{Context} & \begin{bmatrix} \textsc{Info-struct} & \begin{bmatrix} \textsc{in-focus} & \text{i} \\ \text{Clause} & \text{k} \end{bmatrix} \end{bmatrix} \\
\textsc{Sem}|\textsc{frames} & \left\langle \begin{bmatrix} \text{event-fr} & \\ \text{ltop} & k \end{bmatrix} \right\rangle
\end{bmatrix} \\
\text{DTRS} & \left\langle \begin{bmatrix}
\textsc{Arg-st} & \left\langle \text{NP} \begin{bmatrix} \textsc{inst}\ \text{i} \end{bmatrix}, \ldots \right\rangle \\
\textsc{Syn}|\text{VAL} & \left\langle \text{NP} \begin{bmatrix} \textit{overt} \\ \textsc{inst}\ \text{i} \end{bmatrix}, \ldots \right\rangle \\
\textsc{Context} & \begin{bmatrix} \textsc{Info-struct} & \begin{bmatrix} \textsc{in-focus} & \text{i} \\ \text{Clause} & \text{k} \end{bmatrix} \end{bmatrix} \\
\textsc{Sem}|\textsc{frames} & \left\langle \begin{bmatrix} \text{event-fr} & \\ \text{ltop} & k \end{bmatrix} \right\rangle
\end{bmatrix} \right\rangle
\end{bmatrix}
$$

Figure 3.3: Modeling Context and salience

3.3.4 Lexicalized Implicit Roles

Ruppenhofer & Michaelis(2014) proposed that the set of tricks used above – in which a unary construction modifies the valence constraints of a predicate – are not sufficient to capture lexically idiosyncratic implicit role constructions, as seen with the English and Arabic implicit objects (Fillmore 1986). They instead model lexicalized implicit objects using the *optionally covert types* approach of Kay (2004). That approach represents possible implicit roles, not through lexical rules , but through a special kind of "optional" argument that is lexically stored in the VAL list of verbs with lexically optional objects – the (ini) or (dni) instantiation types. Figure 3.4 illustrates such a lexical entry for "eat".

This is formally modeled through the type hierarchy of covert signs; "ini" inherits from *null-comp* signs (which defines it as phonologically null), while adding an existential quantification constraint. This essentially just defines "ini" as something which is phonologically null and which

$$\begin{bmatrix} \textit{eat-trans-lxm} & \\ \text{ARG-ST} & \left\langle \text{NP}_i, \text{NP}_j \right\rangle \\ \text{S}\textsc{yn} & \begin{bmatrix} \text{VAL} & \left\langle \text{NP}_i,\ \text{NP}_j \begin{bmatrix} \textsc{inst}\ \textit{(ini)} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Figure 3.4: Lexical representations with (ini) instantiation type

introduces a new referent into the semantics:

(124) $\text{ini} \Rightarrow \text{null-comp} \ \& \begin{bmatrix} \textsc{sem}|\textsc{frames} \left\langle \begin{bmatrix} \text{exist-fr} \end{bmatrix} \right\rangle \end{bmatrix}$ (Ruppenhofer & Michaelis 2010; Kay 2004)

We can extend this approach further by adding the more fine-grained types of interpretation, such as *type-identifiable* or *script-inferrable*, as subtypes inheriting from these "ini" and "dni" types. One might, therefore, propose a subtype such as TYPE-IDENTIFIABLE, in which a referent is not simply encoded as existentially quantified, but capturing some pragmatic constraints (that it is prototypical or type-identifiable).

(125) $\text{type-identifiable} \Rightarrow \text{ini} \ \& \begin{bmatrix} \textsc{sem}|\textsc{frames} \left\langle \begin{bmatrix} \text{type–identifiable-fr} \end{bmatrix} \right\rangle \end{bmatrix}$

However, we can admit that modeling such pragmatic constraints (especially with issues such as these TYPE-IDENTIFIABLE or SCRIPT-INFERRABLE interpretations) will be hard to model using such general categories (such as whether something is "identifiable" or "expected"), and may require formal work designed to handle quantitative values such as selectional preference or association strength (Guzmán Naranjo 2015).

3.3.5 Pragmatic Enrichment of Generics

The discussion of GENERIC and DEICTIC readings of Chapter 2 noted a common issue in which there is a range of ambiguous situations between GENERIC interpretations and DEICTIC implicit role interpretations. This has commonly been noted as being an enrichment of generic readings – a sentence such as "that is annoying" might be interpreted as "annoying (to me)" or "annoying (to people in general)”, as first noted by Lambrect and Lemoine (2005) for French. Following Lyngfelt (2012), we suggest that this can be modeled as a kind of “pragmatic enrichment” of the explicit GENERIC reading. Such enrichment seems to also go in the other direction, as in the generic interpretations of second-person pronouns or second-person implicit role constructions. Were one to attempt to model that within the grammar, one might represent such enrichment as a simple unary lexical rule, as proposed in Figure 3.5, wherein a sign with a generic semantic frame (assumed to be *people-in-general-fr*) could be coerced into referring to a speaker or addressee.

$$\begin{bmatrix} \textit{pragmatic-enrichment} \\ \text{MTR} \begin{bmatrix}\text{\textsc{Context}} & \begin{bmatrix}\text{C-INDS} \begin{bmatrix}\text{\textsc{Speaker}} & \text{i}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{DTRS} \left\langle \begin{bmatrix}\text{\textsc{Sem}}|\text{\textsc{frames}} & \left\langle \begin{bmatrix}\text{people-in-general-fr} \\ \text{ltop} \quad i\end{bmatrix}\right\rangle\end{bmatrix}\right\rangle \end{bmatrix}$$

Figure 3.5: Pragmatic

3.4 Comparison of Constructions

The problem with such formal models of implicit role constructions is that reference phenomena in general, and implicit roles in particular, hinge upon features such as importance, predictability, expectation, topicality, coherence or cognitive activation. Even if these features were easy to measure, they are continuous dimensions, and most representations of formal grammatical models (such as HSPG or SBCG) do not have frameworks to handle such values.

The representation of implicit roles is therefore closely related to larger questions regarding how to characterize syntactic constructions using latent, continuous representations. The remainder of this chapter discusses the advantages of attempting to model such continuous scales (such as “accessibility”), illustrates the issues with squeezing these into one or two dimensions, and will then discuss one possible alternative to those problems, that of *exemplar-based semantic maps.*

3.4.1 Implicit Role Interpretation Types Reflect Different Scales

The most well-studied continuous dimension of referential behavior is the scale of "referential reduction" applicable to SALIENT implicit role constructions (such as implicit subjects), generally framed as a gradient scale with implicit, reduced forms at one end, and full nominal forms (such as indefinite noun phrases or proper names) at the other. However, as noted in Chapter 3, this is not a single scale, but a complex bundle of highly correlated features. Different scholars have linked this scale to different underlying pragmatic features or psycholinguistic dimensions, ascribing roughly the same scale to accessibility (Ariel 1988; Ariel 2004), topic continuity (Givón 1983), assumed familiarity (Prince 1981), or givenness[5] (Gundel **et al.** 1993). Figure 3.6 illustrates the high-accessibility end of the Accessibility scale (as formulated for Hebrew; Ariel 2013), which is illustrative of such a referential scale. For those approaches that view this as a continuous scale, that means that the pragmatic acceptability of a particular construction such as "Arabic implicit subject" would be defined essentially as a threshold on that scale – so that Arabic would have some parameter α for the implicit subject construction, and antecedents with accessibility beyond that α could be referred to using an implicit subject, and those less than α could only use pronominal mentions.

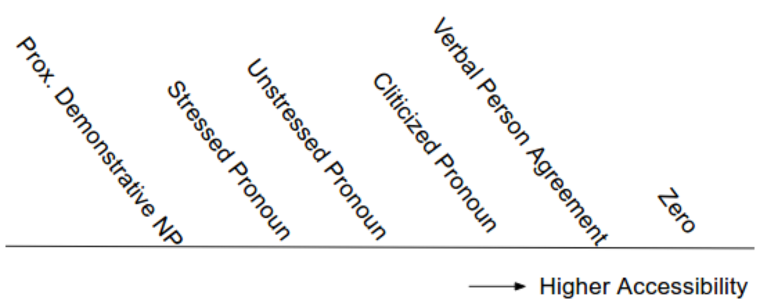

Figure 3.6: A depiction of (one end of) the accessibility hierarchy as implemented for Hebrew (Ariel 2013)

This notion of a threshold upon accessibility is explored in the dataset of Cole (2010), which

[5] This characterization of scales as "continuous" does not extend to the Givenness Hierarchy of Gundel **et al.** (1993), which proposes an ordered list of discrete categories, each of which is "upward entailing" in allowing all forms of reference that are more explicit than required by a particular point of givenness

also illustrates the complexities of such an approach. Cole (2010) built a small elicitation corpus of sentence pairs testing different givenness situations, translated those pairs into a set of six languages, and collected acceptability judgments from speakers of each language. This data illustrated that across different languages with "implicit subject" constructions, one could find clearly different criteria for how "salient" or "given" those implicit subject antecedents needed to be. For example, both Italian and Greek speakers would consistently leave implicit the subject in prototypical examples (referring to the last mentioned subject), shown in examples 126a and 126b, but would differ when the antecedent was the demoted agent of a passive (as in examples 127a and 127b), with Greek speakers still using implicit subjects, but with Italian speakers tending towards pronominal subjects.

(126) a. *Stis 3 Iouliu, o proedros ipegrapse to simvoleo. Tin epomeni merea, aftos/∅ ipegrapse ena kenurgio simovoleo*
on 3 July, the president sign-3sg.pst the contract. The next day he/∅ sign-3s.pst a new contract

"On the 3rd July the president signed the contract. The next day, ∅/he signed a new contract. (Greek; Cole 2010:305)"

b. *Ogni mattine Gianni visita la mostra. Nel pomeriggio ∅ visita la universitá*
each morning Gianni visited the museum. In-the afternoon ∅/he visit-3s.pres the university

"Every morning, Gianni visited the museum. In the afternoon he visited the university. (Italian; Cole 2010:305)"

(127) a. *Stis 3 Iouliu, to simvoleo ipograftike apo ton proedro. Tin epomeni mera, aftos/∅ ipegrapse ena kenurgio simovoleo*
on 3 July, the contract was-signed by the president. The next day he/∅ sign-3s.pst a new contract

"On the 3rd July the contract was signed by the president. The next day, he signed a new contract. (Greek; Cole 2010:309)"

b. *Ogni mattine la mostra é visitata da Gianni. Nel pomeriggio lui/*∅ visita la universitá*
each morning the museum is visited by Gianni during-the afternoon he visit-3s.pres the university

"Every morning, the exhibition is visited by Gianni. In the afternoon he visits the university (Italian; Cole 2010:308)"

Although Greek speakers tend to express this as implicit and Italian speakers as pronominal, this is not an all-or-nothing measure, but one of average preferences; 79% of Greek speakers preferred an implicit role for example 127a, but only 22% of Italian speakers preferred an implicit role for the Italian equivalent (example 127b). In other words, we would want something with a softer and more probabilistic version of a "threshold" than simply modeling a single hard cut-off point.

Figure 3.7 illustrates a modified version of the implicit role data collected from Cole (2010), converting that survey data to a 1-5 scale (with 1 meaning that an implicit role is required, and 5 showing that a pronoun is required) so that average preference scores can be represented. This depicts different preferences over the borderline cases in the data, starting with example E, the demoted passive agent examples listed above as examples 127a and 127b above, and going to even less accessible examples where the antecedent is very oblique, such as F (*Mayumi*$_1$*'s father*$_2$ *is a wonderful person. She*$_1$ *loves him*$_2$) or after dislocation constructions such as example H (*Juan*$_1$*s mother*$_2$*, he*$_1$ *hates her*$_2$). This representation lets us characterize each kind of example with a single score for its average acceptability, and allows one to more quantitatively depict how the implicit subject behavior in each language changes as the accessibility of the antecedent decreases. For example, one can see that Greek keeps implicit subject usage very high regardless of situation; that Japanese generally disprefers implicit subjects across all of these borderline examples, and that one can see the steep decline in Serbian implicit subjects across the same examples. In other words, it lets us more clearly move from the oversimplistic notion that a given language "have an implicit subject construction" to a specific notion of how a given implicit subject construction might compare against others.

3.4.1.1 Other scales of reference

While that referential scale associated with SALIENT/RECENT implicit roles is the most well-studied, we could postulate a range of other such scales that might model the licensing of other implicit role interpretation types.

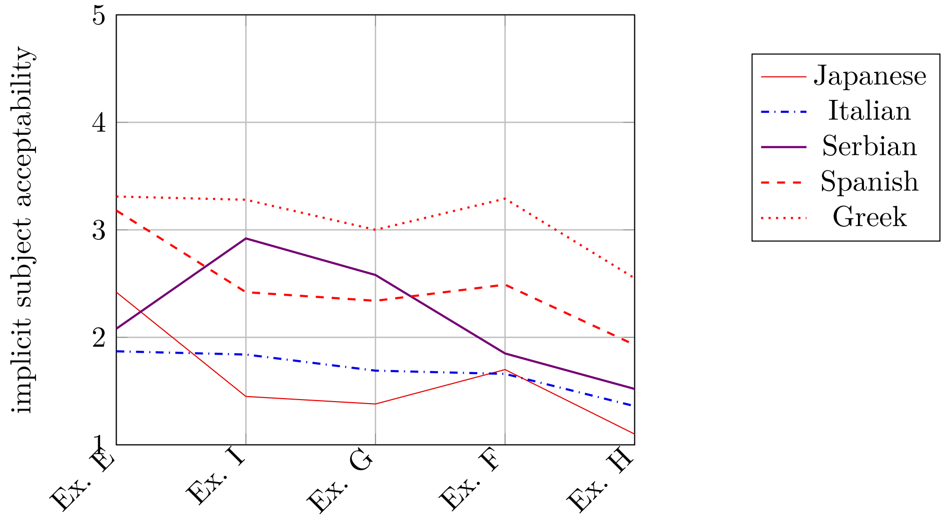

Figure 3.7: Scale from 5 (implicit role required) to 1 (pronoun required) for the edge cases of Cole (2010)

One scale that is likely to show clean cross-linguistic differences in the manner seen above would be a general scale associated with DEICTIC references. Like the SALIENT/RECENT implicit roles above, one could postulate a gradient scale where one end would be a instances in which almost every language leaves out the deictic implicit role – i.e. a backgrounded first-person or second-person mention, where it's very obvious that the speaker or addressee would be the referent (as in the experiencer of mental verbs). At the other end of such a scale would be instances where the implicit deictic referent would be either foregrounded, topicalized, or be an unlikely referent of that particular role. This would, naturally, find languages like English as usually encoding the locuphoric expressions explicitly (except in limited diary-style examples (Haegeman 1997; Ruppenhofer & Michaelis 2010)), and find dramatic amounts of omission for languages such as Arabic.

One example of such a dimension would be the information-gain model of licensing proposed by Resnik (1993, 1996) for implicit object constructions. While this is often connected to selectional preference, the "scale" would be an amount of information gain that would be provided by the use of an explicit mention – e.g. the extent to which "I was reading a book" contributes information beyond what is expressed by "I was reading ∅".

Both of these illustrate the advantages and disadvantages of proposing such "scales" for each interpretation type. Having such a scale could allow us to more accurately model how and whether languages vary regarding a particular kind of implicit role. Moreover, such an "accessibility" scale has also been discussed in terms of formal SBCG representations of constructions in Guzmán Naranjo (2015), and one could postulate actually defining formal constructions which could be defined in relation to these scales, or learning models which could predict such scales from observed behavior (as explored further in Chapter 5).

3.4.1.2 Limitations of Modeling Implicit Role Thresholds

However, as discussed above, a single dimension of variation is often a collection of conflated features, rather than giving us insight into a single dimension of variation. For example, while one might postulate that the rough scale that we can see in the examples from Cole (2010) maps onto some deep underlying feature like "accessibility", one might gain more insight by learning such a scale, and then learning how it would related to a range of underlying features such as discourse coherence, topicality, or subject prominence. Indeed, for many of these interpretation types, we can postulate that they map onto a range of different features, rather than a single objective score; Table 3.2 illustrates the range of features which might be involved in the licensing of each implicit role type.

3.4.2 Exemplar-based Semantic Maps

One alternative to these pre-defined features is that of *semantic map* approaches, in particular *exemplar-based* semantic maps (also called "proximity maps"). While "classical" semantic maps – as with the representation of the Arabic referring expression constructions in Figure 3.1 – provide one way of illustrating how constructions or lexical items interact with meanings, they cannot portray more gradient notions about when a particular construction is more appropriate or less appropriate. The examples noted above from Cole (2010) regarding implicit subjects illustrate a different approach: rather than defining a construction in terms of its compatibility with functions

SALIENT/RECENT	Coherence/discourse structure (Walker **et al.** 1994; Ueno & Kehler 2010; Ueno & Kehler 2016)
	Agreement (Jaeggli & Safir 1989; Taraldsen 1978)
	Last focus in dialogue (Grosz **et al.** 1983; Dahl 1986; Rao **et al.** 2015)
	Referential competition (Skarabela & Allen 2010)
	Implicit causality bias (Ueno & Kehler 2016)
	Physical co-presence (Allen **et al.** 2008)
	Joint attention (Skarabela & Allen 2010; Allen **et al.** 2008)
	Discourse topicality (Huang 1984; Kwon & Sturt 2013)
	Accessibility (Ariel 1988; Ariel 2004)
	Giveness of referent (Skarabela & Allen 2010; Gundel **et al.** 1993)
	Grammatical role of antecedent (Carminati 2002; Arnold 1998)
TYPE-IDENTIFIABLE	Selectional pref. and predictability of type (Resnik 1993)
SCRIPT-INFERRABLE	Schema-based likelihood (Chambers & Jurafsky 2009)
Information-structural factors linked to both	Deprofiling (Goldberg 2001; Goldberg 2006)
	Aspect/telicity (Olsen & Resnik 1997; Rappaport Hovav & Levin 1998)
	Intent to emphasize event (Rice 1988)
GENRE-BASED DEFAULT	Default argument within genre (Ruppenhofer & Michaelis 2010)
GENERIC	(obviousness of non-episodic predicate status)
	predicate non-episodic (Govindarajan **et al.** 2019)
ARBITRARY/DEPROFILED	Deprofiling (Goldberg 2001; Goldberg 2006)
CATAPHORIC	
DEICTIC	Physical co-presence (Allen **et al.** 2008)
	Frame-semantic tendency to be deictic

Table 3.2: Factors proposed to be connected to implicit roles which may apply to specific interpretation types

or senses (such as the "interpretation types" used here), one can define a construction in terms of how appropriate it is in individual contexts. The Cole data presented in the line chart of Figure 3.7 illustrated a very simplistic form of this, defining terms by their compatibility with a small set of hand-crafted elicitation pairs, but one would not need to be limited to such a small set of hand-crafted examples.

"Proximity" or "exemplar" semantic maps take this form of scalar representation to its natural conclusion, by means of a scatter plot in which each point corresponds to an exemplar, often a sentence with the phenomena in question which was translated into many languages. The ability of the semantic map to express cross-linguistic generalizations comes from how those points are arranged: two points in such a map should be close together if they tend to be expressed using the same linguistic form or construction, and should be far apart if languages tend to distinguish between them – roughly that "Meanings are closer if speakers use the same lexical item for them more often across languages." (Zwarts 2010).[6] While this has often used a set of elicitation sentences to learn the underlying similarity space (Hartmann **et al.** 2014), one could use as many points as desired, or might even use all the instances within a parallel corpus, as long as one could get alignments between the phenomena of interest (Wälchli & Cysouw 2012; Cysouw 2007; Wälchli 2010).

One example of such representations can be seen in Figure 3.8 from Wälchli (2010), which illustrates two different views of the same "semantic map" representing spatial adpositions. Each point is an instance (in a corpus of sentences with multi-lingual translations) in which a spatial relationship is asserted. The clusters define situations where most languages all use the same adposition or case marker to express that spatial semantic role. However, the colors and legend within each view of this map illustrate how a particular language realizes these contexts. In this case, the map on the left shows how these points are expressed in French, and the map on the right illustrates how those same contexts are realized in Tok Pisin. Thus one can see that where French roughly makes a distinction between the clusters that roughly correspond to SOURCE(using *de*) and COMPANION (using *avec*), Tok Pisin expresses meanings using the same grammatical encoding. It also illustrates the appeal of such approaches as *exploratory* approaches to the data – so that instead of defining *a priori* what the types or categories are that we see in the data, the clusters can be emergent out of a bottom-up exploration of the data, with labels such as "companion" applied after the fact.

[6] This similarity is often calculated with Torgerson Multi-dimensional Scaling (Torgerson 1952), which maximizes the rank correlation between the Euclidean distance between any two points, and their cosine similarity in a larger multidimensional matrix – usually a sparse matrix where each column is a possible realization in a possible language, such as "English proximal demonstrative". However, the MDS representation is only one way of capturing the more general idea of representing linguistic meanings through their translation equivalents in various languages.

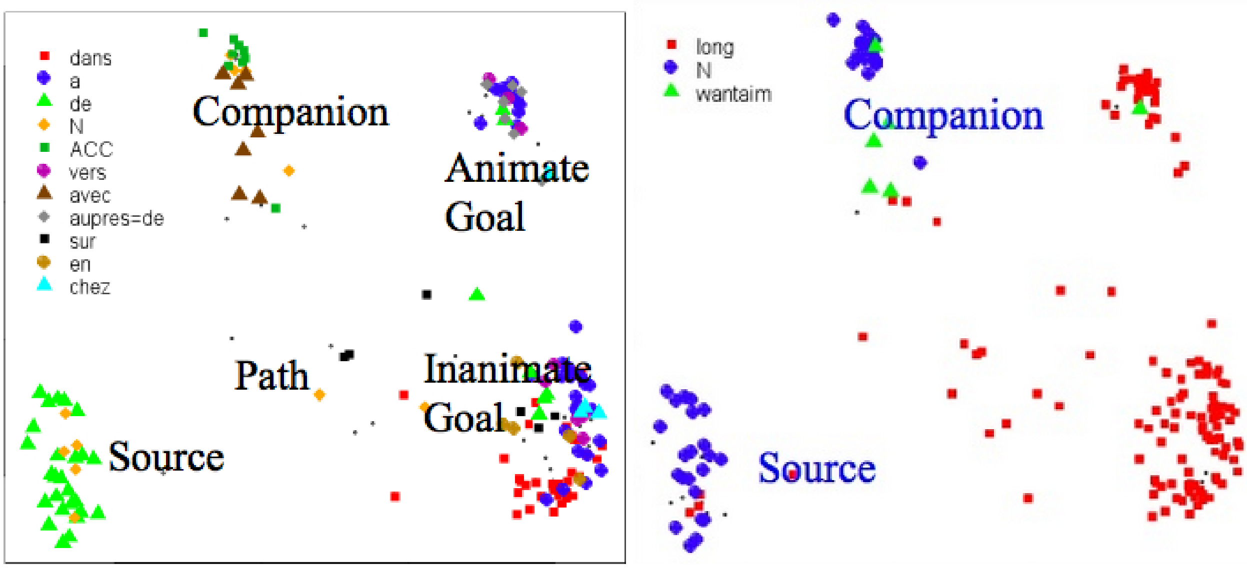

Figure 3.8: Maps of spatial adpositions over the same examples from French and Tok Pisin, from Wälchli (2010)

We explored the possibility of such an exploratory approach to implicit role resolution, both as a more bottom-up alternative to the interpretation types proposed herein, and as a way of looking into more nuanced, non-discrete issues that determine the acceptability of different implicit role constructions. Doing so would, by necessity, mean defining implicit role resolution situations in contrast with *explicit* instantiation – so that a particular SALIENT/RECENT implicit role is defined, not be a label, but by how it cross-linguistically alternates with pronouns, and a TYPE-IDENTIFIABLE implicit role instance might be studied in terms of how it alternates with indefinite noun phrases and indefinite pronouns such as *something*.

In order to establish a simple baseline system for this, we built a small dataset of examples from parallel corpora. The primary resource used was aligned data from the LORELEI language packs (Strassel & Tracey 2016), augmented with some data from biblical translation (Christodouloupoulos & Steedman 2015), using ten languages in LORELEI. A set of instances of semantic roles – where a particular referent is being referred to – were selected, and then for each language, that semantic role was annotated for how the semantic role was being expressed in that language – e.g. pronominally, with a demonstrative, with an implicit role and agreement marking, or implicitly

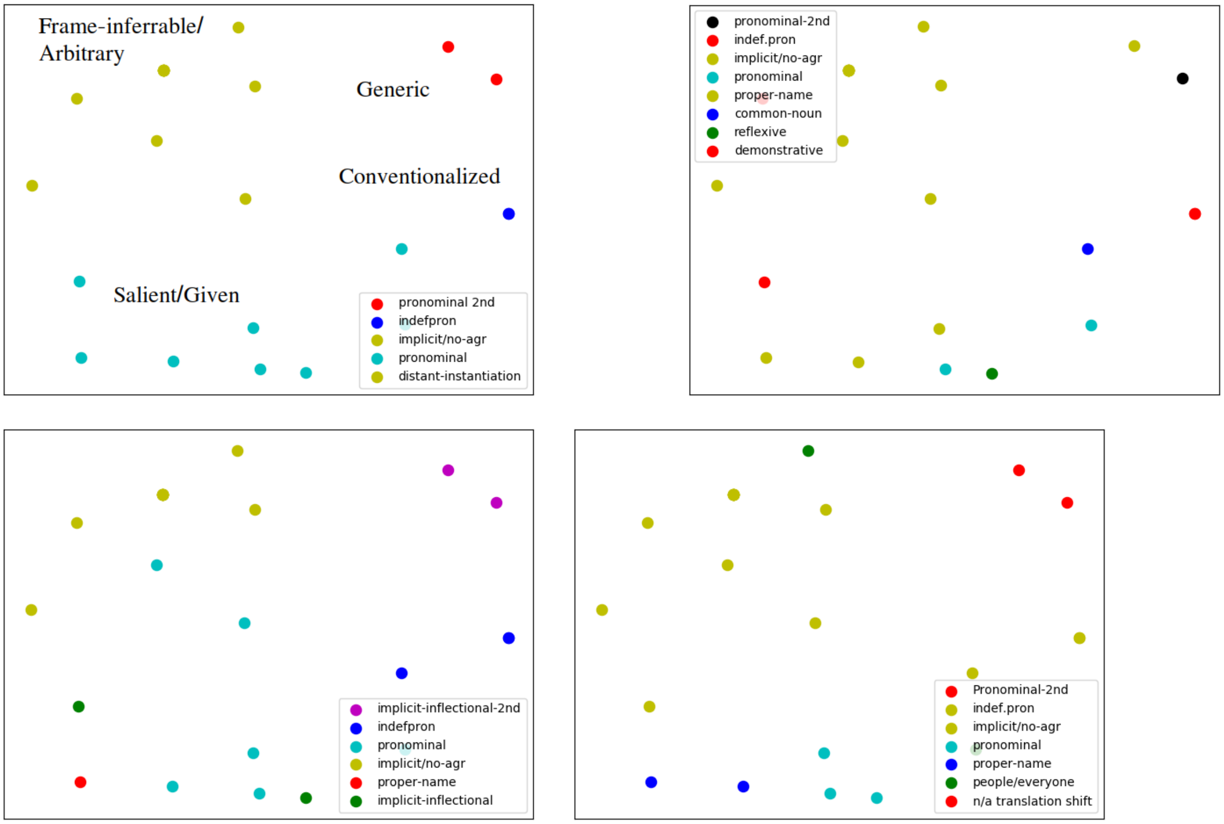

Figure 3.9: Basic MDS analysis of English (top left), Arabic (top right), Italian (bottom left) and Chinese (bottom right) implicit roles

without marking – using translations, dictionaries and grammars. This set of aligned realizations was converted into a two-dimensional semantic map via multi-dimensional scaling, presented in Figure 3.9. A large cluster at the bottom left of SALIENT/RECENT implicit roles is expressed as pronouns in English (turquoise dots in the English map) and as a mixture of implicit and pronominal forms in Arabic. One can also see a region of generic and deictics in the top left (as generalizations in the data were commonly expressed with second person), and a third rough cloud of instances in the top left (labeled "frame-inferrable/ arbitrary"), which encompass both nonspecific, arbitrary implicit roles, and SCRIPT-INFERRABLE implicit roles, as instances in both cases were almost never in alternation with any other referential forms.

It is hard to view a particular semantic map color-coded with the realizations of a particular language as actual hypotheses about the use of that language – one must simply spot patterns in the map. One can resolve that by also learning models which characterize each actual referential construction within a given language (by predicting which referential form will be used, based upon this reduced point). Hartmann **et al.** (2014) introduced an approach of adding constructional boundaries to such maps using such a function, via geostatistical Kriging, we use predictions from the similar Gaussian Process Classification (Neal 1998) .

We illustrate such an attempt to directly mark the boundaries of particular constructions, by looking only at a set of instances annotated as SALIENT/RECENT from that larger dataset; these are presented in Figure 3.10 below. One can see that points in this data form a number of rough clusters (which are the same across all languages), and then that the colored regions define the region of the use of a particular construction or referential form in a particular language. While English obviously illustrates a simple distinction between pronominal reference (the green region encompassing most instances) and a small set of SALIENT/RECENT coordinated constructions, in Arabic and Spanish one can see the distinctions between pronoun use (the leftmost instances) and implicit subject constructions with agreement marking (the colored regions on the center right in both language).

For these, the instances on the right (labeled "high accessibility") correspond to prototypically given, salient main-clause subjects. Another cluster on the left shows a mixture of less accessible instances, such as non-volitional and inanimate subjects (as in example 128) or with semantic roles of embedded predicates – often in infinitival or nominal form – as in example 129.

(128) Residents say they received food and other assistance after the cyclone but that **it**/$\emptyset$ stopped coming months ago and they desperately need more.

(129) ...where he stayed for seven years until he was permitted to $\emptyset_{he}$ return under the terms of the Oslo Accords

The region at the top of the map, in constrast, mostly consisted of constructions in which the implicit role was not a subject at all, and were instances where the SALIENT/RECENT label was somewhat marginal while still referring to a very recently mentioned referent, as in 130 – 131:

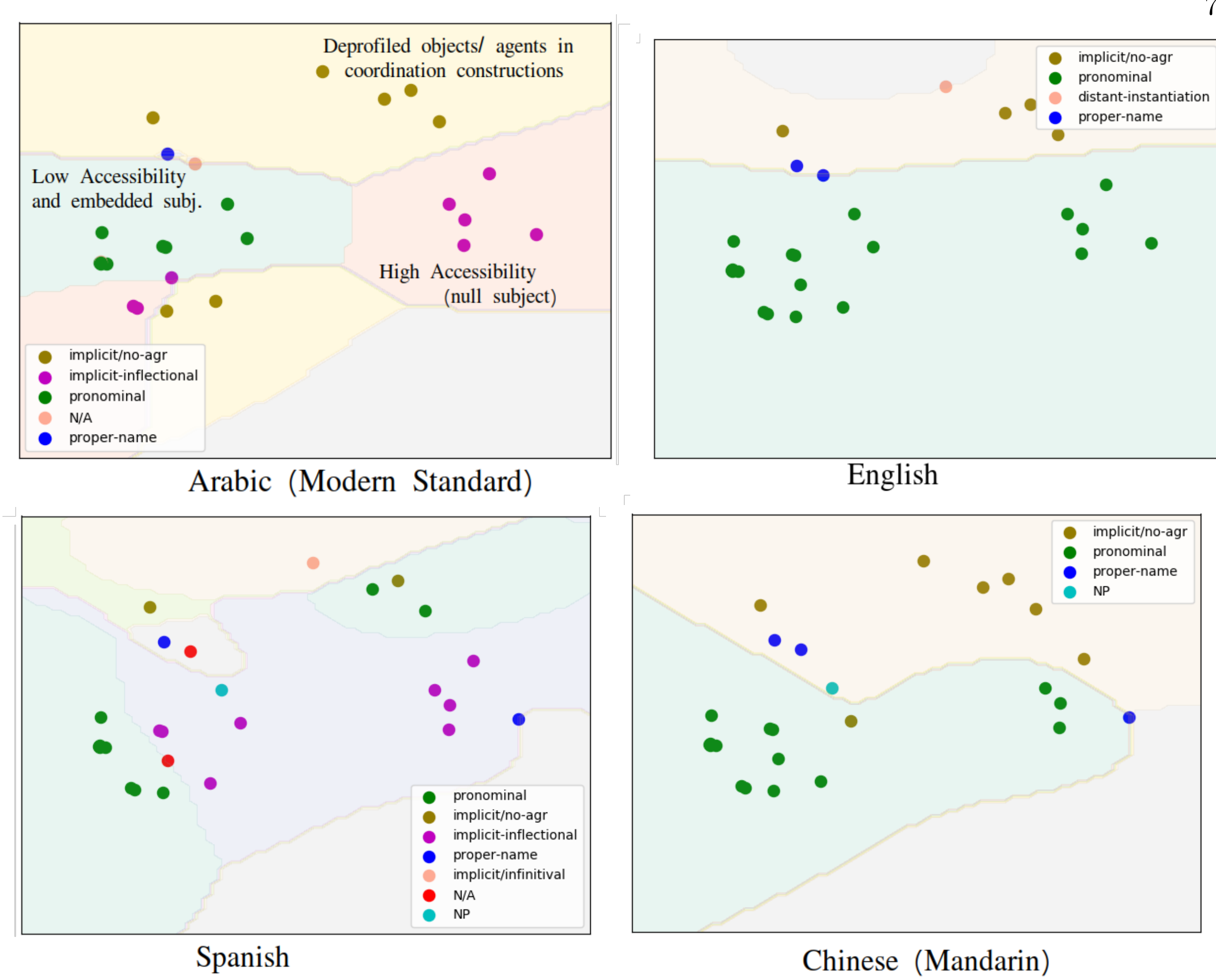

Figure 3.10: Arabic (top left), English (top right), Spanish (bottom left) and Chinese (bottom right) implicit role constructions for SALIENT/RECENT contexts

(130) ... they set bread before him, and he did eat $\emptyset_{thebread}$ (New Testament data)

(131) he was arrested by Israel and deported $\emptyset_{byIsrael}$ to Jordan (LORELEI data; (Strassel & Tracey 2016)

These allow us to clearly see the kinds of underlying implicit role phenomena in the data, and to have different subtypes of these phenomena emerge without *a priori* assumptions about how they should be divided up. But they also illustrate the difficulties inherent to any endeavor to learn rich underlying representations of implicit role interpretation by looking at patterns of cross-linguistic variation. One might note that the most clear-cut patterns noted above do not illustrate underlying kinds of variation we might expect – such as topicality or coherence – but rather capture more grammar-internal questions, such as whether a given representation is being represented as a main or embedded clause, or whether it gets agreement morphology in that language. Therefore, while we still expect that such exploratory approaches may be a useful tool for visualizing how licensing interacts with subtle pragmatic factors unique to each construction, it may be hard to do so without controlling for all of the other parts of grammar that reference interacts with.

Chapter 4

Implicit Role Corpora

This chapter pivots towards implicit role corpora – going beyond language-level description of constructions and rules, and focusing on datasets of annotated implicit role phenomena. Such datasets are important for building computational models of implicit role interpretation, as explored in the next chapter. However, as discussed in prior chapters, implicit roles are not a homogenous phenomenon, as there are many reasons why an argument might be left implicit. Moreover, there are many decisions to be made in designing and constructing a corpus of annotated implicit roles, and these decisions can different corpora to annotate different kinds of implicit roles, even if dealing with the same language.

This chapter will overview the landscape of these corpora. It begins by introducing the landscape of implicit role corpora across all languages, and attempting to overview the fundamental differences in how the different corpora are annotated. Then it overviews the major corpora annotated for English outside of the current thesis. This is followed by a discussion of four implicit role resources developed connected with this thesis, which dramatically increase the data available for implicit role resolution systems. Finally, over this landscape of annotations, we then present some analysis of these interpretation types, and attempt to illustrate the true differences in implicit role behavior that we see in different implicit role corpora.

4.1 Implicit Role Corpora Across Languages

Table 4.1 outlines most of the major corpora annotated with implicit roles. There is a wide variety of corpora which have been annotated with non-local implicit roles, which vary widely both in their structure and in size – with Chinese, Japanese and Korean (Han 2004) having large corpora, but with English and other languages being less well-resourced. Certain annotation decisions are made in the development of each corpus, and those decisions shape what kinds of implicit roles are actually annotated. Even within the same language and the same domain, there will be a huge difference between a corpus which uses a very narrow definition of implicit role (such as only looking at anaphoric implicit subject constructions) and a corpus using a very wide definition of implicit roles (e.g. capturing all semantically inferrable roles, even of nominal predicates).

The most dramatic kind of difference between corpora is in the definition of roles – between those which focus only upon syntactically missing grammatical roles, and annotations which look at any semantic role which would be viewed as inferrable. Some annotate either only implicit subjects, or implicit subjects and some implicit objects, either whenever an annotator labels the role as being syntactically missing (such as the Chinese or Arabic implicit role corpora (Pradhan **et al.** 2011; Weischedel **et al.** 2011)), which rely upon syntactic empty category annotations), or when defined by agreement morphology (as in the Hungarian implicit role corpus, which labels implicit subjects, objects and possessors, as they are all indexed by agreement morphology (Vincze **et al.** 2018)). The other extreme is to annotate all inferrable implicit roles, including nominal predicates (Peris **et al.** 2013; Gerber & Chai 2010). This is connected to another set of constraints – whether one is limited to only annotating the arguments of verbs, or also annotating semantic roles of nominal and adjectival predicates.

The second set of major differences in how corpora are built and defined in regard to the nature of how recoverable referents are labeled: whether they are labeled at all, and whether non-recoverable implicit roles are annotated. Non-recoverable roles (both "indefinite" implicit roles and definite implicit roles without an anaphoric referent) are annotated in a few corpora, such as Czech

PCEDT corpus (Hajičová & Ceplová 2000), which captured nonspecific and generic mentions; the NAIST Japanese corpus (Iida **et al.** 2007a), which captures exophoric, generic and nonspecific; and corpora in Spanish (Peris & Taulé 2012; Peris **et al.** 2012; Rello & Ilisei 2009), Portuguese (Pereira 2009), Romainan (Mihaila **et al.** 2010) and Italian (Rodrıguez **et al.** 2010), as well as a few smaller English corpora (Moor **et al.** 2013a; Baker **et al.** 1998; Ruppenhofer **et al.** 2010a), all of which captured some indefinite null instantiations.

Finally, corpora may also differ in the procedure used for their annotation. Some corpora are annotated in during the process of coreference annotation – adding empty categories which represent particular implicit roles into a text, and then during the process of clustering coreferent mentions (such as labeling the antecedents of pronouns) the annotators also cluster empty categories. Other corpora are annotated by prompting annotators with a particular instance of an implicit role, and asking annotators to label all spans which might fill that implicit role. Unlike the other differences, these differences are not guaranteed to entail differences in the nature of a corpus, but can lead to different biases regarding which kinds of implicit roles are identified.

4.2 Detailed Discussion of English Implicit Role Corpora

As English has few implicit role constructions (such as null subjects or pronoun-like implicit objects), there have been no annotations which focus upon such syntactically absent implicit roles. However, this means that English implicit role corpora provide the challenging frontier of annotating more pragmatically entailed implicit roles – such as the arguments of nominal and adjectival predicates, and core oblique arguments. This means that annotation of Engish implicit roles is challenging to annotate, but also means that they provide an excellent proving ground for whether one can computationally handle complex pragmatic phenomena.

4.2.1 FrameNet and Semeval-2010/10

The first major datasets for implicit role resolution in English come from the FrameNet lineage of work on implicit roles. FrameNet annotation of implicit roles formalizes the “null in-

Language/Corpus	instances (+non-recov.)	role definition	referent	annot. method
Chinese (Weischedel et al. 2011)	18,431	syn. roles of verbs	anaphoric	Coref.
Japanese (NAIST; Iida et al. 2007)	18371 (41974)	case roles of verbs+nouns	any	Coref.
Korean (Han 2004)	12522 (12633)	syn. roles of verbs	anaphoric	Coref.
Czech (Hajičová and Ceplova 2000)	syn. roles	any	any	Coref.
English (Banarescu et al. 2012)	3302 (converted)	sem. roles	same-sentence	Coref.
English (O'Gorman et al. 2018)	2824	sem. roles	anaphoric	Coref.
English (Gerber and Chai 2010)	1172	sem. roles of nouns	anaphoric	Prompt
English (Semeval-2010-10)	245 (580)	sem. roles	all	Prompt
English (This chapter)	(860)	sem. roles	not resolved	n/a
English (Baker et al. 1998)	sem. roles	any sem.	not resolved	n/a
Spanish (Rello and Ilisei 2009)	1202	syn. roles of verbs	anaphoric	Prompt
Spanish (Perin et al. 2012)	469	sem. roles of nouns	anaphoric	Prompt
Romanian (Mihaila et al. 2010)	997	syn. roles of verbs	anaphoric	Coref.
Hungarian (Vincze et al. 2018)	1243	syn. roles of verbs	anaphoric	Coref.
Arabic (Weischedel et al. 2011)	2633	syn. roles of verbs	anaphoric	Coref.
Italian (Rodriguez et al. 2010)	1885	syn. roles of verbs	anaphoric	Prompt
Portuguese (Pereira et al. 2009)	1076 (1477)	syn. roles of verbs	any	prompt

Table 4.1: Different Implicit Role Corpora

stantiations" of Fillmore (1986) into three types – lexicalized definite null instantiations (DNI), lexicalized indefinite null instantiations (INI), and constructional null instantiations (CNI) such as demoted agents of passives. The FrameNet corpus annotated these DNI, CNI, and INI types upon examples and full-text annotations across the corpus (Baker **et al.** 1998), but did not resolve the actual referents of those implicit roles, when they can be resolved in the text. This made it a resource for detection of DNI and INI instances, but not viable for learning systems that would do implicit role resolution. We can see some examples below:

(132) Seventh : **Provide** labor force . (frame = Supply, supplier=DNI, recipient=DNI)

(133) The comment could be interpreted as a **warning** that Russia would oppose, for example, an attack on Iraq. (frame=Warning, Addressee=INI, Speaker=INI)

For the Semeval 2010 shared task #10 (hereafter simply "SemEval-2010-10") (Ruppenhofer **et al.** 2010a), this FrameNet approach to implicit roles was augmented with annotations which labeled the identity of the referents, so that one might actually build and evaluate systems which would find these implicit role referents. Two "Sherlock Holmes" texts by Arthur Conan Doyle – a short story used as training data ("the Tigre of San Pedro"(Doyle 1917)) and two chapters from "The Hound of the Baskervilles"(Doyle 1884) – were annotated with this additional information. These were annotated first with the original DNI, INI and CNI FrameNet labels, and then instances of DNI frame elements were linked to their referents in the text when possible. This resulted in a quite small set of examples – for recoverable mentions, being only 250 training examples and 250 test examples. One can see the sizes of the FrameNet and SemEval data in Table 4.2; while there is ample data annotated for the "DNI" and "INI" types, the actual referent annotations provided by SemEval-2010-10 are quite limited.

	INI	DNI	recoverable DNI
FrameNet 1.7	19716	18499	N/A
SemEval-2010 train	277	303	245
SemEval-2010 test	361	349	259

Table 4.2: FrameNet-labeled Data

These SemEval-2010-10 annotations exhibit some peculiarities of the fiction genre. For example, we see large amounts of narrated conversation, and those direct-speech mentions have conversational phenomena, such as implicit roles which refer to a recently mentioned proposition, as one might see in conversational English:

(134) "You 're right$_{proposition}$, Mr . Holmes ."

(135) " Oh , very good$_{theme}$" said Holmes .

(136) "But WE know$_{about \quad the \quad crime}$ "

Another common and unique characteristic of this data are implicit arguments that might be said to rely on "speaker tracking"; implicit addressees and sources of communication verbs, so that one needs to remember who is talking to whom:

(137) " Henderson , " the inspector **answered** $\emptyset_{addressee}$, " is Don Murillo , once call the Tiger of San Pedro . "

(138) "But how come you into this matter , *[Miss Burnet]*$_j$?" **asked** $\emptyset_{addressee}$ Holmes .

4.2.2 "Beyond NomBank" and ONV5 Corpora

A separate approach to implicit role annotation looks at small sets of hand-selected predicate types, but annotated meaningful numbers of instances of each predicate, so that there was sufficient training data for each predicate (at the expense of only covering a small number of different implicit roles). This approaches was introduced in an annotation over nominal predicates (Gerber & Chai 2010) using the NomBank inventory (Meyers **et al.** 2004) and followed by a similar approach over verbal predicates (Moor **et al.** 2013a) using the FrameNet inventory. The nominal data (called the "Beyond NomBank" corpus or BNB), as it is larger than the wide-ranging SemEval-2010-10 data, has therefore become the *de facto* evaluation standard for more recent implicit semantic role labeling models.

The "Beyond NomBank" corpus (Gerber & Chai 2010; Gerber & Chai 2012a) selected ten predicates with the same goal of "high implicit role frequency", ranking nominal predicates by both their frequency and by the difference between the average number of explicit arguments of each noun vs its verbal equivalent. A separate corpus Moor **et al.** – sometimes referred to as "ONV5", as it was annotated over the Ontonotes 5.0 verbs – was annotated using a similar approach, but using verbal implicit roles from the FrameNet inventory, also attempting to maximize the recoverability of the implicit roles. Moor **et al.** (2013a) utilized a study of FrameNet data as a starting point, focusing upon detecting which arguments were most likely to have a recoverable argument, and settling upon five frames (Departing, Bringing, Commerce_Pay, Giving and Placing) which should have high recoverability rates, annotating one low-polysemy predicate from each frame. Feizabadi & Padó (2014) also attempted targeted annotation of a small set of predicates.

As one might expect, these corpora which are optimized to have high rates of recoverable implicit roles show higher rates of recoverability than full-text annotations, as will be discussed below. The size of these corpora can be seen in Table 4.3; one can see that they are all somewhat limited in size.

	predicates	non-recoverable	recoverable
Gerber and Chai (2010) – train	10	0	1172
Moor et al. (2013)	5	388	242
Feizabadi and Pado (2014)	10	0	250

Table 4.3: Predicate-specific annotations

One defining characteristic of the BeyondNomBank data in particular is that a large percentage of implicit roles are what we are calling REMEMBERED-EVENT implicit roles: instances where the actual semantic role has been uttered previously, and one needs only to resolve the event coreference link to the prior verbal instance. We can see an example of this (intervening text in ellipses) below.

(139) ... it will ask a U.S. bankruptcy court to allow it to hire Lazard Freres & Co. to help it *sell* **its leasing unit**$_j$ ....

The real estate and mortgage banking concern had hoped to use proceeds from the *sale* (theme = $\emptyset_j$) to reduce its debt .

Another characteristic of the Beyond NomBank data is a range of implicit role instances in which the process of prompting annotators for implicit roles either encouraged annotation of non-eventive instances, or led to annotation of local arguments which were simply not annotated in the original NomBank data. This often led to marginal interpretations of the arguments of generic mentions – such as finding the payer of “costs” in “health-care costs” (as in example 140) or the commodity being sold in “sales positions” (as in example 141).

(140) In the 12 months ended in September , wages and salaries of [**private-sector**]$_j$ workers rose 4.4 % , while health insurance costs spurted by 13.7 The consumer price index climbed 4.3 % in the same period . Despite the big increases in health-care costs $\emptyset_{cost\ \ to\ \ whom\ \ =\ \ j}$, wages still account for a far greater share of overall labor costs .

(141) The acquisition combined the country 's second-largest [**security**]$_j$ company , Pinkerton 's , with 1987 sales of $ 410 million ... eliminating about 31 % of the company 's 2,500-person administrative staff , including more than 100 sales $\emptyset_{thing\ \ sold\ \ =\ \ j}$ positions .

This also leads to issues such as the agent of “investors” as in example 142. The term “investors” is a predicate in the NomBank annotations (mapped to invest-01 in PropBank), but this leads to an annotation issue in which NomBank annotators did not annotate the arg0 (as it was incorporated in the predicate), and thus it was a “unstated” argument for the annotators to resolve. While these examples are supposed to be omitted from evaluation – as per instructions upon the website for the data, `http://lair.cse.msu.edu/projects/implicit_annotations.html` – it is often unclear whether all systems removed them from the training and evaluation data:

(142) Big Board officials have been under siege from both **investors** and the exchange 's own floor traders since the Dow Jones Industrial Average 's 190-point tumble on Oct. 13 .

These illustrate a number of challenges which remain a hazard for all annotation of implicit role resolution corpora, and of nominal predicates in particular.

4.2.3 Limitations of the Current Landscape of Implicit Role Corpora

The resources outlined above constitute the data available before the current thesis for models which hope to learn how to detect implicit roles in English. However, not only are they limited in size, they are particularly limited in terms of "full text" annotations, in which all implicit roles in a document are annotated. Only the approximately 250 instances of implicit role annotations available in the SemEval-2010-10 data currently behave as such full-text data, as most of the other datasets only provide resources for a limited set of predicates. This paucity of data makes it extremely difficult to build any systems oriented around this task, and difficult for the current annotations to act as anything more than a test set.

4.3 New Implicit Role Corpora

To resolve these ongoing issues regarding implicit role data, we outline four new (or adapted) corpora. The first two are adaptations of existing corpora – extracting implicit role annotations from the Abstract Meaning Representation corpus (Banarescu **et al.** 2013), and converting the NomBank, ONV5 and FrameNet data into the modern PropBank inventory. The third new corpus is the Multi-sentence AMR corpus, which enabled annotation of implicit roles at scale on top of existing AMR data, and which provides a new wide-coverage resource for implicit role resolution and detection. The final corpus is a set of annotations which provide the Chapter 2 interpretation types annotated upon all unstated arguments, providing some insight into the landscape of different implicit role types. These annotations dramatically increase the size of the annotated data available for implicit role resolution, particularly for the purpose of full-text annotation.

4.3.1 Abstract Meaning Representation Within-Sentence Links

One potential source of implicit role annotation data is the Abstract Meaning Representation(AMR) corpus (Banarescu **et al.** 2013). AMR annotation is a formalism for representing the meaning of a sentence in a simple readable graph, designed to be understandable enough for human annotation. Figure 4.1 shows a very simple version of such an AMR for the sentence "The dog whined at the owner for a walk"; AMR captures and sense-disambiguates all predicates, links co-referent pronouns to their antecedents within the sentence, and provides semantic roles between all of those predicates, even when the referent is only pragmatically inferrable, as with implicit roles. Each of the numbered arguments (such as "arg0" or "arg1") are uniquely defined by their predicate (as in the "individual thematic roles" notion mentioned in Dowty (Dowty 1991) – so that in this case, the arg0 of "whine-01" is "whiner" and the arg3 is "subject matter of the whining".

This example illustrates two possible "implicit role" instances – the annotation disambiguates the possessee of "the owner", and identifies both the agent and patient of "a walk". But as one can see from that example, these are not specifically labeled as implicit – they are simpled encoded as semantic roles. Moreover, the AMR annotation does not provide a layer of derivational syntax; there is no representation of how the string is converted into this graph.

```
(w / whine-01
  :ARG0 (d / dog)
  :ARG2 (p / person
    :ARG0-OF (o / own-01
      :ARG1 d))
  :ARG3 (w / walk-01
    :ARG0 p
    :ARG1 d))
```

Figure 4.1: Simple AMR for "The dog whined at the owner for a walk"

These two facts – that a large amount of implicit role information is annotated within AMR representations, and that those implicit roles were not marked as such – motivated the creation of a derived resource in which those implicit role were identified and then converted to a non-AMR format that might be used alongside other implicit role corpora. This required linking all semantic roles in AMR back into the original text, and then using other resources to differentiate between explicit and implicit semantic roles.

Converting AMR concepts back into predicates and spans in the text was done using automatic statistical alignments (Pourdamghani **et al.** 2014) over the upcoming AMR 2019 public release. For quality control, predicates were double-checked against lexical representations of all their possible aliases. This resulted in the set of all semantic roles (including implicit roles). Then explicit semantic roles were detected using two methods; a hand-crafted set of argument identification rules designed for this task over dependency trees, and the predictions of a full SRL system SRL system (AllenNLP version of Lee **et al.** (2017)), trained upon the PropBank Unified corpus (O'Gorman **et al.** 2018a) so that nominal and adjectival predicates (and informal text similar to the AMR genre) would be represented in the training data. Arguments which were not categorized as being a local, explicit mention according to either metric were categorized as implicit.

The output of this provides a relatively high-precision collection of implicit roles, and one where the "errors" would still be valid semantic role arguments, but simply explicit roles. Examples 143 – 144 illustrate examples of the generated within-sentence roles:

(143) Analysts worry that Indonesia 's military is now so degraded it can no longer control the borders of [the far-flung archipelago]$_k$, allowing for easy **infiltration** $\emptyset_{place \quad infiltrated=k}$ by extremists .

(144) It took 6 months of twice weekly physiotherapy (very painful) as well as bi - weekly **visits** $\emptyset_{agent=k}$ to a chiropractor and exercises at home twice daily before my$_k$ arm was back to " near " normal .

To evaluate the quality of this output, we manually assessed 57 examples of implicit roles produced by this process. Over that sample, we find that 57% of the converted AMRs are valid implicit role instances, with 20% actual errors (either locally instantiated mentions or alignment errors), with another 23% being either correct implicit roles with partially correct spans, or being complex, long-distance coreference which might still be viewed as implicit. This results in a large corpus of somewhat viable implicit roles, totally 3302 recoverable implicit role examples in total.

4.3.2 Conversion Methodology for NomBank and FrameNet data

As noted above, the BNB data (Gerber & Chai 2010), ONV5 (Moor **et al.** 2013a) data, and Semeval 2010-10 FrameNet (Ruppenhofer **et al.** 2010a) data all utilize different inventories of semantic roles, and are not compatible with the PropBank inventory used in the new corpora presented below. In order to treat these as a compatible set of annotations, one must therefore either provide discrete mappings from one resource to the other, or must utilize more complicated models (such as multi-task learning) to fit them into the same conceptual space, as done for SRL in Fitzgerald et al. (2015). We focus upon the simpler task of building direct mappings, as it can result in clean datasets which might be used without additional work. Prior work in using domain adaptation between SemEval and NomBank has taken this approach (Feizabadi & Padó 2015), using the original FrameNet-PropBank mappings provided within the SemEval task (those of Palmer 2009).

There is prior work (Ruppenhofer **et al.** 2010a) in mapping these datasets to each other. Ruppenhofer et al.(2010a) utilized SemLink (Palmer 2009) in order to establish mappings from some FrameNet frames and arguments into PropBank annotations. However, there exist a number of predicates within FrameNet which are not treated as predicates within PropBank. There are also a number of arguments in FrameNet corpora which would be modeled as adjuncts in the PropBank representations, so that there is some loss of arguments. To exemplify what is omitted from FrameNet, we randomly sampled 10 implicit roles where the predicates have no clean PropBank conversion, illustrated in Table 4.4.

waiting at the station all the **week**$_{calendric\_unit}$
I don 't want you to commit yourself **too**$_{Sufficiency}$ far unless you are sure .
the rise of **to-morrow**$_{Calendric\_unit}$'s sun
the singular and formidable inhabitants , the unknown **dangers**$_{Risky\_situation}$ of the approach
what is the **next**$_{Relative\_time}$ step ?
But another will come , and yet **another**$_{Increment}$, until some day ...
The **dictator**$_{Leadership}$, his two children , his secretary , and his wealth had all escaped them
and the **avenger**$_{Revenge}$ might find him .
we had discovered in the transformed Henderson the fallen **despot**$_{Leadership}$,
all three fired with the same reasons for **revenge**$_{Revenge}$

Table 4.4: FrameNet-labeled Data

A second task is the conversion of FrameNet implicit roles into a PropBank inventory is the conversion of the FrameNet "Frame Elements" (characterizations of semantic roles in the FrameNet inventory) into the PropBank arguments, which is viable because FrameNet tends to be more fine-grained than PropBank. For example, FrameNet distinguishes often between agents and non-agentive causes, or distinguishes between plural reciprocal arguments (as in "the two brothers fought") and simple plurals ("the two brothers fought the oil company"). Such arguments being conflated in PropBank, it can be problematic to convert from PropBank arguments to FrameNet arguments without manual intervention. However, there is usually a single correct PropBank argument for each FrameNet Frame Element, allowing that one-directional mapping to be easily provided. However, some subset of implicit roles are omitted due to mismatches, primarily in cases wherein the Frame Elements do not correspond to any core numbered arguments in the equivalent PropBank sense. For example, adjectives such as "kind" or "horrible" in FrameNet have a possible "Judge" role (kind/horrible according to whom), but the corresponding adjectival rolesets in PropBank have no such role. Table 4.5 shows the total statistics for this conversion, along with that of NomBank and ONV5.

	original recoverable	pred. mismatch	role mismatch	converted	total candidates
SemEval-train	245	24	72	**149**	728
SemEval-test	259	84	26	**149**	863
NomBank-train+val	718	0	0	718	1969
NomBank-test	310	0	0	310	1045

Table 4.5: FrameNet-labeled Data with conversion to PropBank labels

These new converted corpora are also converted into a consistent representation format in which the implicit roles are represented in stand-off annotation, used in the same format as the other two corpora discussed here, so that one might use the three converted corpora alongside the MS-AMR and within-sentence AMR datasets presented here.

4.3.3 The Multi-sentence AMR Corpus

As noted above, AMR representations capture implicit roles and other coreference relations while capturing the general propositional content of a sentence (Banarescu **et al.** 2013), but do not represent any implicit roles across sentence boundaries. The Multi-sentence AMR (MS-AMR) corpus (O'Gorman **et al.** 2018b) is an annotation on top of AMR, which adds coreference and implicit role annotations between individual AMR sentence representations, to provide coreference over an entire document. This is therefore a useful new resource for implicit role annotations, and constitutes a dramatic increase of "full text" implicit role annotations (annotating all implicit roles in a document), which previously were only annotated in the SemEval-2010-10 annotations discussed above.

4.3.3.1 Annotation Methodology

The MS-AMR annotation used gold, human-annotated AMR graphs (as described in section 4.3.1 above) as a starting point for this coreference annotation. On top of those annotations (using the upcoming AMR public release), annotators added a separate layer of coreference relations and implicit role relations. Figure 4.2 illustrates a basic example of coreference in MS-AMR, where the variables highlighted in red would be linked together into one "coreference cluster" (a list of mentions referring to the same referent or event), and those highlighted in blue would be linked together into another such cluster.

Implicit roles were annotated by first augmenting graphs with candidate implicit roles from the PropBank lexicon, and then linking to those candidates during coreference annotation. Because within-sentence AMR annotates all predicates using the PropBank sense lexicon (Palmer **et al.** 2005), we may consult that lexicon for the list of all arguments which we might expect for that predicate, and present predicate-specific definitions for each role. For example, in the example in Figure 4.2, the predicate "arrive-01" has three arguments that were not filled (ARG2: *extend*, ARG3: *start point* and ARG4: *destination*), and these were therefore added as "'implicit-role" nodes for the sake of coreference. This allows annotators to naturally consider these during the task of annotation, rather than individually prompting annotators for each such implicit role.

```
Bill left for Paris
  (l / leave-11
    :ARG0 (p / person :wiki - :name (n / name :op1 “Bill”))
    :ARG1 (i / implicit-role :val “starting point”)
    :ARG2 (c / city :wiki “Paris” :name (n / name :op1 “Paris”)
He arrived at noon
  (a / arrive-01
    :ARG1 (h / he)
    :ARG2 (i / implicit-role :val “extent”)
    :ARG3 (i2 / implicit-role :val “start point”)
    :ARG4 (i3 / implicit-role :val “destination”)
    :TIME (d / date-entity :dayperiod (n3 / noon)))
```

Figure 4.2: Example of MS-AMR annotation; annotators link coreferent variables (such as marking a relation between between p and h (in red)) and implicit roles, here linking the destination (arg4) in the second sentence to the previous variable c (in blue)

In the context of different annotation decisions discussed at the start of this chapter, this can be viewed as a combination of approaches. This approach to annotation can be viewed as a combination of the coreference-focused annotations (which traditionally propose syntactic empty categories and link to them during coreference) and the more semantically defined implicit role annotations such as Ruppenhofer **et al.** (2010a) or Gerber & Chai (2010), which utilized semantic criteria for what roles would be implicit, but individually prompted annotators with each semantic role, so that they would re-read and re-annotate a document for each instance of each implicit role. This MS-AMR methodology uses predicate-specific semantic role definitions, but still clusters them during a coreference pass.

An example of the actual annotation interface can be seen in Figure 4.3, using the Anafora toolkit (Chen & Styler 2013). The annotation guidelines used for this annotation are publicly available at `https://github.com/timjogorman/Multisentence-AMR-guidelines/`. By annotating on top of static AMRs, this annotation tool can allow annotators to add an implicit role to a coreference chain with a single click, allowing for relatively rapid annotations.

In order to make this data compatible with those other annotations, we therefore provided

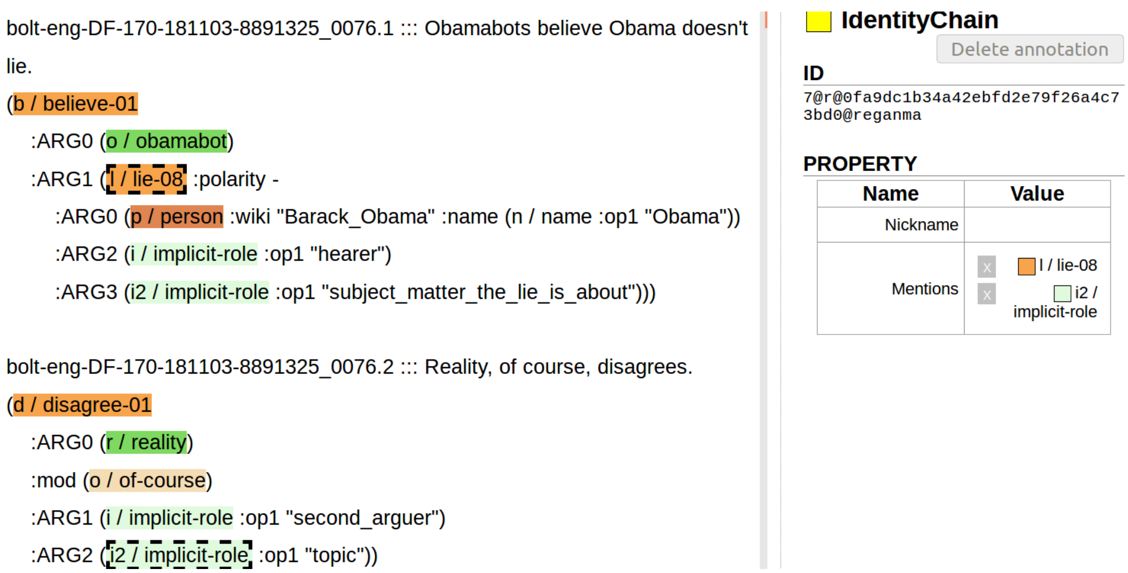

Figure 4.3: Annotation interface, illustrating implicit role links. Annotators click on boxes within the AMR (left) to add them to coreference chains (full chains shown on the right), as with the link between the implicit topic ("i2") and the earlier "l / lie" mention.

conversion of this dataset into a span-based annotation framework, following the same format as the conversion of other datasets discussed above. To do so, we utilized alignments to the string (Pourdamghani **et al.** 2014) to map predicates and referents into the text, and utilized automatic parsing (Gardner **et al.** 2018) in order to convert the referent heads to full syntactically-defined spans. This produces a set of 2,142 implicit roles for training, and a small set of test documents, within the AMR test split. It also provides a large set of 29095 semantic roles which were presented as options to annotators, but which were not added to coreference chains, providing negative examples (this high rate of non-recoverable implicit roles will be discussed further below).

	original recoverable	pred. mismatch	role mismatch	converted	total candidates
MSAMR-train	2446	**2142**	29095		
MSAMR-test	67	**64**	281		

Table 4.6: MS-AMR conversion

A set of 43 documents within this annotation were also double-annotated, in order to measure inter-annotator agreement and check for persistent errors (although they were both annotated over the same gold AMR). The current annotations, following the same evaluation assumptions, found a Cohen's kappa of κ=0.59, which is a testament to the difficulty of the task. The most comparable annotations are those of Gerber and Chai (2010), which found a slightly higher but comparable Cohen's kappa (Cohen 1960) of κ=0.64. As found in Gerber and Chai (2010), the bulk of this type of error is in the task of discerning whether a given referent is implicit at all. When both annotators agreed that a given implicit role was present, κ=0.85.

Examples 145–149 illustrate randomly sampled recoverable implicit roles in the data, to illustrate the range of phenomena seen in the data. One can see that these rely upon a wide variety of pragmatic knowledge:

(145) [**I**]$_j$ enjoy a good political rhubarb as the next person , but this is getting out of hand . (And some of you thought Claus was bad !) To **put** $\emptyset_{agent=j}$ it bluntly , this kind of behavior is precisely the kind of thing the Koch Brothers or Bill Maher get accused of : belittle the opponent , degrade their humanity , malign the intent of their opponents , and on and on .

(146) There 's a lot to **learn** $\emptyset_{agent=j}$ in this path , and [**I**]$_j$ 'm still trying to feel my way through .

(147) He just texted me to **ask** $\emptyset_{personasked=j}$ if any of [**you**]$_j$ French hipforumers are in the area and would be kind enough to give him a lift southwards ?

(148) OP trying to [**win**]$_j$ brownie points from other clueless students who hate the Conservative party just because .
And **failing** $\emptyset_{failing \quad at \quad what=j}$.

(149) [**Why**]$_j$ won't the people of this country wake up and stop allowing Cameron to brainwash them ?
You need not to tell us why , as we already **know** $\emptyset_{theme=j}$.

4.3.4 FiGref: Fine-grained annotations of referential interpretation types

A final dataset proposed here is a set of annotations labeling these "interpretation types" proposed in Chapter 2 of this thesis. This moves beyond simple annotation of only recoverable implicit roles, to characterizing different kinds of non-recoverable roles. The proposed distinctions presented in Chapter 2 were formalized into a hierarchy with examples from English and with tests to distinguish difficult cases; these guidelines are included in Appendix B. This is the first such fine-grained annotation of implicit role interpretation types, but it is comparable to the FrameNet annotations of DNI/INI labels (Baker **et al.** 1998; Ruppenhofer **et al.** 2006), as well as annotations of definiteness or information-structure (Nissim **et al.** 2003; Bhatia **et al.** 2014)

While this largely follows the types proposed in Chapter 2, one major shift is the addition of many types of genuinely invalid implicit roles, as we suspect that this may be useful for characterizing various kinds of non-recoverable roles. One common instance of such non-implicit arguments are instances in which the PropBank sense has an argument which is not valid in that particular context, as with causative-inchoative alternations; a PropBank event such as "open" or "grow" has a role ARG0 for the agent of a given event, but inchoative instances such as "the flower opened" or "the son grew to be tall" don't have such causative agents. Another class of non-implicit arguments are "explicit-only" roles, where a predicate-specific semantic role is unique to that predicate, but is so low-prominence and low-frequency that one would not ever infer it without explicit linguistic material – such as the "coworker" argument of "work", or the "start state" argument for "reform".

A set of 856 instances of implicit roles – 126 recoverable implicit roles and 730 non-recoverable roles – were annotated with this inventory, over documents annotated from the MS-AMR training data, the SemEval-2010-10 training document, and Beyond NomBank (Gerber and Chai 2010). More detailed discussion of the distributions seen with this data is discussed below.

To get some measure of whether this can be consistently annotated, 100 instances of these implicit roles were annotated by a second annotator trained upon the task using a defined set of guidelines. Annotators had a Cohen's κ(Cohen 1960) of 55.2 at the fine-grained classification into the 14-way classification (ten interpretation types and four kinds of invalid roles). Converting these into more coarse-grained four-way representations (definite, indefinite, invalid, or the "partially definite" edge cases) provides κ=58.1, which belies the inherent ambiguity of the implicit role data. As these were scores with minimal annotation and no iteration or adjudication, future annotations could be expected to achieve higher agreement (both in better training and in adding details to the guidelines). The most common disagreements were between REPEATED-EVENT and SCRIPT-INFERRABLE – annotator disagreements about when a particular role is previously mentioned in a prior event or simply inferrable from that prior event.

4.3.5 Updating the English Implicit Role Landscape

Table 4.7 illustrates the state of all data available for training an implicit role resolution system, after the contribution of the current annotations. We suggest that while this is still a small amount of data for training many systems, it should be enough to transfer actual systems for full-text annotation of implicit roles. Moreover, all of the corpora presented below have been converted to a simple standard stand-off format, rather than the current complex formats linked to the syntactic spans within Framenet or Penn Treebank, and the act of conversion to PropBank provides a slightly more consistent treatment of what constitutes an actual predicative event. Data for these corpora (or systems to convert existing data into these modified forms) will be made publically available at `https://github.com/timjogorman/isrl-data-and-conversions`, alongside specifications for the format.

	training recov.	training non-recov	test recov.	test non-recov
SemEval-2010-10	149	579	149	714
BeyondNomBank	718	1251	310	735
ONV5 (Moor **et al.** 2013b)	197	340	16	50
MSAMR	2142	26953	64	217
Within-sentence AMR	3302	9417	-	-
Total	6508	36795	573	1716

Table 4.7: All English data with conversions to PropBank lexicon

4.4 Analysis of Implicit Role Data

There are a number of ongoing empirical questions about the implicit role phenomena seen in English, and we suggest that the interpretation types corpus mentioned above allows one way for us to explore what is going on with this data. This section therefore dives into three more complicated questions regarding implicit roles – the nature of differences in recoverability between different corpora, the nature of differences in referential distance across different corpora, and the relationship between these implicit roles and actual grammatical role phenomena.

4.4.1 Rates of Recoverable Implicit Roles

This first question of the rate of recoverable implicit roles relies upon the "candidates" for implicit roles. We treat a "candidate" implicit role as any semantic role that is core (predicate-specific "numbered" arguments, in PropBank) and has no explicit realization within the text. We can then view the *recoverability rate* as simply being the percentage of those candidates which can be linked to an actual prior linguistic unit in the text.

Table 4.8 illustrates that recoverability rate for the core English corpora. One can see that Beyond NomBank and ONV5 (Moor et al. 2013) have dramatically higher rates of recoverability than what is seen for either full-text corpus, and that the MS-AMR corpus has an extremely low rate of recoverability even when compared to the SemEval dataset.

Corpus	Candidates	Recoverable	rate of recoverability
MSAMR	30211	2206	6.8%
Nombank	3014	1028	34.11%
Semeval	1591	298	18.7%
ONV5 (Moor et al. 2013)	390	213	35.3%

Table 4.8: Rates of recoverability for different corpora. Bottom rows show subsets of MS-AMR chosen to reflect predicate-selection rules from NomBank

There are a number of alternative hypotheses regarding what could lead to this dramatic difference in recoverability rate, although they are likely all partial culpable:

(1) Predicate Bias: The specific predicates selected within NomBank and ONV5 have inherently higher rates of recoverability than the wide range of predicates used in MS-AMR.

(2) Domain Bias: The newswire domain used in NomBank and ONV5 has a much higher rate of recoverability than fiction or discussion forum discussions.

(3) Annotation Method Bias: The differences in annotation methodology (in which MS-AMR is coreference-focused, and the others specifically prompt annotators for each implicit role) lead to more annotations in the prompt-oriented annotations.

While we suspect that the predicate-selection methods of Gerber & Chai (2010) and Moor **et al.** (2013b) lead to a high rate of recoverability, looking at those same predicates within the MS-AMR corpus does not reveal pursuantly high rates of recoverability – if looking at nominal instances of the predicates used in NomBank (e.g. sale, fund, investement), we see a recoverability rate of 6.1%; and see a 9.9% recoverability rate when looking in the MS-AMR corpus for the ONV5 (Moor **et al.** 2013b) data. Thus, variation in predicate recoverability from predicate to predicate is not likely to be the only driving factor in this difference between the MS-AMR corpus and other corpora.

The other measure that we could test was to consider whether the NomBank annotation methodology was leading to annotation of implicit roles which would not be captured by MS-AMR annotators. As an initial, exploratory consideration of this, we did a quick assessment annotation of 10 documents of annotation from NomBank (52 implicit roles total), with one annotator experienced with the multi-sentence AMR annotation, and labeled each implicit role for whether that particular implicit role would be labeled as explicit using the MS-AMR approach. Of these, 34 instances (66.7%) were labeled as acceptable annotations within the multi-sentence AMR framework. Therefore, we can suggest that one part of this discrepancy is emergent from differences in annotation methodology itself. Examples 150–151 below illustrate a number of examples which would not get implicit role status in MS-AMR, which illustrate a mixture of instances where the annotation is simply a far reach (as in the planner in example 150), or examples in which the

	Within-sentence	Cross-sentence Anaphoric	Cataphoric	Within-sentence rate
MS-AMR	701	1154	354	31.7%
NomBank	570	444	14	55.4%
Semeval	86	164	-	34.4%
ONV5	88	125	-	41.3%

Table 4.9: Cross-sentence rates for different corpora

referent is mentioned within the local scope of annotation (as in example 151, where "its" was not annotated, but would have been in the normal scope of NomBank or PropBank annotations).

(150) **[The Democratic Leadership Council]**$_j$, a centrist group sponsoring the **plan** $\emptyset_{planner=j}$ surely thought it might help ...

(151) If the market wo n't pay for [it]$_j$, they argue , [it]$_j$ ca n't be worth *its* **cost** $\emptyset_{theme=j}$

4.4.2 Differences in same-sentence vs anaphoric implicit roles

A second major difference in datasets is the quantity of within-sentence implicit roles vs cross-sentence anaphoric implicit roles. Table 4.9 shows the size of these datasets when distinguishing between cross-sentence and same-sentence implicit role annotations. One can see that the MS-AMR data shows a low rate of same-sentence implicit role labels, and that the NomBank data sits at the other extreme, being dominated by within-sentence links.

4.4.3 Qualitive Analysis – How English Implicit Roles are Expressed

One final way of looking at these questions is to examine the annotated data provided by the interpretation type annotations. Figure 4.4 illustrates the frequency of different interpretation types over non-recoverable implicit roles for the three corpora. This gives us a different kind of characterization of the differences between the datasets – while the MS-AMR data has far more non-recoverable roles, this is less due to a dramatic number of indefinite implicit roles, but far more examples of possible but non-implicit roles, such as these "subsenses" where the role is not valid in context (e.g. the agent of freeze in “my computer is **frozen**”), or “low importance” roles (such as the secondary predication role for “my computer is frozen”). In contrast, we can see that much of the Beyond NomBank (Gerber & Chai 2010) implicit role instances are nonspecific or type-inferrable mentions – which, since these are financial domains, often refer to unstated amounts of assets used to purchase things, or unstated amounts of money invested into things.

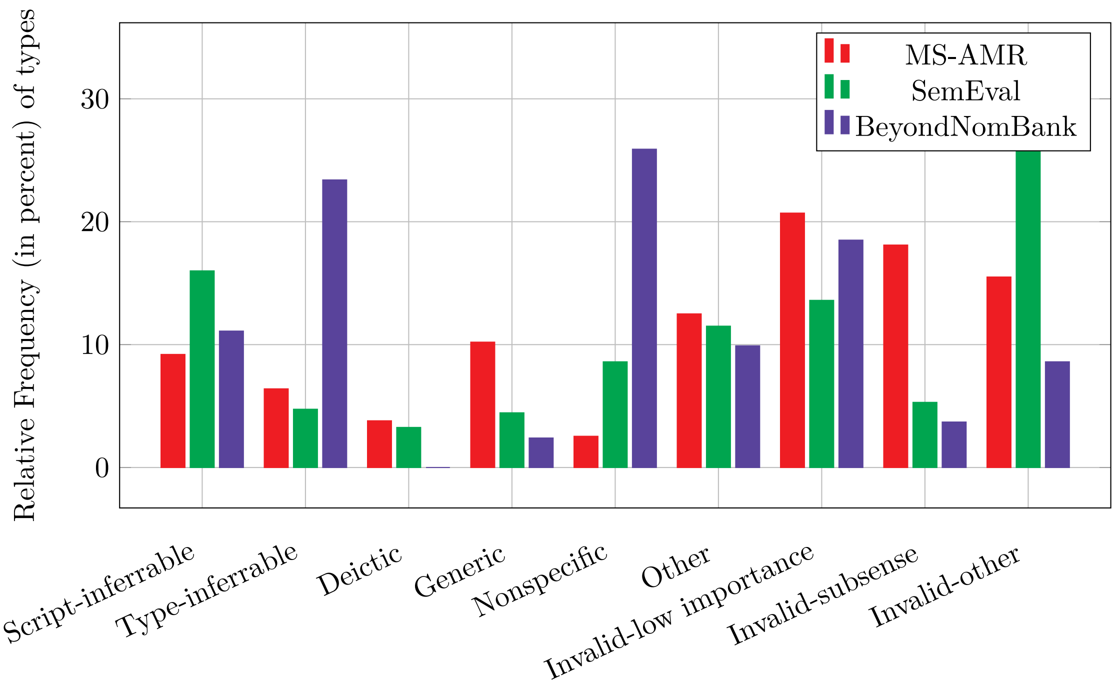

Figure 4.4: Distribution of interpretation types over non-recoverable roles. Other encompasses the long tail of cataphoric/“establishing”, ISNA, and Remembered Event (exophoric) non-recoverables

We can look at a similar distribution over interpretation types for the actual recoverable interpretation types, shown in Figure 4.5. One can see from this that while all three corpora have large amounts of “script-inferrable” implicit roles, the NomBank data is dominated by the REMEMBERED EVENT interpretation type, as well as by certain types of annotation errors (mentions which are within the explicit SRL scope of the predicate). This illustrates how systems trained or developed with the NomBank can be targeting a different task than what is seen with other two corpora.

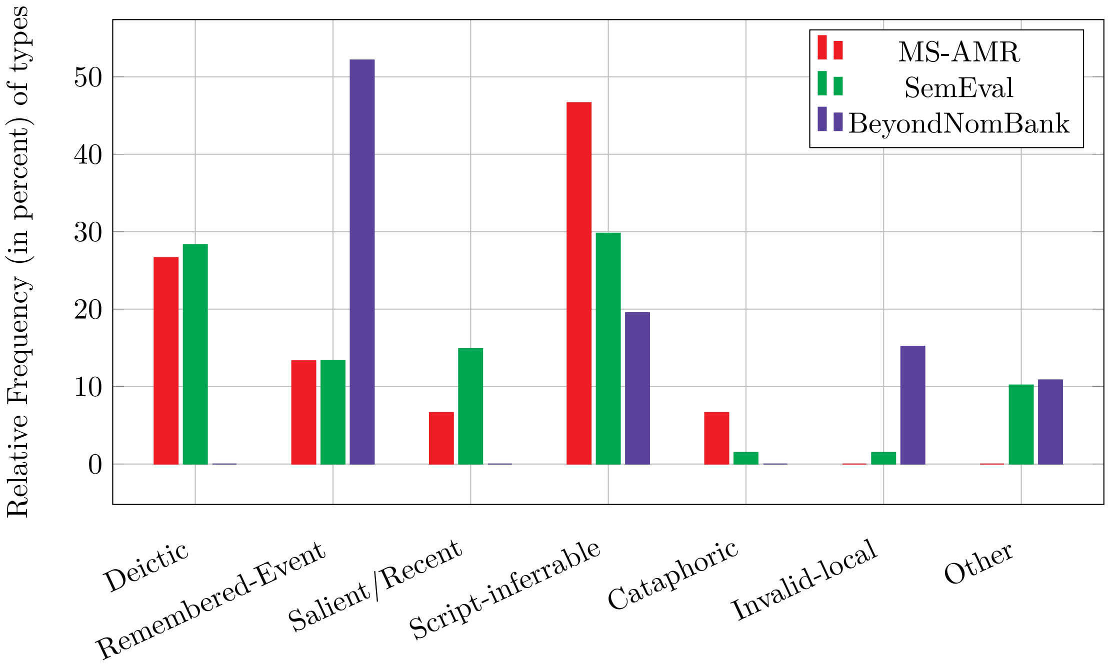

Figure 4.5: Distribution of interpretation types annotated on data with recoverable referents, primarily showing the prominence of "remembered event" mentions in the Beyond NomBank corpus. Invalid-local refers to implicit roles that would be annotated in traditional SRL labeling.

4.4.4 Qualitative Examination – Implicit Role Types and Constructions

The final underlying question with this data is whether these particular kinds of implicit role interpretations can be linked to specific grammatical constructions. One starting point was to analyze all instances labeled as SALIENT/RECENT within this annotation. We do find a small amount of conversational, diary-style implicit subjects, both first person and third person, although they often also have explicit mentions in other parts of the sentence:

(152) $\emptyset_i$ just looking for some more **advice** $\emptyset_i$ on my$_i$ situation .

(153) $\emptyset_j$ 'Sold his soul to the devil in exchange for money , ' says Warner , 'and $\emptyset_j$ **expects** his$_j$ creditor to come up and claim his own . '

However, the majority of salient implicit roles categorized in the data could be referred to as "last proposition" examples – implicit complements of a verb referring to the last mentioned proposition. These occur primarily within the SemEval corpus, which is fiction and therefore abounding with direct-speech quotation:

(154) “To you they are like crimes committed in some other planet. But WE know $\emptyset_{aboutthecrimes}$.”

(155) “I know $\emptyset_{thingknown}$, sir, I know; but it shook me, sir, and there ’s no use to deny it .”

(156) “You ’re right$\emptyset_{accurateproposition}$, Mr. Holmes .”

There are some mentions that aren’t such propositional content mentions, but which are on the border of being “explicit” mentions, as they refer to a very recently mentioned referent in a manner that would be described as pragmatic control (as in example 157), or might be viewed as being part of a multi-word expression (as in example 158).

(157) **I** was recently made redundant a week after $\emptyset_{agent}$ announcing my pregnancy .

(158) **Inspector Baynes** ’s small eyes twinkled with $\emptyset_{experiencer}$ pleasure .

We can similarly review kinds of DEICTIC arguments. Many of these metalinguistic communication events – communication verbs that refer to the current discussion, and which therefore automatically map to the current speaker and the current addressee:

(159) and needless $\emptyset_{forme}$ to **say** , there was no open casket .

(160) When I saw your longest entry , I guessed what your choice was without you **explaining** it $\emptyset_{to \quad me}$.

(161) As I **said** $\emptyset_{to \quad you}$: ..

4.4.5 AMR Re-entrancies as Implicit Roles

As these AMR re-entrancies can refer to implicit roles, it is possible that AMR parsers automatically learn to predict these implicit roles. However, AMR re-entrancies do not only express implicit roles, but are used for any situation where a participant is referred to more than once in a sentence. Therefore, we provide a small exploration of how these implicit roles are distributed, both in gold AMRs and in predicted AMRs, to make clear how common implicit roles are within the AMR data.

For determining whether implicit roles are being captured by AMR parsing models, we annotated the AMR development set with the Zhang **et al.** (2019b) parser, the parser with the highest reported performance upon reentrancy prediction at the time of writing. Using that parser (trained upon the AMR 2019 release), all sentences in the development set were parsed, and for those where a predicted re-entrancy matched a semantic role in the original AMR, we examined those instances manually regarding whether it actually referred to the correct referent, and was an instance of an implicit role. From a set of 750 initial predicted re-entrancies used, we found 50 correct implicit role instances – approximately 6.6% of all postulated re-entrancies being actually reflective of an implicit role.

Another way of viewing the same question is to examine the correct re-entrancies found in an AMR, and to study which phenomena phenomena are being captured. We established 100 instances of re-entrancies from the manual annotations of the AMR development set, and 100 correctly predicted re-entrancies from the Zhang **et al.** (2019b) AMR parser. Of these, we categorized each re-entrancy into one of seven types. Of those correct re-entrancy predictions, the predictions of the Zhang **et al.** (2019b) parser were implicit roles approximately 20% of the time, as seen in Figure 4.6. Examples of the kind of implicit roles captured by AMR parsing are shown in examples 162–164 below, with the predicate in red and referent in blue:

(162) However, over the last month or two [I] 've found myself becoming really anxious about things, and I 'm not talking about a mere out of the blue feeling of nervousness.

(163) After strategically ' excising ' bits of Australia in order to thwart people smugglers (you can't land in Australia if we pass a law to say it IS N'T Australia this week), [our pint-sized PM] has decided to think big .

(164) Retain 30 strategic [nuclear submarines] with the ability to inflict (ARG0) three devastating nuclear strikes against enemies.

Beside those true implicit roles, three of the other types of reentrancies seen would be generally categorized as local instantiation, such as *pronoun* use, *coordination* of subjects, and *referring expressions*, where a participant is mentioned nominally multiple times in the same sentence. One can see in Figure 4.6 that such very local mentions constitute a large portion of all re-entrancies, particularly for gold AMRs.

(165) Coordination: [It] invests heavily in dollar - denominated securities overseas and is currently waiving management fees , which boosts its yield .

(166) Ref. Expression: Removing rules doesn't help [business] be more competitive, it means that [businesses] must race to the bottom to counter new businesses that start there.

(167) Pronoun: [My husband] is a veteran. He said if a war takes place, he would stand in the very frontline of the war - rather die on the war field than live on in degradation

Finally, there are many instances of control phenomena, and non-deterministic control phenomena, which are also captured by AMR parsing. Alongside these, another kind of re-entrancy captured is a bundle of phenomena associated with comparative constructions. In the most recent models for AMR comparative constuctions (Bonial **et al.** 2018), the entity being modified by an adjective is modified twice – so that the AMR for "Blair is taller than Sam" would refer to Blair twice, both to assert the comparison itself ("Blair has more <tallness> than Sam") and also to assert the attribute ("Blair is tall").

(168) Control: [The House] has voted to raise the ceiling to $ 3.1 trillion , but the Senate is n't expected to act until next week at the earliest .

(169) Pragmatic Control: [It] can compete with articles from the Cultural Revolution, magnifying a problem to become a question of principle or of ideological line, buttoning the hat.

These results show that AMR parsing models do learn to make some implicit role predictions. Perhaps more surprising, this shows that the AMR parsing models learn the full range of different kinds of re-entrancies seen in the data, rather than simply learning the most predictable systems such as coordination or control. However, the limited number of implicit roles seen, and the variety of kinds of re-entrancies seen in the data, support the postulate of the within-sentence implicit role AMR corpus, that only a tiny fraction of AMR re-entrancies should be treated as implicit roles.

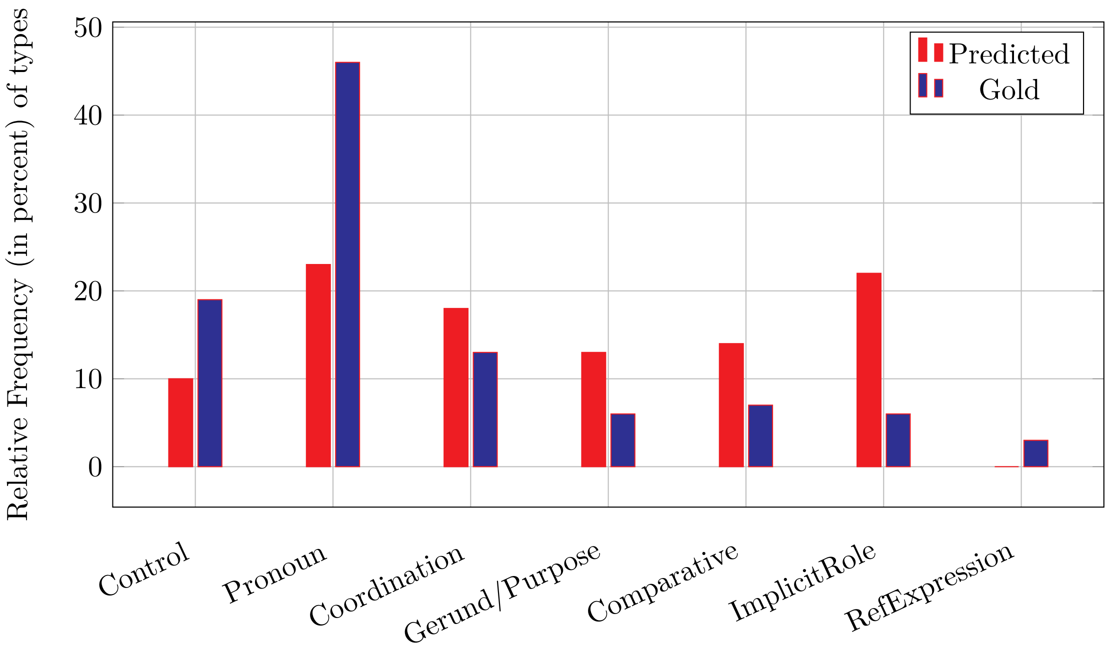

Figure 4.6: Distribution of types of (correct) reentrancies, from both predicted and manual AMRs. This illustrates both that models trained on AMR do capture implicit roles, but that the majority of re-entrancies are more explicit or syntactically defined re-entrancies

4.5 Conclusions

This chapter provides a strong starting point for new models of implicit role resolution, but it also hopefully illustrated the complexity of this task, even with the addition of new datasets.

Because implicit role resolution phenomena are quite heterogeneous, it is unclear how to get a model to handle this wide range of phenomena. Moreover, the type annotations illustrate that some datasets can end up having dramatically different characteristics – such as the very low rates of recoverability seen with the MS-AMR dataset, or the very high rates of REMEMBERED EVENT implicit roles that seem to exist in the Beyond NomBank corpus.

Chapter 5

Computational Models of Implicit Role Resolution

This chapter presents models of implicit role resolution for the English datasets discussed in Chapter 4. This chapter explores the development of iSRL systems aimed at running upon all of these implicit role corpora, even with all of them together. In doing so, we also explore attempts to transfer knowledge from larger related coreference and SRL datasets while maintaining some extent of interpretability, by building separated approximations of the major latent features postulated to be relevant to the task. We present a set of individual models for these components underlying implicit role resolution – such as selectional preference, salience and the recoverability of the implicit role – and a set of actual implicit role models which use these to train upon all the iSRL corpora.

The rest of this chapter will break down as follows. The underlying model relies upon a series of separate feature-specific models trained on larger corpora – selectional preference, narrative schema predictions, salience, mention referentiality, and a deixis score. The first section will outline each of these feature models, attempting to establish the use of competitive models for each of these components. The second section will outline an architecture for implicit role resolution and analyze its performance. A third section presents ablations and analyses of these models, attempting to illustrate what is missing from the current approaches and the relative contribution of different components. While the results of the models presented in this chapter are limited, they present a new frontier in training upon the full implicit SRL task and illustrate some characteristics of the implicit role resolution task.

5.1 Details of the Task

There are many ways to model implicit role behavior. For example, one might build models for when implicit roles are licensed, or build models to predict what indefinite implicit roles refer to. However, we focus upon one subset of this task: the prediction of *recoverable* implicit roles, where an implicit role is not only definite, but the participant it refers to is explicitly mentioned within the text.

The following models will also assume a specific form to this task of recoverable implicit role labeling, in which a list of "candidate" implicit roles are given, and the task is to provide a single span which corresponds to the referent, when recoverable. This follows the formulation of this task in prior implicit role resolution tasks (Ruppenhofer **et al.** 2010a), but makes more explicit the notion of starting with a pre-defined list of candidate implicit roles.

Specifically, the input to a model is then a raw text, the location of a predicate, and the label for a semantic role, providing a full PropBank numbered argument and roleset. For each provided predicate-role pair, the system would then provide either "None" or would provide a particular span which is the model prediction for a single span which it postulates to refer to the antecedent. The annotated data for each recoverable implicit role contains a list of each spans which refers to that referent, and we therefore aim to measure the model in terms of whether the single span it predicted is within the correct coreference chain.

The particular methods of scoring this prediction uses the SemEval-2010 task 10 (Ruppenhofer **et al.** 2010a) scoring metric for evaluating each implicit role[1] . Each implicit role instance r has a span for the predicted referent (the list of unique characters $Predicted_r$) and a list of correct spans C_r. The score is then the highest Dice score (Dice 1945) from the set of correct spans.

$$score(S_r, p_r) = \max_t^{C_r}\left(\frac{2 \mid Predicted_r \cap t \mid}{\mid Predicted_r \mid + \mid t \mid}\right)$$

The final precision is then the sum of this score over the number of predicted spans, and the

[1] This is simplified form if this metric, as some implementations add the additional constraint that the score for a given (predicted, true) pair will be 0 if the predicted span does not contain the headword of the true span.

final recall is the sum of this score over all implicit roles with at least one correct span.

$$precision = \frac{\sum_r^R score(S_r, p_r)}{\sum_r^R 1 if \mid p_r \mid > 0}$$

$$recall = \frac{\sum_r^R score(S_r, p_r)}{\sum_r^R 1 if \mid S_r \mid > 0}$$

A baseline implementation of this scoring function is provided at `https://github.com/timjogorman/isrl-data-and-conversions` to encourage replicability, and details of the assumed format will be included therein.

5.2 Feature Models

The first section looks at ways of learning models for particular latent features, using larger available resources such as SRL corpora or coreference corpora. We focus on building separate machine learning models for each of these features, so that one might treat them as separate components within a larger model, using a total of five features: *selectional preference*, *narrative schema effects*, *deictic tendencies*, *salience*, *mention importance (referentiality)*. Because end-to-end models which are trained on the task in a more general manner can learn all of these tasks, it can be unclear what recent deep learning models (such as Do **et al.** (2017) or (Cheng & Erk 2018a)) are learning. This approach, therefore, attempts to separate these different components, while still attempting to train individual components upon larger, separate resources using SRL and coreference corpora. Additional details and parameters used for each model are included in Appendix A.1.

5.2.1 Linguistic Interpretability and Feature-based Models

There are some possible uses of implicit roles for downstream tasks, such as during translation to and from languages with high levels of implicit role usage (Chung & Gildea 2010; Xiang **et al.** 2013)), or capturing long-distance links needed in information extraction tasks such as ERE (Song **et al.** 2015). However, the best-performing systems for implicit role labeling in English cannot exceed 50 f_1 without gold parse and SRL information, and performance upon this "full text" task (as with SemEval iSRL, or the newly presented MS-AMR data) remains below 20 f_1. The low performance of implicit role resolution models in English motivates focusing not upon maximizing performance on the task, but instead upon exploration of the task: to be able to provide some measure of interpretability regarding how these implicit role tasks work, what current models are learning, and what elements of a model need the most improvement.

There are many ways to attempt to get insight into model behavior. Due to the limits of the data landscape for implicit roles, we focus upon updating a traditional, feature-oriented approach to interpretability – attempting to use simpler and more additive models wherein one can examine the contribution of particular features. In doing so, we follow the approach of Generalized Additive Models (Hastie & Tibshirani 1987), in which features are combined additively (as in logistic regression), but are transformed by a nonlinear function, such as smoothing functions (the original model), tree boosting (Lou **et al.** 2012; Caruana **et al.** 2015), or small single-input MLPs (De Waal & Du Toit 2011). This gives a measure of the kind of predictive power seen in logistic regression over small sets of features while adding some additional predictive power.

While it is hoped that the current feature-driven approach provides some level of insight, we will also hope to keep in mind the more recent advances, in which feature-based approaches have been largely replaced by *diagnostic classification* or *probing* (Tenney **et al.** 2019b; Tenney **et al.** 2019a; Ettinger **et al.** 2016; Adi **et al.** 2016) tasks. In such tasks, one builds deep learning models to do a particular task (potentially even a general task such as language modeling) and then studies the representations learned by those deep models by using those representations as inputs to small, models applied to auxilliary tasks each corresponding to a single phenomena, so that performance upon such a task illustrates whether that representation has inherently learned the information involved in that task. This chapter will end with exploration of how to pivot from treatment of this task as a traditional supervised task, to establishing the necessary data for a future *probing task* of this phenomenon, using the interpretation types presented in prior chapters to help define the task.

5.2.2 Representation of Selectional Preference

Broadly, selectional preference of an argument gives some measure of whether a particular entity would be likely to fill a particular role within an event. This is a complicated issue to measure, however. Early models of *selectional restriction* or *plausibility* would measure the plausibility of a referent fitting into a particular event role – whether it is theoretically capable of filling that role (Katz & Fodor 1963). While such plausibility-focused work was common with rule-based systems (Wu & Palmer 1994) and continues to this day in more advanced forms (Wang **et al.** 2018; Pustejovsky **et al.** 2017), most research on selectional preferences have found it hard to cleanly separate the grammatical plausibility restrictions from more pragmatic constraints provided by probability and world knowledge (Wilks 1978; Johnson-Laird 1983; Fodor 1975; Resnik 1993), and therefore the bulk of modern computational research on selectional preference has focused on measures that roughly express whether a particular (predicate, role, mention) is *likely* to be mentioned.

We operationalize selectional preferences as the compatibility of a particular semantic role of a predicate with a particular mention (for our purposes, a headword). For a particular pairing of (predicate, role, mention), a selectional preference model would give a score for how often that pairing might occur. While one could measure the likelihood that "lasagne" will be the object of "devour" by counting the number of times it occurs in a large corpus (Chambers & Jurafsky 2010) note such direct count models are still a surprisingly good baseline), such a count would fail to provide generalization about foods generally being the object of eating events. Different algorithms for selectional preference largely differ in introducing different ways of adding this generalization. These can loosely be broken up into three kinds of approach: a first generation of approaches that provided clustering over discrete sets of words or concepts – usually referencing WordNet (Fellbaum 1998), which provides a good starting point for such generalization, a second generation of approaches in which mentions were modeled distributionally and compared using vector similarity measures such as cosine similarity, and a third generation of approaches in which both the "mentions" and the "roles" could be modeled with neural methods.

Models of learning clusters or distributions over WordNet concepts developed a range of techniques for doing so, such as simple clusterings of words and concepts, partitions of the WordNet tree (finding a correct parent node), determining paths from WordNet roots, or learning latent topics that provide distributions over those paths or partitions (Abney & Light 1999; Abe & Li 1993; Clark & Weir 2002; Chen 2006; Van Durme **et al.** 2009; Ritter **et al.** 2010; Séaghdha & Korhonen 2012; Wu & Palmer 2015). Some also developed generalizations over words (instead of WordNet concepts) (Rooth **et al.** 1999; Séaghdha 2010), or postulated latent classes over the roles as well (Pereira **et al.** 1993; Séaghdha & Korhonen 2012). Indeed, one might also consider hand-crafted or semi-supervised approaches to generalizing over words, such as VerbNet (Kipper **et al.** 2000) or lexical sets (Hanks & Jezek 2008; Jezek & Hanks 2010; Hanks 2006), as a related way of generalizing over individual words.

The issue of generalization can be modeled using distributional semantics instead, by representing a mention using vectors learned from large corpora. While this has been discussed speculatively since the earliest computational models (Resnik 1993 proposed SVD vectors for this), the first prominent model was that of Erk (2007). This approach learned vector representations for mentions, and then represented the selectional preferences of each actual predicate+role pair by a bag of mentions seen filling that predicate+role pair. Thus, if one were to consider "devour" and "lasagna", one could have a bag of prior words that occurred as the object of "devour", compare the similarity of each of these exemplars with a vector for lasagna using a measure such as cosine similarity, and then aggregate those scores (i.e. with a weighted average). Extensions of this model have been proposed, such as doing some aggregation over these seen exemplars (Schenk & Chiarcos 2016; Baroni & Lenci 2010), or using such approaches as a backoff in combination with simple count-based models (Chambers & Jurafsky 2010).

The natural extension of vector-based models is to directly learn both vector representations and the functions for comparing those representations, using neural nets instead of a simple similarity function. Van de Cruys (2014) presented models that would concatenate a vector for a predicate and an object (for a "two-way" model) or subject and object ("three-way" model), by training the model to perform the so-called "pseudodisambiguation" task – taking observed predicate-role-mention triples from a corpus (e.g. "devour+lasagna") and scoring them higher than randomly generated ones ("devour+kayak"). Do and Bethard (2017) also trained a neural network to do a selectional preference task, although instead of a negative sampling task, they directly trained a model to predict mentions given a verb (and its prior context), treating the task as something like a language modeling task.

5.2.2.1 Selectional Preference Feature Model

To develop a selectional preference model to use for iSRL, we start with a re-implementation of Van de Cruys (2014) and make a series of modifications to match the constraints of the current iSRL task. The original model assumes a single grammatical role (object), learns a vector for each predicate and each mention, and learns a scoring function which concatenates and compares the two. This is a simple feedforward neural network:

$$score(p, r, m) = \text{FFNN}([Embedding(m); Embedding(p)])$$

Specifically, this uses a 100-dimension hidden layer, 50-dimensional embeddings, tanh non-linearity, and is trained with pseudodisambiguation using a hinge loss where:

$$loss = max(0, margin + Score(p, r, m_{negative}) - Score(p, r, m_{correct}))$$

We follow Van Der Cruys(2014) in training and evaluating using "pseudodisambiguation", comparing positive examples with randomly selected negative samples, re-implemented using the syntactic annotations of APW section of the Annotated Gigaword (Napoles **et al.** 2012). We compare the re-implementation (just trained on direct objects) with the same scores from the Van Der Cruys, using what appears to be the sampling method from Van Der Cruys (2014) uniformly sampling from the vocabulary. Table 5.1 illustrates the performance on that task, showing approximately the same performance.

model	role representation	sampling	acc.
Reported in Van Der Cruys (2014)	objects	uniform	0.88
Re-implementation	objects	uniform	0.906

Table 5.1: Comparison to Van Der Cruys (2014), to illustrate general replication of the model

The first of two modifications to make this compatible with the implicit SRL task was to extend the model from only working with direct objects to working with all grammatical roles. We modify the Van Der Cruys(2014) model slightly, replacing the embedding representing each predicate (e.g. "eat") with an embedding representing each (predicate, role) combination (e.g. "object of eat" or "instrument of eat"). In other words – the original Van Der Cruys two-way model would start with a predicate-object pair such as "eat" and "sandwich", randomly sample a word such as "truck", and would be trained so that the score S(eat, sandwich) should be higher than the score S(eat, truck). Doing so simply entails changing that task to one in which the model is comparing S("object of eat", "sandwich") to S("object of eat", "truck"). We can compare this against the "surprisingly effective baseline" of Chambers and Jurafsky (Chambers & Jurafsky 2010), and we achieve slightly better performance with the current model, although there are small differences in the corpora used which make them not perfectly comparable. Nevertheless, we suggest that getting roughly similar scores supports the general approach.

Model	sampling	acc.
This Re-implementation + role-predicate embeddings	bucket	93.7
Chambers and Jurafsky(2010) count baseline	bucket	91.7
Chambers and Jurafsky(2010) count + Erk smoothing	bucket	92.6
Chambers and Jurafsky(2010) count + Google smoothing	bucket	91.9

Table 5.2: Pseudodisambiguation over all core and prepositional dependencies (subject, dobj, prepositions) using bucket sampling. Comparison not exactly equal (evaluation on Gigaword vs NYT)

Finally, we then modified this model to also handle PropBank semantic roles and senses (Palmer **et al.** 2005), specifically the PropBank 3 inventory (O'Gorman **et al.** 2018a). This involves a pivot from ambiguous pairs of dependency relations and predicate lemmas such as "subject of break" into

Model	Role dataset	PB validation	PB test
Current model	Gold PB	90.47	86.5
Current model + dep	Gold PB + Gigaword	89.4	85.15
Current model +similarity constraints	Gold PB + Gigaword	**92.1**	**86.63**

Table 5.3: Pseudo-disambiguation against gold Propbank arguments

more specific characterizations such as "arg0 of break.01" (the agent of breaking). However, the amount of data with gold PropBank labels is relatively small. While many different approaches were attempted, our final model used the same simple approach taken with dependency labels – simply replacing dependency and lemma pairs such as "object of eat" with PropBank senses and numbered arguments such as "arg1 of eat.01". We then trained using the data annotated with gold PropBank labels from OntoNotes (Hovy **et al.** 2006).

We also attempted a variety of methods to leverage automatically labeled data to augment that limited set of gold PropBank data. One method that was attempted was a multi-task learning approach (Caruana 1997): training a model to do both PropBank selectional preference and dependency selectional preference, using the larger amount of automatic dependency data from the Annotated Gigaword. This approach by itself did not improve performance, but did help the selectional preference performance when combined with a constraint that those PropBank arguments (such as "arg1 of break.01") should be similar (via cosine similarity) to dependency arguments that express them (such as "direct object of break"). This allows a model to learn tendencies from the much larger corpus, as illustrated in Table 5.3.

This model is therefore used as the approximation for selectional preference in the final implicit SRL models, and might be judged based upon its performance in that downstream task. We can also measure the correlation in other ways, such as against the human plausibility scores of Padó **et al.**(2007) – where this model has a Spearman correlation of r=0.290 with the human judgements. The kinds of patterns learned by such a model are illustrated in Table 5.4 below; this illustrates that these models learn both general world-knowledge biases as well as specific high-frequency combinations (such as "come to a close" or "deal with it"):

arg1|offer.01	arg0|outline.01	arg2|wait.01	arg2|deal.01	arg4|come.01
thing offered	outliner	waiting for what	what it dealt with	destination/goal
comment	leader	help	violence	close
money	lawyer	work	strike	place
approval	statement	return	it	point
apology	she	call	arrest	time
help	analyst	vote	problem	end

Table 5.4: Examples (selected from top 20 of high-frequency args)

5.2.2.2 Selectional Preference Factors Omitted here

No current model of selectional preference handles all issues that might be considered important for selectional preference and implicit roles. We suggest a few issues that are worth mentioning even though they were not used in the current model:

- This model trained systems using the head-word of a mention alone – ideal models might consider larger representations such as representations of the rest of a mention span, representations of named entity type (Schenk & Chiarcos 2016), or modern representations of word meaning in context such as ELMO or BERT (Peters **et al.** 2018; Devlin **et al.** 2018).
- Wang et al. 2018 notes that models of selectional preference do not capture plausibility effects, one might want a separate system for modeling implausible, rather than simply uncommon, referents.
- This scores single event-role-mention triples out of context, without considering information from the larger context, such as considering multiple arguments at once (Le & Fokkens 2018; Do **et al.** 2017; Van de Cruys 2014). This is especially important when selectional preferences are essentially encoding world knowledge; one might expect a model to give very low scores to "eat houses", but might hope to give high scores to particular combinations such as "termites eat houses".
- Finally, as will be mentioned later, this also omits interactions between selectional preference and script information. We omit such interactions in order to cleanly separate

the effect of selectional preference from the expectations provided by narrative schema, but many models could be said to capture some selectional preference information within those models of scripts and frames (Ferraro & Van Durme 2016; Cheng & Erk 2018a; Chambers & Jurafsky 2009; Chambers & Jurafsky 2011; Do **et al.** 2017).

5.2.3 Narrative Schema Features

Narrative schema models look at how a particular event role fits into a larger script or storyline – reflecting human world knowledge about how various events lead to other, related events and scenarios. There are a range of details here in the long lineage of work on "schemas" or "scripts"(Schank & Abelson 1977; Johnson-Laird 1980; Rumelhart 1975), but we focus on the subset of this literature which can characterize a participant, and their likely future actions, based upon previous events. While the original works on narrative schemas postulated hand-crafted lists of scripts, many modern approaches have focused upon learning these models from data (Chambers & Jurafsky 2009; Ferraro & Van Durme 2016; Jans **et al.** 2012). Figure 5.1 shows a famous illustration from Chambers and Jurafksy which illustrates the kind of learned templates developed by these models – which would predict, e.g., that if one is the direct object of "raid", then they will also be predicted to be the direct object of "arrest" and "charge".

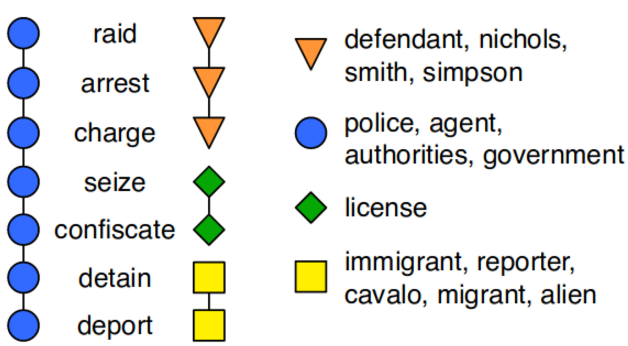

Figure 5.1: Figure from Chambers and Jurafsky (2010) illustrating generated narrative schemas

Chambers and Jurafsky (2009) proposed to evaluate these models with the *narrative cloze* task – in which we take a referent in a document participating in a series of events, and then need to predict the next (event, role) pair. While the details of this task have been modified over time [2] , this "cloze task" has become the common method of evaluating these models.

5.2.3.1 Narrative Schema Model Background: The Pivot towards Event Language Models

Rudinger **et al.** (2015) noted that this narrative cloze task could be reframed as a simple language modeling task – in which one does not need to build discrete clusters for schemas, but one can instead simply take a series of events (event-role pairs), and predict the next occurrence in that sequence. They were able to adapt traditional neural sequence-prediction approaches, using a log-bilinear model developed for predicting the next word in a language model (Mnih & Hinton 2009). This line of research – viewing such "script" effects as a simple prediction task over sequences – was expanded with the addition of more complex, LSTM-based representations of sequences. Examples of such more complicated approaches included adding more rich representations of the events and arguments, (Pichotta & Mooney 2016; Pichotta 2017) , making predictions of entity-argument pairs rather than just events (Do **et al.** 2017), or using multi-head attention over events and arguments (Cheng & Erk 2018a).

5.2.3.2 From Narrative Cloze to Argument Cloze

The prediction of the next predicate+role "event" given a sequence of prior predicate-role pairs, is easily affected by the frequency of the predicate-role pair being predicted, so that it is often hard to beat a simple model that ranks events by their frequency (Reisinger **et al.** 2015; Jans **et al.** 2012; Chambers 2017). Yet in tasks such as coreference or implicit SRL, the "next" event is given. In that sense, tasks such as coreference can be more directly modeled with a slight reformulation of the same measurement: instead of assuming a particular referent history and predicting the next event-role pair, one can start with that next event, and predict which sequence

[2] The main change in how people evaluate narrative cloze is that while original evaluation studies how often the model exactly predicts the next event, Jans **et al.**(2012) evaluates how often the correct event occurs within the top k arguments

of prior events (which referent) that event came from. This is proposed by Cheng and Erk (Cheng & Erk 2018b) and framed as the "argument cloze" task. We adopt this reformulation of the task: although it has natural parallels to the main "narrative cloze" task, it also is naturally closer to the coreference task that we focus on here.

5.2.3.3 Narrative Schema Model and Evaluation

We follow this "argument cloze" approach of (Cheng & Erk 2018a). As with models for selectional preference, this model also shifts to the use of Propbank semantic roles, rather than the more common grammatical roles used in prior work.

We use an LSTM (Hochreiter & Schmidhuber 1997) to traverse a series of prior event-role pairs. Each event instance is characterized by an embedding for the role, and an embedding for the roleset. This is therefore similar to the general RNN approach (applying an LSTM to a sequence of event-role pairs) used in some prior models (Pichotta & Mooney 2016). The model is trained with cross-entropy loss, to maximize the likelihood that an event-role pair would be assigned to the correct coreference chain. However, we train and test upon sets of 9 randomly sampled distractor coreference chains, instead of over a full set of possible antecedents from the original documents. Table 5.5 reports the argument cloze evaluations for this task. As with the selectional preference feature model, this approach also utilized Propbank data, and therefore focused upon the limited amount of data annotated with SRL and coreference labels. Models used primarily relied upon data annotated with Propbank semantic roles and human-annotated coreference chains, using the coreference data released with the CoNLL shared task (Pradhan **et al.** 2011). 2 million words of additional automatically labeled coreference and propbank were added to that training set, but no improvement was found with that data augmentation. Similarly, a series of attempts were made to transfer knowledge from larger automatically annotated coreference annotations with dependencies (the AFP section of the Annotated Gigaword (Napoles **et al.** 2012)), but experiments showed negative results.

model	val	test
LSTM (PB-GOLD)	0.272	0.268
LSTM (PB-GOLD + (PB-AUTO)	0.272	0.268
random chance	0.10	0.10

Table 5.5: Score on argument cloze vs random negative chains

5.2.3.4 Narrative Schema Future Work – Details Omitted from the Model

We omit the interaction between narrative schema effects and discourse structure. If you had an event such as “X was arrested”, then true human script knowledge would provide one set of expectations if the next event was marked as a continuation “and then” (e.g. “ and then X went to trial”), a different set of expectations if the next event was marked as contrastive with “but” (e.g. “but they had to let X free”), and other different expectations for an explanation relation with “because” (e.g. “because X had robbed Z”). Storylines models and the EventStory dataset (Caselli & Vossen 2017) or the “story cloze” task of Mostafazadeh **et al.** (2016) (a variant of narrative cloze designed to have more prototypical, script-oriented event sequences) are oriented towards models that can handle such discourse or narrative structure. Peng 2018 showed one approach to such issues, directly adding discourse relations alongside events within the sequence of events being modeled.

5.2.4 Deictic Tendency of Arguments

An additional feature we explore (based upon the discussion on Chapters 2 and 3) concerns the use of implicit arguments to refer to deictically present referents, particularly the speaker or addressee. The motivation for such a feature is to have a feature which would remain at 0 for non-deictic antecedents, but which would provide a score for speakers and addressees, proportional to how biased the semantic role is towards referring to deictic referents. A model for a “deictic tendency” would therefore simply be a model wherein high scores are given for semantic roles which are likely to refer to a speaker or addressee (such as the cognizer role of mental state verbs, or the adjudicator role of evaluation roles).

We measure this score as zero for non-deictic antecedents, and as a simple combination of the total selectional preference assigned by that role to all locuphoric pronouns (e.g. "I", "you", "me", etc.) when an antecedent has any such locuphoric pronouns in their coreference chain. There are theoretically other ways of approaching this – such as measuring the likelihood of a verb participating in first-person narratives (e.g., Pavlick & Nenkova(2015)) or by also factoring in modal modifiers. The consideration of modal situations which might entail deictic implicit roles has been noted to be very important for this measurement for other languages, particularly those without any person agreement upon verbs: Nakaiwa and Shirai (1996) attempted to predict deictic implicit roles in Japanese using co-located deontic modal verbs and the membership of the predicate in certain verb classes. Han (2006) also looked at similar issues for Korean and found deictic implicit roles to be correlated with Korean grammatical moods (imperative/ exhortative) – both in general and for resolving distinctions between 1st and 2nd person.

5.2.5 Mention Referentiality Models

One common feature used in coreference systems is to approximate a measure of *referentiality* – whether a mention is capable of participating in coreference chains and being anaphorically referred to.

As with selectional preferences and restrictions, there are many clear-cut examples of things in the literature that are *categorically* non-referential, such as expletive pronouns and certain kinds of mentions with only local referential scope (Karttunen 1968), but in practice we conflate that categorically non-referential material with a more general measure of reference likelihood than encompasses questions of importance and topicality – e.g. "the hat" in "the man in the hat" is not categorically non-referential, but it relatively unlikely in to be mentioned again.

Modern models, therefore, learn to predict whether a particular phrase or span of text should be treated as "a mention", by predicting the more concrete measure of whether that span will occur in a coreference chain at all (Bengtson and Roth, 2008; Rahman and Ng, 2009; Recasens, de

Marneffe, and Potts 2013, Wiseman et al. 2016). Such measures naturally only *correlate* with any true measure of referentiality, since not all referential mentions are actually anaphoric, but this is commonly viewed as a useful correlation.

In building the actual feature model for iSRL, we model this as a prediction for whether a mention will be coreferential, using the coreference chains in the OntoNotes coreference data (Pradhan **et al.** 2011). For this, we start with a simple set of features largely following Recasens **et al.**(2013) – the part of speech of the headword, its dependency relation, the lemma, and a sequence of the headwords and dependency relations of each dependent of this mention, and the lemma of the mention's head. This captures many features used in prior models, such as negation or the presence of definite and indefinite determiners. All of these features were embedded (d=20, with Dropout of 0.2, (Srivastava **et al.** 2014)) and then simply summed, with a final hidden layer making a scalar prediction. A variety of more powerful architectures were attempted (e.g., LSTMs), but this approach performed the best. Table 5.6 reports how well this model did upon predicting anaphoric mentions in the OntoNotes validation test, and compares it to the reported scores from Recasens **et al.**(2013); while our model does not beat their well-engineered system, it approaches that general performance.

Model	val. precision	val. recall	val F1
Feed-forward Simple	79.53	65.4	71.8
Recasens **et al.** (2013)	76.4	76.6	76.5

Table 5.6: Models for predicting importance of mentions

5.2.5.1 Models of Salience

One might contrast the idea of a general measure of an entity's importance to a much more context-specific measure of how salient or prominent that mention is within a particular point in the discourse. As discussed in prior theoretical chapters, this has been traditionally split into two kinds of approaches – coherence-oriented approaches and activation-oriented approaches. Coherence-type models generally can be treated as hypothesizing that we maintain ongoing lists of discourse referents, particularly in terms of how they have related to the ongoing discourse, and that discourse structure and pronoun use act jointly – so that pronominal mentions can technically be defined not simply as something that is most active at a given point, but also as something which helps to best aid in the coherent understanding of a discourse. Computational implementations of such models were the first representations of salience to be used in implicit role resolution, as the PUNDIT system (Palmer **et al.** 1986; Dahl 1986; Dahl **et al.** 1987) deployed a version of the "focus lists" of Sidner (1979) in their reference system. Later models using centering theory, or using approaches such as the Hobbs algorithm, have also been attempted for early implicit role resolution systems (Iida **et al.** 2006; Converse 2005).

"Activation"-oriented models estimate a score of a referent as an approximation of some psychological characterization, such as activation in memory. A simple version of this kind of approach is used in Laparra and Rigau 2013, which initialize each mention in terms of an initial score (determined by grammatical role, etc.) and then decrease that score by a small increment after each sentence, using a damping factor α. This score was calculated as :

$$score = initialized - score - 100 + 100 * (\alpha^{\text{sentence-distance}})$$

More complex models for referring expression form have been used with rigid predictions for referential activation, as in Vogelzang et al. (Vogelzang **et al.** 2015), which implementing the referential status using the ACT-R model of activation (Van Rij **et al.** 2013; Anderson **et al.** 1997). However, these measures of activation are often somewhat ad-hoc. A measure that is often more ideal is an approach in which one doesn't simply guess regarding referential status, but builds a model which predicts some latent status of the antecedent, where that status predicts the referential choice – whether the mention will be pronominal or explicit. Data-oriented models have been proposed(Grüning & Kibrik 2005; Khudyakova **et al.** 2011; Kibrik **et al.** 2016), as well as the Bayesian model of Orita **et al.**(2014), and we follow those models in using data-oriented approaches.

Following these data-oriented approaches, we develop an actual model of the salience of an antecedent by assuming that highly salient antecedents will be realized as pronouns, and that a pronominal mention will refer to the most salient antecedent. Conveniently, that means that such models can be trained upon any datasets in which there are coreference annotations, such as the Ontonotes coreference corpus (Pradhan **et al.** 2011). For each instance of pronominal anaphora, the last mentioned antecedent of that pronoun was used as a training example, and all prior competitors were added as negative examples – essentially treating this as a simple version of a coreference resolution task, in which only the salience information is available. We outline a simple feed-forward model which learns embedded representations of the headword, the part of speech of the headword, the dependency relation of the headword, the number of mentions within that mention's coreference chain, and whether that grammatical relation matches that of the current pronominal mention. These were concatenated into a single feed-forward network with one hidden layer and tanh nonlinearity. Table 5.7 illustrates the performance of this model on the simple task of predicting the correct prior mention using nothing but salience, with ablations for each of these features.

Model	Validation F_1	Test F_1
All features	47.9	48.7
All features except POS	43.2	44.6
All features except ref. distance	39.2	42.1
All features except coref. chain size	36.7	37.1
All features except dependency role	36.9	35.4
All features except gram. parallelism	40.8	39.1
Only distance and chain length	29.1	28.2

Table 5.7: Simple Models for Predicting Salience, with removal of various features; prediction is most impacted by the referential distance, dependency role of the antecedent, and size of the coreference chain it is in (1 for singletons)

The interesting outcome of these ablations is that no one characteristic is the single determining factor of the salience of a prior mention. This is in contrast with single-factor arguments, such as those that focus on the importance of grammatical parallelism (Carminati 2002) or those that focus upon a preference toward prior subjects (Crawley **et al.** 1990). This variable set of factors also supports arguments that many different factors are involved in the referential status of antecedents, rather than any one component.

5.2.6 A Model for Referential Status of Implicit Roles

The other important feature is the estimation of whether a particular implicit role is likely to be recoverable in context. This can be viewed in terms of the direct prediction of whether an implicit role is recoverable, or may be aimed at indirectly by training to predict FrameNet interpretation types (Definite Null Instantiation, Indefinite Null Instantiation, or Constructional Null Instantiation (Baker **et al.** 1998)), or even predicting the fine-grained implicit role interpretation types proposed for this thesis.

A major issue involved in these tasks is the extent to which one can learn to predict recoverability (or the interpretation type) of a possible implicit role by looking at the semantic role itself, or whether that interpretation is construed by particular syntactic constructions. Discussions of implicit role constructions of linguistics – as provided in Chapter 3 – illustrate that one might generally expect that both lexical (or frame-semantic) information will be expected to help in general, and that for verbs, one might expect syntactic constructions to also be very relevant, as syntactic representations can modify expectations of arguments and valency. In particular, Ruppenhofer & Michaelis (2014) suggest that lexical and frame-semantic generalizations over the semantic roles themselves will provide an important factor in determining the interpretation type of referents, and suggest that tendencies provided by definiteness tendencies over those frames may be a signal to such biases.

Various models have approached these tasks with a variety of syntactic and semantic features. Cheng & Erk (2018b) and Gerber & Chai (2012b) presented recoverability modles (sometimes called "fill or no fill" models) to determine whether to use an implicit role was recoverable for the Beyond NomBank data. Tonelli & Delmonte (2011) predicted the FrameNet DNI and INI types instead, using FrameNet specific information such as the FrameNet core/periphery distinction, and the likelihood of a given argument being definite in data (having some similarity to the proposal in Ruppenhofer & Michaelis (2014) regarding a link between DNI/INI tendency of a semantic role and the definiteness of explicit arguments that fill that role).

For learning these types, we used a combination of the features proposed in prior work on FrameNet DNI/INI detection (Tonelli & Delmonte 2011), with a simple neural network designed to characterize each implicit role in context, using the three different tasks (recoverability, DNI/INI/CNI detection, and implicit role type detection) in multi-task learning. There are four ways of characterizing the implicit role itself – either as a simple numbered argument such as “arg0” (NARG), with the addition of a predicate-specific argument such as “arg0 of lose.02” (PB), with the addition of lexical resources such as VerbNet and FrameNet (LEX), with the addition of information regarding how often the semantic role is explicitly realizied (EXPL), and with the addition of a bias about how often the semantic role is definite, when explicitly realized (DEF). Table 5.8 illustrates models with various amounts of this semantic informaton about the role, and one can see a meaningful improvement in recoverability when using richer resources.

There are also many ways of looking at the syntactic context determining the implicit role. We can consider no context at all (NO-SYN), can consider only the part of speech and dependency role of the predicate (PRED), can consider the constructional situation of the predicate, as defined by a sequence of the dependency roles and parts of speech of its dependents (DEP). Finally, becuase abstract syntax may omit information which can be aquired from a representation of words in context, we also utilize a richer representation based upon an LSTM applied to the whole sentence context (ALL), in which a vector representation of each semantic role was learned as a query and then used to key-value attention over an LSTM representation of the sentence. One can see in Table 5.8 that the addition of more syntactic information does improve the prediction of recoverability. In partciular, one can see a larger impact of syntax when predicting the recoverability of verbal arguments, as one might expect from the stronger valency restrictions provided by verbs.

For all of these, the result was passed through a shared feed-forward layer, and then different prediction weights used for each different type in the last layer. All results assume multi-task training, alternating between instances of each task. For FrameNet type detection, all FrameNet data was used; for recoverability, 10% of the AMR training data was used. "Recoverable" mentions were upsampled to be one third of the recoverability data.

The score produced by such a model was then used by the models presented below, in order to determine whether a particular mention is recoverable. However, it should be noted that full implicit role models have access to a third piece of information important for this task, as these models are independent of the *referent* of that implicit role. We assume that both for computational purposes and also for linguistic analysis, the constructional information provides some preliminary biases about whether an implicit role will be recoverable, and what its interpretation type will be, but that a final decision is best made in context. We note that it is also extremely curious how dependent the prediction of interpretation types and FrameNet information seems to be, in this formulation, upon lexical and semantic information rather than syntactic cues, and more exploration is required to understand whether that finding generalizes beyond the currently-presented models.

syntax	semantics	Ch. 2 types	DNI/INI/CNI	Recoverability
no syn.	NARG	44.94	83.30	63.02
no syn.	PB	**56.2**	90.19	65.8
no syn.	PB+LEX	56.18	90.26	65.57
no syn.	PB+LEX+EXPL	55.06	**90.34**	66.40
no syn.	all (above +DEF)	53.93	90.26	65.36
pred.	all	53.93	89.73	66.67
pred+DEP	all	52.81	89.36	67.33
pred+DEP+LSTM	all	53.93	87.86	**70.91**
No Multi-task: all	all	47.19	88.68	72.69
Verbs Only				
no syn.	PB+LEX+EXPL	59.09	**90.59**	59.72
pred.	PB+LEX+EXPL	**61.36**	90.29	60.17
pred.+DEP syntax	PB+LEX+EXPL	56.82	89.67	**63.67**
pred.+DEP+LSTM	PB+LEX+EXPL	56.8	89.88	61.89

Table 5.8: Multi-task learning models for predicting referential status, illustrating impact of both constructional and lexico-semantic information about the implicit roles in context

5.3 Implicit Role Labeling Models

The models outlined in the previous section each provide specific models for approximating the kind of underlying information we assume to be required for implicit role resolution. However, this leads to the natural question of how far we can get with such a set of separated features for modeling this data, and whether there are kinds of information about implicit roles which go beyond our current characterizations. The current section presents both a simple and somewhat interpretable model using these proposed features, and larger models attempting to leverage anything available from available surface features.

5.3.1 Background of Existing Implicit Role Labeling Systems

All models for handling implicit role resolution (or "iSRL") could be viewed as applying some kind of scoring function between an implicit role and a set of antecedents, sometimes poorly scoring antecedents along the way.

Early rule-based models used combinations of rule-based ranking heuristics – such as lists of grammatically salient recent mentions – and ruled out invalid candidates using hard constraints such as selectional restrictions. The first of these, PUNDIT (Palmer **et al.** 1986; Dahl 1986), maintained "focus lists" of recent mentions and used rich lexical information for filtering possible referents. Other rule-based models used variants with Centering Theory (Walker **et al.** 1994), used heuristic searches through the prior syntactic and discourses structure such as the Hobbs algorithm (Converse 2005) or combined heuristics about semantic role matching with cognitive estimates of referent activation (Laparra & Rigau 2013). However, as many of the implicit role phenomena in general English are pragmatically defined implicit roles (such as arguments of nominal predicates), it is unlikely that one could model such issues with rules alone.

These were followed by statistical models that were trained on implicit semantic role data, primarily using the nominal annotations of Beyond Nombank (Gerber and Chai 2012) or Semeval data (Ruppenhofer **et al.** 2010a), using large sets of feature for dense annotation (Iida **et al.**

2007b; Tonelli & Delmonte 2011; Silberer & Frank 2012). Most define a wide range of features for each possible antecedent and how those antecedents combine with expectations from the implicit role, such as characteristics of the antecedent (NER tags, POS labels, headword, events that it participates in, etc.), characteristics of the implicit role and its predicate, and interactions between those two feature sets. These can learn evocative but extremely domains-specific; for example, Gerber and Chai(2012) reported that the most important feature they found was (word, arg_n, predicate) triples, such as “oil & arg1 & price”. A set of models trained upon the NomBank data with such approaches (Gerber & Chai 2010; Gerber & Chai 2012a), with others taking similar approaches to the FrameNet SemEval or ONV5 datasets (Tonelli & Delmonte 2011; Silberer & Frank 2012; Moor **et al.** 2013a; Gorinski **et al.** 2013). However, especially when these models are being developed for very restricted domains and small sets of predicates, it can be unclear how much feature engineering for these corpora actually generalizes into better understanding of the larger task.

The most recent approaches to implicit role resolution (Do **et al.** 2017; Cheng & Erk 2018b) build end-to-end systems trained upon explicit language data trained on large corpora. The simplest form of this is the simple learning of vector representations for the sake of selectional preferent for iSRL (Schenk & Chiarcos 2016), but others focused upon rich neural networks that can learn to do SRL or coreference-related tasks which can transfer into the implicit role labeling task. Do and Bethard (2017) developed a model focused upon selectional preference – reading in a sequence of explicit (or predicted implicit) arguments and making a language model prediction about the next word-argument pair, so that the model would start an LSTM reading “sale:pred , a0:company , a1:units ” and predict the next word-role pair such as “a2:buyers” or “a2:investors”; it seems likely that such a model essentially learns to do selectional preference in context, but with a focus upon making this directly useful for the NomBank task. Cheng and Erk (Cheng & Erk 2018b; Cheng & Erk 2018a) also built an end-to-end system (albeit with far more components) giving a score to a mention using explicit data, but also used components which could leverage information from other information in the document, providing the possibility of learning script-like information –

either by using a single "context" event alongside the event under question (scoring event-argument pairs using negative sampling) (Cheng & Erk 2018b), or by applying multi-hop attention over the prior events of the document (Cheng & Erk 2018a). We assume that systems such as these – learning a single task over large amounts of explicit SRL or dependency data – are likely to be the actual future of implicit SRL models, but that it is hard to characterize what kinds of knowledge are actually being learned by such systems. These approaches also resemble recent deep approaches to iSRL in other languages such as Chinese (Chang **et al.** 2017; Yin **et al.** 2016; Yang **et al.** 2019), where there is sufficient training data for models to be trained upon the implicit role resolution information alone.

5.3.2 Implicit Role Resolution Models

In order to attempt to provide slightly more illuminating analyses of the impact of various components, we present both a simple "interpretable" model aimed at keeping separate the contribution of different implicit role features, as well as denser, deeper models which can give some representation of the improvement gained by access to the full surface forms. Perhaps unsurprisingly, we will find that simple approximation of the core features of implicit role behavior to be not sufficient, in their current form, to exceed the performance of models that also have access to raw text.

5.3.2.1 Model 1 – Interpretable Model

The "interpretable" model aims to make predictions which are decomposable into the contribution of these individual models (selectional preference, narrative schema information, mention importance, salience, and deixis), such that one can separately consider the contribution of each part. We use a simplistic model that takes a (potentially weighted) average over each of these scores, applying a minimal non-linear transform to each score first. For each implicit role linking decision, we assume that score for each pairing of the antecedent and the implicit role (optionally considering a set of additional features such as referential distance or the number of intervening entities, ϕ). If we view each feature like "selectional preference" as a function f from a list of M different little models, we might initially represent the score for any antecedent as the sum of the predictions from these models, i.e.:

$$s(a_i, r, \phi) = \sum_{k=1}^{M} f_k(a_i, r, \phi_i)$$

Because the scores from those different models are not being optimized to be used for this current task, their correlation to the actual desired scoring function is not necessarily linear. We modify each with a tiny feed-forward neural network (with a single input, a layer of hidden units (we use 7 in the reported models), and a single output), roughly following work on generalized additive neural networks (De Waal & Du Toit 2011). This leads to the following model:

$$s(a_i, r_j, \phi) = \sum_{k=1}^{M} \mathrm{FFNN}(f_k(a_i, r_j, \phi_{i,j}))$$

The most direct way of adding different "implicit role interpretation types" to this model is to re-weight these predictions based upon different implicit role constructions. For example, an "implicit subject" construction may pay more attention to salience factors, whereas a nominal argument may have more of a bias towards factors such as selectional preference and script information. We define a simple list of implicit role constructions (described later in section 5.5.0.2) and learn an embedding of weights for each construction, in order to learn weights for these types.

$$s(a_i, r_j, \phi) = \sum_{k=1}^{M} w_{construction(r_j)} \mathrm{FFNN}(f_k(a_i, r_j, \phi_{i,j}))$$

The primary value of adding such weights is not simply predictive, but rather as a way of studying how individual constructions and reference contexts – such as implicit subject constructions – align with particular underlying features. The study of the individual weights learned by each given grammatical structure, discussed in the analysis section later in this chapter, illustrates the kind of biases that this model learns. This approach is somewhat similar to the recent work of Zhang **et al.** (2019), who look at how to weight different information sources for pronoun resolution

– using selectional preference, plurality agreement and gender and animacy agreement – but they use a full feedforward network with mention context to determine the weights.

For that simple interpretable model, we then predict whether an implicit role actually occurred using the predictions from the separate models for recoverability discussed above, along with predictions for FrameNet DNI/INI detection, and interpretation type predictions, fed into a single linear predictor. As one additional issue with implicit role detection concerned the low rate of recoverability (particularly within the MS-AMR corpus wherein less than 10% of MS-AMR possible implicit roles are actually recoverable), we deal with that data imbalance by proportionally reweighting the importance of recoverable instances (Morik **et al.** 1999), so that the loss of missing a recoverable instance is much greater than the cost of a false positive, weighted according to that recoverability rate. More details for the model and its parameters are included in Appendix A.2.1

5.3.2.2 Model 2 – Simple Dense Features

We also implemented a system with more traditional syntactic and semantic features (Gerber & Chai 2012a; Silberer & Frank 2012) – such as the part of speech of the antecedent, whether the numbered argument of the implicit role matches any roles of the antecedent, and even sparse features such as the headword of the antecedent. The results reported below refer to this version as the DENSE model. These are hashed into a fixed-width array, combined with the outputs of the individual feature models and features from the recoverability model, and then passed through two layers of a feedforward network heavily regularized with dropout, producing a single score for each possible antecedent – therefore having the ability to memorize specific patterns in the data, and to gain from interactions between features.

In this dense model, also predict recoverability using both the outputs from our external model (predicting recoverability, DNI/INI and fine-grained interpretation type predictions), but also with simpler features which the model may learn from, such as the implicit role itself, its syntactic context, the presence of quantifiers on the predicate, and more. Those features and the recoverability model predictions are concatenated, and the result is then passed through a feedforward neural network (regularized with Dropout) for a prediction regarding recoverability. More details for the model and its parameters are included in Appendix A.2.2

5.3.2.3 Model 3 – ELMO model

We present a third model oriented towards directly learning representations, to evaluate whether the existing implicit role data is sufficient to learn a relatively deep model. This model evaluates mentions using a preliminary scoring metric, selects a top set of candidates, and then goes beyond simple representations of heads out of context, modeling those top candidates using self-attention over the entire span and using pre-trained representations of word meaning in context (Peters **et al.** 2018).

The preliminary scoring metric starts with a single vector $\mathbf{g}_r$(i) characterizing a given implicit role *j*, and then for each candidate antecedent *i*, a vector representing that mention $\mathbf{g}_m$(j) , embedding characteristics such as NER label, POS and headword, and finally a vector of additional features (such as distance, selectional preference, etc.) ϕ(i,j). We use a dot product comparison between the vectors for the mention and implicit role, and combine that result with a learned score using those other comparison scores, calculating the preliminary score as follows, where · represents dot product and FFNN a feed-forward neural network:

$$s_{preliminary} = FFNN(\mathbf{g}_m(i)) \cdot FFNN(\mathbf{g}_r(j)) + FFNN(\phi(i,j))$$

We then prune all but the K highest-scoring possible antecedents. Each of these more likely candidates is then characterized using a more computationally expensive approach, in which a representation of the sentence in context is generated (using the ELMO representation of the sentence (Peters **et al.** 2018), followed by one layer of trainable LSTM weights) and then the mention span is merged into a single vector, by generating a self-attention scalar for each token within the mention span, and taking a weighted average of those tokens in the span, following the approach of Lee et al. (2017) for coreference resolution. This self-attention mechanism is generally considered to learn a soft version of syntactic headedness, so that the resultant characterization of the mention provides a robust representation of the mention in context. This follows the equations below, wherein x_i represents the sentence corresponding to a particular mention, $\hat{\mathrm{x}}_t$ the representation of each sentence t in context, and $\mathrm{a}_{i,t}$ the attention score corresponding to each mention:

$$x_t = \textsc{lstm}(\textsc{elmo}(sentence_t))$$
$$\alpha_t = w_\alpha \textsc{ffnn}_\alpha(x_t)$$
$$a_{i,t} = \frac{exp(\alpha_t)}{\sum_{k=start(i)}^{end(i)} exp(\alpha_k)}$$
$$\hat{x}_i = \sum_{t=start(i)}^{end(i)} \alpha_{t,i} \cdot x_t$$

We use this process both to characterize both each mention span, and the implicit role span, and then re-calculate compatibility between each candidate and the implicit role, with the addition of these in-context representations:

$$s_{final} = s_{preliminary} + \textsc{ffnn}(SA_m(i); \mathbf{g}_m(i)) \cdot \textsc{ffnn}(SA_r(j); \mathbf{g}_r(j))$$

The estimate of whether the implicit role was *recoverable* was also estimated using features similar to those used in Model 2, along with the score and representation of the highest-scoring referent. This model, although likely overly complicated for the quantity of annotated available, was able to learn general representations of the iSRL task, trainable across all datasets, and might be rich enough to pre-train using explicit SRL (as done with the multi-hop attention model of Chang and Erk (2018b)). More details for the model and its parameters are included in Appendix A.2.3

5.3.2.4 Preprocessing, Training, and Optimization

For each dataset – Nombank, MS-AMR, ONV5, and Semeval – we preprocess raw input text (using no gold information except tokenization) using the AllenNLP toolkit implementations of coreference, SRL and dependency parsing systems (Gardner **et al.** 2018; Lee **et al.** 2017; Dozat

& Manning 2016; He **et al.** 2017). SRL was modified to include nominal and adjectival predicate detection (following the deterministic rules used in PropBank annotation) and re-trained upon the BOLT discussion forum text (Song **et al.** 2014) for AMR and Semeval data, and upon a mixture of Ontonotes WSJ, the WSJ financial subcorpus, and NomBank, to achieve some approximation of the ConLL 2009 data with modern PropBank annotation. All other Allennlp tools used the base models from Allennlp v0.5.0. Mentions were selected using recall-focused heuristics (essentially capturing all nouns, referential adjectives and verbs) and passed through the feature models outlined above. Those outputs were provided to the second model making the implicit role resolution predictions.

This model (using gold implicit role detection) then examined all implicit roles with antecedents, scoring every candidate antecedent, and was trained to maximize the likelihood of a correct antecedent for each implicit role, using Adam (Kingma & Ba 2014) and the AllenNLP toolkit (Gardner **et al.** 2018). The training process was always halted when accuracy on the validation set failed to increase for a number of epochs, and the best-performing model on the validation set was used.

5.4 Evaluation of iSRL systems

5.4.1 Full Implicit Role Resolution

Table 5.9 shows the performance of these models over the full task of implicit role resolution. The performance upon existing datasets – especially the Beyond NomBank data – are not as impressive as the current state of the start. However, they provide the first models trained upon the challenging MS-AMR dataset, and are the second model (after Feizabadi and Pado 2015) to develop systems designed to handle multiple implicit role corpora, rather than a single corpus. For all models reported in Table 5.9, the models were trained upon either the individual corpora listed in the "trained on" column, with "all annotated" referring to the set of all manually annotated implicit role training corpora (Beyond Nombank (Gerber and Chai 2010), Semeval-2010 (Ruppenhofer et al. 2010), ONV5 (Moor et al. 2013) and MS-AMR (O'Gorman et al. 2018)), and that set was sometimes supplemented with the addition of the converted within-sentence AMR corpus ("WS-AMR") discussed in Chapter 4.

Unsurprisingly, the datasets trained upon NomBank do not transfer well to MS-AMR, and vice versa. Moreover, the MS-AMR corpus prediction performance was quite low, regardless of the models used. As the current state of the art of the other wide-coverage corpus, SemEval-2010-10, is only 18.0, and the MS-AMR dataset poses new challenges due to the very low rate of recoverable implicit roles, it is likely that the MS-AMR implicit role resolution task will be similarly difficult.

In looking at differences between various models tried, one can see that for MS-AMR and SemEval-2010 datasets, simply increasing the amount of data used for training generally helps with model performance, and the deeper ELMO model seems to perform better than simpler approaches. However, since all models are outperformed dramatically by other models when looking at the Beyond NomBank data, any generalizations to be made about model performance are limited.

5.4.1.1 Linking-only Results

Especially when considering datasets such as SemEval and ONV5, it is common for models to only evaluate their model performance upon the subtask of implicit role "linking", wherein one assumes gold detection of which implicit roles are recoverable, and only evaluates whether a system can discern the correct referent. Table 5.10 illustrates the performance of these models upon this linking task. This potentially shows instances in which the added data of MS-AMR and the within-sentence AMR conversions make an impact, as we can see improvements in accuracy over SemEval-2010 data (which has often been reported as a linking-only metric) – training a model upon all of the datasets, and fine-tuning upon MS-AMR, shows higher scores than previously reported. However, both here and with the scores listed above, it should be noted that the scores for SemEval-2010 are not directly comparable to prior work, because of the conversion from SemEval into PropBank-style annotation, which omitted some non-eventive frames (such as "Tuesday" as Calendric_Unit frame or "House" as a Building frame).

Model	Trained on	NomBank	Semeval	ONV5	MS-AMR
Baseline	(with current system)	11.03	1.74	9.9	1.43
Model 1 (Interpretable)	NomBank	16.8	11.75	8.69	4.04
Model 1 (Interpretable weights)	NomBank	13	18.06	7.4	3.70
Model 2 (Dense)	MS-AMR	3.65	4.69	0(0)	5.32
Model 2 (Dense)	NomBank	22.82	2.75	4.65	0.88
Model 2 (Dense)	All annot.	20.94	**12.00**	3.33	7.53
Model 2 (Dense)	All annot. + WS-AMR	19.9	10.82	7.4	7.62
Model 2 (Dense)	All annot., finetune on each	21.28	7.50		**8.25**
Model 3 (with Elmo)	NomBank	**27.6**	3.1	0.0	3.9
Model 3 (with Elmo)	all, fine-tuned on each	31.0	5.3	0.0	8.08
Laparra and Rigau (2013)	NomBank	45.8			
Laparra and Rigau (2013)	SemEval		14.0		
Feizabadi and Pado (2015)	Semeval+NomBank	21	18		
Gerber and Chai (2010)	NomBank	42.3			
Schenk & Chiarcos (2016)	NomBank	32.5			
Do et al. (2017)	NomBank	**46.1**			
Cheng & Erk (2018b) (GCAuto)	NomBank	44.5			
Cheng & Erk (2018a)	NomBank	**49.6**			

Table 5.9: Score on full task, including role detection, using F1 with partial spans defined in Ruppenhofer et al. (2010)

5.4.2 Resolution Performance At Different Referential Distances

An additional question regards the performance of implicit role models as the distance from the implicit role varies. Using the models presented above, we measure the performance of these models at different distances between the implicit role and an antecedent. Figure 5.2 shows the recall of each of these models, with 0 being in the same sentence. This shows that while the performance upon the Beyond NomBank newswire data is much higher in within-sentence contexts, all corpora observe relatively similar prediction scores on long-distance implicit role resolution. Such an observation may also be important as existing within-sentence explicit SRL models (Zhou & Xu 2015; Ouchi **et al.** 2018; He **et al.** 2017) might be easily extended to also predict SRL arguments

Model	Train on	Nombank	Semeval	MS-AMR
Interpretable	Nombank	15.1	17.33	5.03
Interpretable-weights	MS-AMR	12.05	16.45	17.88
Interpretable-weights	MS-AMR	12.05	16.45	2.99
Model 2 (Dense)	MS-AMR	13.19	17.35	12.84
Model 2 (Dense)	NomBank	29.55	11.23	3.75
Model 2 (Dense)	All annot.	29.2	22.6	25.45
Model 2 (Dense)	all 5, fine-tuned on NomBank	30.17	19.75	9.76
Model 2 (Dense)	all 5, fine-tuned on MS-AMR	33.15	**40.35**	**29.03**
Model 2 (Dense)	all 5, fine-tuned on SemEval		28.67	
Model 3 (Deep)	all 5, fine-tuned on each	**37.3**	15.0	19.3
Silberer and Frank (2012)			27.7	
Gorinski et al. (2013)			25.0	
Schenk and Chiarcos			26.4	

Table 5.10: Score on linking-only task; SemEval scores not directly comparable due to conversion to PropBank

for within-sentence SRL, it is useful to get a sense of how difficult the remaining cross-sentence task would be.

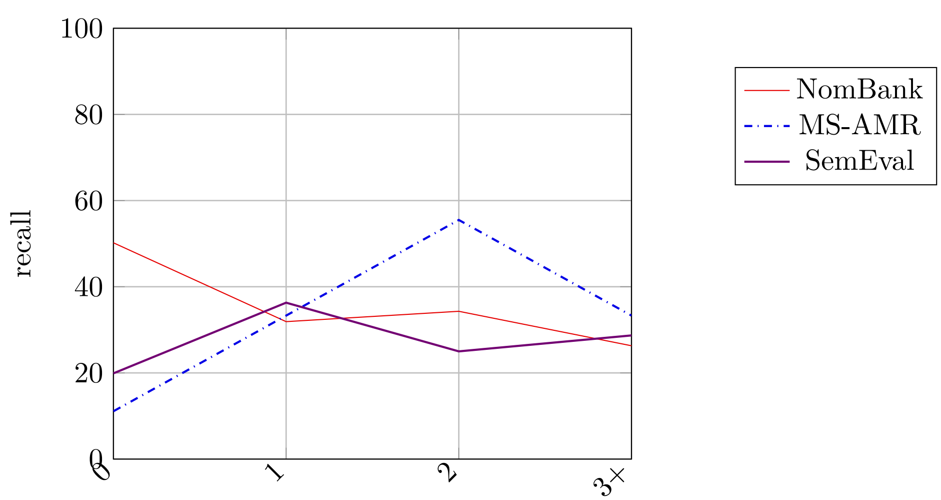

Figure 5.2: Scores based upon distance to antecedent, with 0 being in the same sentence (distance calculated on the gold data)

5.4.3 On the Difficulty of Replicating Beyond NomBank scores

One issue with the current results is the low performance of these models upon the Beyond NomBank dataset. Although there are mitigating factors in comparison to some of the prior models (many prior works, such as Gerber and Chai (2010, 2012) and Laparra and Rigau (2013), used gold parse and SRL features), this poor performance is nevertheless a worrisome issue.

One measure which may illuminate this issue is the baseline proposed by Gerber and Chai (2010) (in line with the high amount of REMEMBERED EVENT instances seen in the Beyond NomBank data), wherein one simply looks for prior versions of the same semantic role mentioned in the text, and links to that role when seen. Gerber and Chai define this as "Fill $iarg_n$ for predicate instance p with the nearest constituent in the two-sentence candidate window that fills arg_n for a different instance of p, where all nominal predicates are normalized to their verbal forms." The models presented here, running that same baseline, find a score of 10.6 on the Beyond NomBank test set (precision of 29.6 and recall of 6.5), which is much lower than that seen with Gerber and Chai (2010), who report a 26.5 f_1 using this baseline using gold trees, gold NomBank SRL, and gold PropBank SRL. That score with gold pre-processing is, curiously, itself lower than the 28.0 f_1 of Do et al. (2017) against the same baseline using automatic verbal predicate predictions from MATE (Björkelund **et al.** 2010).

This suggests one possible culprit for low scores upon this model; that either the underlying SRL systems used in the current model (potentially in combination with candidate selection) is much less effective than that of gold SRL and parse selection, or that characteristics of the evaluation setup differ between different models. The difference between this reported baseline vs. the gold systems would hint that errors caused during automatic SRL labeling (or candidate selection) lead to this discrepancy, suggesting the need for better SRL on nominal instances, or a focused use of the in-domain financial Propbank data. However, we should admit that the curiously high baseline of Do et al. (2017) does not support that characterization.

A second possible reason is that this discrepancy could be due to simple differences in evaluation or data preparation. The NomBank task has not been evaluated as a shared task, and the data release format is ambiguous enough that research systems have many degrees of freedom for evaluation decisions. For example, incorporated arguments are included as normal iSRL instances in the data, but there is a note on the website which instructs users to remove them. There is also ambiguity regarding how one should model negative or "candidate" instances of implicit role data, which can be complicated if one is not assuming perfect knowledge of explicit SRL arguments. This baseline may also be very uniquely tuned to the NomBank data – evaluating it against the MS-AMR data reveals a baseline of 1.44 f_1,

It is hoped that the possible issues regarding evaluation and data preparation can be addressed by having a single repository in which the current datasets are represented in as simple of a format as possible (pointing to spans within the text, rather than phrases within parse trees) and where the assumptions about possible "candidate" implicit roles are made explicit. By standardizing assumptions about these datasets, it hoped that we could simplify the actual research cycle regarding implicit roles, and therefore potentially resolve such questions in the future.

5.5 Analysis

5.5.0.1 Ablation Tests

Using that "interpretable" model, we ablate each component in order to characterize the impact of each factor upon the model behavior – removing the feature from the underlying model and retraining a system without access to that feature. As these models do not have access to surface-form features, they cannot attempt to approximate the feature otherwise, therefore giving us some representation of the actual contribution of those features to the result. Table 5.11 illustrates the behavior of these models over both the validation and the test sets.

Selectional preference and salience consistently show contributions to the current models. In contrast, there is a great deal of variance in deixis, mention, and script/narrative schema preferences. This supports the vein of recent works focusing upon the relevance of selectional preference

Corpus	without ablation	-sel. pref	-salience	-deixis	-mention importance	-script info.
NomBank test	18.29	13.67	13.14	17.6	19.4	16.25
	Δ	-4.62	**-5.15**	-0.68	+1.11	-2.03
NomBank val.	26.1	22.9	19.5	24.5	25.7	29.99
	Δ	-3.2	**-6.6**	-1.6	-0.4	+3.9
MS-AMR test	17.88	12.84	11.9	13.2	8.55	17.2
	Δ	-5.04	-5.94	-4.68	**-9.32**	-0.68
MS-AMR val.	17.6	13.2	12.0	21.6	16.0	13.6
	Δ	-4.4	**-5.6**	+4.0	-1.6	-4.0

Table 5.11: Accuracy Scores when each feature is removed from the model. For each model, the feature whose removal most impacted the model is bolded.

models for implicit role resolution (Le & Fokkens 2018; Do **et al.** 2017), and recent works finding difficulty in establishing a clear impact of narrative schema information for coreference resolution (Pichotta 2017).

5.5.0.2 Learned Weights of Constructions

The "intepretive+weights" model discussed above re-weighted each of these factors according to weights learned from an implicit role context. We used a small set of possible contexts which might loosely characterize the syntactic context of each implicit role. We use statistics from OntoNotes (Weischedel **et al.** 2011), combined with the syntactic situation of a predicate, to determine clear-cut implicit role constructions – primarily main-clause implicit subject constructions (where an argument traditionally realized as the subject is implicit, and the verbal predicate has no subject), or implicit object construction. For nouns, adjectives, and oblique verbal arguments, these constructions must are defined in a more semantic manner, such as the "agentive argument of adjectival predicate" or "goal argument of verbal predicate".

One goal of having a model in which different features are weighted by particular constructions is to provide some kind of insight into the interaction between general pragmatic factors and the local implicit role construction context. By doing so, we can examine the weights given to each score as to what one might expect from a model is learned. Some distributions of these weights are shown in Table 5.12. One can see that implicit subjects of embedded subjects (as in pragmatic control contexts) place a strong emphasis upon salience, being biased towards a main-clause subject. In main-clause implicit subjects, the salience score was surprisingly less prominent, but this was perhaps because these models emphasized deixis instead, as English implicit main-clause subjects are very often deictic. We can also see from these examples that arguments of nominals, as one might expect, would learn to put more weight upon general pragmatic and script factors, as do the weights for English implicit object constructions. These support the characterization from Chapter 2 that these may be generally thought of as SCRIPT-INFERRABLE implicit roles.

Corpus	construction	Sel. pref	Script info.	Salience	Deixis	Mention Referentiality
MSAMR	Implicit Subject (embedded)	17.6	41.3	26.2	12.2	2.7
MSAMR	Implicit Subject (main)	1.2	25.7	6.4	26.2	40.4
MSAMR	Nominal patient/ experiencer	20.1	15.5	3.0	57.5	3.9
Nombank	Nominal agent/ stimulus	3.7	47.9	8.7	23.7	15.9
Nombank	Nominal Goal	19.8	21.5	20.0	36.2	2.5
MSAMR	Nominal agent/ stimulus	8.1	66.8	8.4	22.8	1.2
MSAMR	Verbal Goal	32.7	15.1	16.6	8.4	27.2
MSAMR	Implicit Object	7.8	74.1	5.2	5.6	7.3

Table 5.12: Weights learned by a simple implicit role construction

While the performance of the underlying models is not sufficient enough for such weights to be viewed as strong evidence for how the actual implicit role constructions behave, this illustrates a simple way of approximating how these different implicit role constructions may be behaving in a given context, and of exploring the biases of the implicit roles of a given language.

5.5.0.3 Manual Re-examination of Implicit Role Predictions

Finally, we attempt to explore not simply the effectiveness of current ways of approximating a feature, but also looking at what the biggest gaps are: which components would help the most if given human-grade judgments. We suggest that we cannot really get insight into those questions with ablation alone, as a poor approximation of a particular kind of feature will not provide a proper representation of how it might contribute to a task.

To gain some insight into this, we started with outputs from trained models, so that one might look at a set of the top K candidates and rank those antecedents according to a particular feature – such as how well a given candidate matches the current implicit role for a "selectional preference" scores. Since these models are not perfect, some of those rankings will be clearly wrong. We provide a little interface where a human annotator might correct such errors directly – by re-ranking the possible antecedents in terms of how well they should score along one of those particular features. Figure 5.3 illustrates this annotation (albeit in a more compact and readable form). In this approach, one might see in the second column a ranking of antecedents for the "buyer of sell.01" role, and start with ranking them with "company" highest and "buyers" lowest, and the human annotator would re-rank them so that "buyers" would be the highest scored element, and "portion" the lowest.

The useful thing about such an annotation is that we might view the impact upon the data as one applies corrections to each feature (inspired by error analyses of SRL by He **et al.** (2017, who are able to study the impact of various deterministic corrections). Figure 5.4 illustrates the effect of progressively adding these corrections to adjust the data, until all corrections are made. Such a representation is purely exploratory but suggests that there may be some room in these models for better models of salience, and a very large amount of additional value in the "narrative schema/script information" feature.

A final correction "optimal weights" provides the score if one could re-weight the features arbitrarily – largely measuring whether a particular argument is high-scoring in one or two features, but outscored by something else, such as something plausible being overwhelmed by a more topical and recent candidate. It is also not surprising that this would be a large score, and largely supports the complexity of balancing different factors, and supports directions (such as end-to-end models) in which one might automatically learn to accommodate different factors.

Ports of Call Inc. reached agreements to sell its remaining seven aircraft to buyers that weren't disclosed. The agreements bring to a total of nine the number of planes the travel company has sold this year as part of a restructuring. The company said a portion of the $ 32 million realized from the sales (arg2 "buyers" of sell.01) will be used to repay its bank debt and other obligations resulting from the currently suspended air-charter operations . Earlier the company announced it would sell its aging fleet of Boeing Co. 707s because of increasing maintenance costs .

Narr.Schema	Selpref	Salience	Mention Ref.	Deixis
buyers ↓↑	company ↓↑	Portion ↓↑	Portion ↓↑	buyers ↓↑
company ↓↑	Portion ↓↑	company ↓↑	company ↓↑	company ↓↑
Portion ↓↑	buyers ↓↑	buyers ↓↑	buyers ↓↑	Portion ↓↑

Figure 5.3: Example of interface for annotating decisions by re-ranking candidates (moving them up an down the ranking list). As an example – one would hopefully view "buyers" as being a more appropriate fit, in terms of selectional preference, than "portion" or "company", and move it up

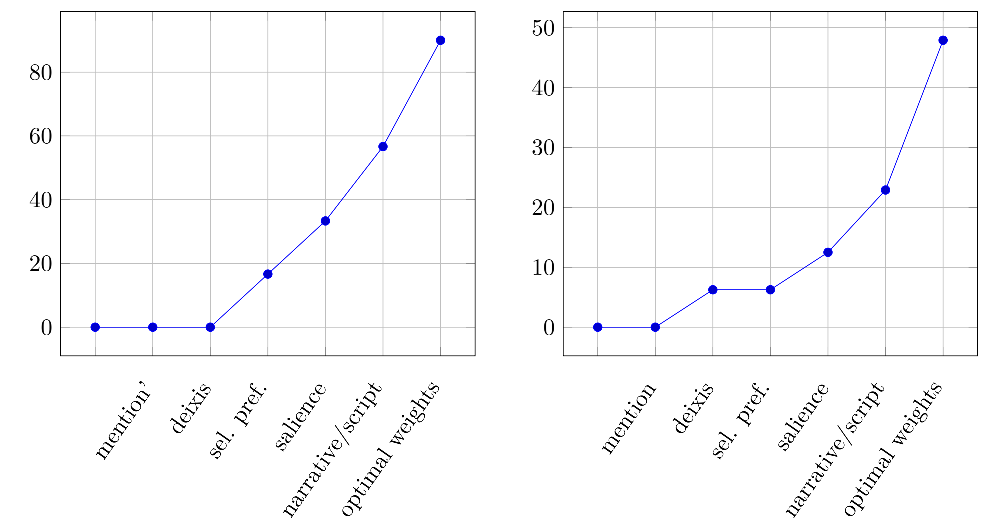

Figure 5.4: Manual correction effects on NomBank (left) and MSAMR (right) data, after human re-ranking of each set of scores (cumulative).

5.5.1 Discussion and Examples of Errors

The models discussed do correctly identify some implicit role antecedents. Examples 170 and 171 from Beyond NomBank and MS-AMR validation sets illustrate examples of the kind of implicit role antecedents which the models do identify:

(170) For third quarter last year , **[Keystone]**$_{predicted+correct}$ reported a \$ 1 million loss from continuing operations and a \$ 200,000 loss from discontinued operations , for a net **loss** $\emptyset_{loser}$ of \$ 1.2 million .

(171) And **[I]**$_{predicted+correct}$ do n't know what to do . Thanks for reading. It 's nice to **vent** $\emptyset_{venter}$.

However, there are far more examples of incorrect model predictions. For the Beyond Nombank, while current models often correctly identify the implicit roles, there are many resolution errors in which the model fails to obey the selectional restrictions of low-frequency implicit roles. For example, in example 172, the implicit role "arg3 of cost.01" refers to the buyer (cost to whom), and should be constrained to animate referents, but the model misses that and links to "A put option" rather than "its holder" below. This cannot be learned from general characteristics about "arg3" instances, but required the model to have acquired a specific model of the arg3 of cost.01, which seems to be missing in the models:

(172) [A put option]$_{predicted}$ gives [its holder]$_{correct}$ the right (but not the obligation) to sell a stock (or stock index) for a specified price (the strike price) until the option expires . Whether this insurance is worthwhile depends on the cost of an option. The **cost** $\emptyset_{to\ \ whom}$, or premium , tends to get fat in times of crisis .

There are also instances where the models find a correct class of referent, but refer to generic mentions rather than the particular version under discussion. For example, in a discussion of sales of a particular toy, a prediction model links to the generic "toy" sales, when the predicate specifically links to "the Ninja Turtles", which both require general document knowledge as well as Named Entity prediction (knowing that "Ninja Turtles" refer to a toy):

(173) Should **sales** $\emptyset_{commodity}$ continue to be strong through the Christmas season , which accounts for about 60 % of U.S. retail toy$_{predicted}$ sales , Mr. Kwan said the Ninja Turtles$_{correct}$ could make 1989 a record sales year for Playmates .

In the MS-AMR data, the more common issues are not such confusions, as the most common set of errors are those in which a non-recoverable implicit role is incorrectly linked to an antecedent.

For example, example 174 has an example of “practice” used intransitively (roughly “practice medicine”), but the model incorrectly attempts to link that “thing practiced” to prior text:

(174) No matter what, he never should have given Michael Jackson that propofol$_{predicted}$. The man **practiced** $\emptyset_{thing\ \ practiced}$ here for many years .

We similarly find many instances where nominal events which do not have a clearly stated agent in prior text are linked to possible agents, as in the agent of “relocation” below:

(175) But creating a balanced community with a mix of housing, offices, shopping and other amenities – allowing people$_{predicted}$ to live close to where they work and play – is an even more worthy goal. As Ed Risse has pointed out , the **relocation** $\emptyset_{arg0,agent}$ of 17,000 jobs creates a rare opportunity to create a genuine balanced communities in Northern Virginia.

Such examples illustrate the sheer difficulty of the task in wide-coverage implicit semantic role labeling – it is often quite difficult to point to simple generalizations which the models should have relied upon to correct a given decision. These models are therefore also quite poor at identifying the correct antecedents of roles which refer to propositions, as in the implicit themes of predicates such as “know”, “answer”. All such patterns suggest that the current models presented here may not be utilizing enough data to learn all of the rich information required for this difficult task.

5.5.2 Conclusions of Analysis

The limited accuracy of the models presented here (and indeed, of all current implicit role models) makes it difficult to make broad generalizations about the current state of computational implicit role resolution. However, we can draw some limited generalizations about the nature of the current models – and running models upon the new MS-AMR data – from the systems presented here.

One takeaway is that different datasets such as the MS-AMR, NomBank, and SemEval-2010 are relatively compatible after conversion to PropBank inventory – in the sense that one can train a shared model upon all of them, and that models trained on MS-AMR data can perform well (at least for the linking subtask) upon the SemEval data. This promise of a larger shared dataset, however, required further testing with models which are closer to the current state of the art. The analysis sections primarily also provide evidence supporting an idea that there is still a large amount of room to grow in terms of phenomena which might be characterized as "narrative schema" issues.

5.6 A Probing Dataset for Implicit Semantic Roles

Diagnostic classification or "probing" techniques (Tenney **et al.** 2019b; Tenney **et al.** 2019a; Ettinger **et al.** 2016; Adi **et al.** 2016; Liu **et al.** 2019) refer to the notion of taking deep models with rich, vector-based representations of a task, and testing whether those representations have learned a particular pattern or abstraction, by using the state of that richer representation as the input for a simple model trained upon an auxilliary task. For example, one may measure whether a rich representation of word meaning in context (Devlin **et al.** 2018; Peters **et al.** 2018) has "learned" part of speech representations of the words in context, by taking the vectors used to represent a particular word, and directly training a small (often linear) model to predict part of speech using that representation. We suggest that the data presented here – and the interpretation types as well – may be useful for building *probing* analyses of such information, helping to expose whether models are learning this kind of information.

The natural starting point for a probing model of iSRL data is to look at how probing approaches to coreference resolution and semantic role labeling have been implemented. Edge-probing tasks (Tenney **et al.** 2019b; Tenney **et al.** 2019a) have focused upon converting complex tasks of predicting relations into simple characterizations where two spans are examined out of context, and a model makes a simple (often binary) prediction using those two spans. As these tasks are often very context-depndent tasks, this therefore evaluates whether the models probed have already incorporated the necessary contextual disambiguations and connections into the underlying representations. Tenney **et al.** (2019b) evaluate the ability of systems to handle coreference (both from the Ontonotes corpus (Weischedel **et al.** 2011) and in Winograd schema (Levesque **et al.**

2012)) by breaking coreference decisions into simple pairwise decisions between two spans, and making a binary decisions about whether those spans are coreferent or not (wherein each span is reduced to a single vector via self-attention over the span (Lee **et al.** 2017)). Example 176 illustrates such a simplified form of the coreference task.

(176) The important thing about Disney_1 is that it_2 is a global brand. → True

Tenney **et al.** (2019b) also evaluate SRL using the same edge-probing methodology, by representing an argument and a predicate, and making a prediction of the semantic role label itself, as in example 177 below. One may note that this does not fully test whether a model is representing SRL information, but specifically evaluates the argument classification subtask of SRL.

(177) The important thing about Disney_2 is_1 that it is a global brand. → Arg1

Based upon such prior work, we develop – and will release – an annotation utilizing the implicit roles presented in this thesis, but oriented towards such probing tasks. This is not presented as a single task, but split into four tasks, defined by four interpretation types most commonly seen in recoverable implicit roles – those requiring *event coreference*, those that involve *deictic reference*, those that involve *larger script knowledge*, and those that involve *salience and other short-distance local phenomena*. We also utilize predictions from the models provided in the current chapter, in order to best convert this task into a binary probing task by providing challenging negative examples.

This data is provided with the intent that future work may actually implement probing models, and some decisions would need to be made by future work. However, we assume a tentative form for such a future probing task, in that this data cannot be directly represented as edge probing (simply looking at two spans), but instead as a labeled form of edge probing: a given instance is a pair of spans and a label characterizing the relationship, and the model then makes a binary prediction of whether that link is valid or invalid. This is therefore analagous to making the coreference prediction tasks discussed above, with the addition of the implicit role label. Do develop the four sub-tasks defined by the interpretation types proposed, we annotated all recoverable implicit role instances within the MS-AMR corpus training data, in order to provide valid instances of implicit roles. For negative examples of implicit roles, different paradigms for each type of task were used, designed to provide an accurate test of whether the underlying phenomenon is being learned. Table 5.13 shows the size of these derived probing datasets. Moreover, these pairs of spans assume representation across an entire document, and are often separate sentences; while this may be evaluated against models of meaning in sentence context, it is assumed that richer representations of word meaning in document context may show better understanding of these phenomena.

Interpretation type	Positive	Negatives	total
Script inferrence	814	1399	2213
deixis	458	920	1378
event coref	217	403	620
salience	120	400	520

Table 5.13: Derived datasets for probing analysis of implicit role data.

The subset of DEICTIC implicit roles refers to mentions of the speaker and addressee in texts. The negative examples from these deictic implicit roles are taken from instances of non-deictic implicit role resolution where a locuphoric pronoun (such as *I*, *you* or *we*) was given a high score as a candidate by Model 3 above. This means that models cannot rely upon selection artifacts (labeling all locuphoric pronouns as correct), and can test whether the representation being probed captures information necessary for deixis (such as grammatical mood, detection of mental state verbs and detection of other deictically appropriate contexts). The examples below illustrate such a task, wherein an actual probing analysis would need to simply predict the binary judgement that the first two examples are correct and the third set is invalid.

(178)

So I ask do you think drugs can be good for some people ? ?	ARG2	True
Not looking for support . Just stating my mind .	ARG0	True
what tests should I have them run ?	ARG0	False

For EVENT COREFERENCE instances, the true instances have antecedents which have been explicitly mentioned in prior context. Negative examples were sampled using the same implicit roles and predicates, but supplying high-scoring alternative candidates from the current models. Examples 179 illustrate examples of this task, which may be difficult to model with current representations, although some amount of the data may be predictable through selectional preference behaviors.

(179)

Thus , the New York Times poll of this week found that all voters , by a 66 to 26 ratio , support the federal requirement that private health care plans cover the full cost of birth control for female patients . Among women , support is 72 - 20 .	ARG1	True
So my qestion is this , should gov't turn to slaves to save money ? @David cv , Depending on country it is already done . Hard labour , chaingangs , prison farms Voluntary work while incarcerated gets minimal pay etc	ARG0	False

Example 180 illustrates the third set of probing instances, corresponding to the phenomena labeled as SALIENT/RECENT in the interpretation types proposed here. Having prior knowledge that these instances are of this SALIENT/RECENT, it may be relatively easy for a model with a rich enough representation of syntactic position and referential distance to predict referents, but such a task would be very hard if a model cannot capture such information. Negative instances were sampled from high-scoring referents in prior context, balancing both a total high score with high scores in selectional preference, in order to avoid situations in which a mention can be disambiguated using selectional preference alone.[3]

[3] A natural addition to this set of tasks, although not currently proposed in this dataset, would be to study long-distance SRL dependencies such as control, topicalization, and wh-movement, which could be extracted from PropBank (Palmer **et al.** 2005). Like these SALIENT/RECENT instances, such information would probe the ability of models to handle relatively long-distance dependencies as they relate to semantic roles.

OP trying to win brownie points from other clueless students who hate the Conservative party just because. And failing .	ARG2	True
Actually.. some states are doing what the Fed won't. and adding Stimulus .	ARG0	True
Hi , I 've a dilema ! Looking to remortgage, and have found suitable product from Yorkshire Building Society . £195 non-refundable application fee, conveyancing and valuation free .	ARG1	False
Well I would argue that by definition when you have 16 % of people unemployed and underemployed it really is not. They 'd do exactly what they 've done since 1913. Lend out " money " that does n't exist , charge interest for it , inflate the hell out of the currency and line their pockets in the process .	ARG0 PAG	False

(180)

Example 181 below illustrates the largest subtask for this probing set, the instances of *script-inferrable* implicit roles. We suggest that these types may be of particular value in the kind of probing most commonly employed currently, in measuring spans over unsupervised representations such as BERT or ELMO, as these instances require a richer representation of how a particular participant fits into a larger series of events. The examples in 181 illustrate the true and false instances of this data, where negative examples are taken fromhigh-scoring but incorrect antecedents in this task.

(181)

The detective in charge has n't answered phone calls and his bag has n't been returned to him , which contained business papers and clothes . DH did need treatment at A and E after the attack It seems pretty clear that the police did not bother with forensic testing and may not even have checked the CCTV footage .	ARG0	True
If you have to drive, the hybrid is a joy. Dead silent, dirt cheap, reliable, little bit of capital cost detriment, sufficient power. 48 mpg if you drive at 75+ like a maniac .	ARG1	True
They took DH 's bag (which they found dumped a few streets away) and said they would forensically test it, and there was a CCTV camera which covered the area and the cabby said he could definitely identify the men as he got a clear view .	ARG0	True
KABUL, Afghanistan (AP) – A military helicopter crashed in eastern Afghanistan, killing 31 U.S. special operation troops and seven Afghan commandos, the country 's president said Saturday. An American official said it was apparently shot down, in the deadliest single incident for American forces in the decade - long war .	ARG0	False
He got a letter in the post this morning, dated Tuesday, saying the police were closing their investigation. The detective in charge has n't answered phone calls and his bag has n't been returned to him , which contained business papers and clothes .	ARG0	False
He was still popular and the country was willing to forgive him and frankly , his being acquitted was a forgone conclusion .	ARG2	False

This dataset therefore will provide the ability to probe whether a given representation of meaning in context is capturing the different kinds of meanings necessary for implicit role resolution. A natural next step for this data would be to evaluate human performance on such true-or-false judgments (which will likely be much higher than current systems), and to evaluate current word representation systems using such as task.

Chapter 6

Roadmap Forward

This work concludes with an examination of the future directions in implicit roles, not only for linguistic issues but also for data collection and computational models.

6.1 Future Directions in Linguistic Analysis of Implicit Roles

This work both provides a preliminary inventory of different implicit role interpretation types while also providing an initial exploration of how to describe language-specific constructions using that inventory. One could extend the linguistic claims presented here either of two promising future directions – either expanding from analysis of referential interpretation to also consider the licensing of implicit roles (determining when implicit roles will be used), or expanding from discussions of implicit role interpretation to related issues seen with explicit referents, such as connecting to interpretation inventories involved with articles or connecting to issues raised in the study of distant instantiation.

6.1.1 Interaction with Licensing Constraints – Coherence, Aspect, and Deprofiling

This thesis focuses upon implicit role interpretations, rather than the "licensing" factors determining when you can use an implicit role construction rather than an explicit form. Although easy of interpretation is an important factor for whether a speaker of a given language will choose to use an implicit role construction, it is not the only factor which affects whether an implicit role construction will be licensed in a particular context. Licensing of implicit role constructions has also been linked to information-structural issues such as deprofiling (de-emphasizing the argument, (Goldberg 2001)) or emphasizing the event (Rice 1988), linked to event-internal conceptual structure representations of which arguments were "essential" (Rappaport Hovav & Levin 1998), and even linked to factors such as discourse coherence (Iida 1996; Tao 1996) or telicity (Olsen & Resnik 1997; Mittwoch 1971).

Some of those postulated factors which have been linked to licensing could be represented with computational models, such as estimating "information gain" of using an explicit argument vs. learning that argument unstated. Others, such as the Goldberg hypothesis connecting implicit arguments to deprofiling, or proposed connections to aspectual implications, might be manually annotated (e.g., one might utilize aspect annotation frameworks such as that of Croft **et al.** (2016)). In that sense, just as one can model constructional choices such as dative alternations using models such as Bresnan **et al.** (2007), one might similarly model the licensing of English implicit object constructions. This would allow one to more rigorously pivot from "single-factor" explanations of implicit role licensing to richer and more corpus-driven ways of approaching the same questions. The same approach could be replicated for implicit subjects as well (as has been done in simple regression studies in sociolinguistics, as discussed in Chapter 1).

All of these factors support the notion of building rich, more predictive models of when implicit roles are licensed, and when they are ungrammatical. There are some prior models of referential choice proposed in the past (Grüning & Kibrik 2005), and regression models from sociolinguistics and psycholinguistics as discussed in Chapter 1, but such approaches have often been unable to model many of the more nuanced pragmatic factors involved in implicit role licensing. The kinds of models presented here, or improvements therein, might allow one to more accurately represent issues such as selectional preference, predictability, and salience. Such models would also necessarily need to learn seemingly arbitrary requirements defined by the syntactic subcategorization of particular verbs (such as the notion that one can say "she loaded the wagon" but cannot have an implicit object for "*she loaded into the wagon").

6.1.2 Connections to Local and Distant Instantiation

A second future direction of work regarding the current inventory of interpretation types would be a more general synthesis between the interpretation types proposed here and inventories of referent types which have been proposed for explicit reference forms, such as Becker's (2018) inventory of articles, Himmelmann's (1996) typology of demonstratives, or the Haspelmath (1997) inventory of indefinite pronoun types. This would allow one to compare behaviors in similar contexts – allowing one to link generic implicit roles into the same class as explicit generics such as *one* or the French *on*, or cluster SALIENT/RECENT implicit roles with pronominal usages.

Similarly, one might attempt to connect implicit role interpretation behaviors to other kinds of long-distance semantic role instantiation, such as control, topicalization, or wh-gap resolution. Because certain kinds of implicit role interpretations (particularly some of what we are calling SALIENT/RECENT roles) can be thought of as syntactically deterministic (Carminati 2002; Auer 2014), one may examine how these deterministic effects across sentence boundaries interact with the more officially "controlled" effects in within-sentence distant instantiations.

6.2 Future Directions for Data – How should future iSRL corpora be annotated

Another future direction is to simply improve the size and quality of data available for implicit role resolution. The current work has dramatically increased the size of data annotated with implicit role phenomena. However, in spite of this increase of implicit role data (6452 recoverable instances if one counts the within-sentence AMR conversions, which is nearly an order of magnitude more than the existing NomBank annotation), these datasets are still too small, and too noisy, to do more than fine-tuning of models. Therefore, the primary future direction for implicit role data is to determine methods for expanding the implicit role landscape beyond this limited situation. Naturally, these goals are intertwined with ways of automatically augmenting data – such as training datasets upon SRL or coreference tasks – but it is also possible that genuinely useful and insightful implicit role resolution systems will require both improvements from augmented data as well as increases in targeted, manual implicit role annotations.

6.2.1 Crowd-sourcing, Annotating Recoverability, and Confirmation of Annotations

There are many recent works which have explored more lightweight approaches to coreference or SRL annotation, including preliminary works on the crowd-sourcing of implicit roles. These may illuminate ways of scaling up the annotation of new implicit role resolution data. For coreference annotation, Sankepally **et al.** (2018) found sets of potentially coreferent mentions (mentioned in different documents) and asked annotators to judge on a scale from 1-5 the likelihood of those referring to the same referent. Paun **et al.** (2018) used a games-with-a-purpose framework to develop coreference annotations, by asking annotators whether a given referent was recoverable, asking them to highlight a prior mention, and then confirming those annotations with a separate confirmation task looking at individual pairs QA-SRL (Fitzgerald et al. 2018, He et al. 2015) annotated SRL with simple wh-questions of annotators, asking them to highlight spans in response to a question, also using a preliminary pass in which people would judge the validity of particular questions. Finally, Feizabadi and Pado (2015) presented a small pilot annotation involving the crowd-sourced annotation of implicit roles, experimenting both with asking annotators for spans corresponding to an argument, and asking them to fill out a template (such as "They went from ______ to ______ "). Many of these approaches share some of the same insights – that there are advantages to separately annotating things like recoverability, and having separate annotations of individual annotation pairs which one might want to confirm.

6.2.2 Noisy Methods For Preliminary Implicit Role Links

We suggest that one future path for annotation in a rapid manner would be to focus upon lightweight annotations in which non-recoverable implicit roles might be easily annotated as non-recoverable, and wherein annotators would provide basic clues regarding the referent of the implicit role. The low rate of recoverability seen in the MS-AMR dataset supports such approaches. One way of approaching such a lightweight annotation of recoverability might involve small lists of ”highly informative” options, so that one might ask an annotator whether an argument is recoverable while also checking for answers which might be easily understood and linked to, such as likely referents (such as “the speaker”). One might even present noisy ways of getting an annotation of the referent, such as the “fill in the blank” option introduced with Feizabadi and Pado (2015), under the assumption that any conclusions would be passed through confirmatory quality control annotations. Figure 6.1 illustrates one possible way on might frame such a prompt.

I think a lot more people feel the same way as you do , but medicate themselves by using friends . Whilst it probably is n't a ' solution ' as such , being around friends whilst being depressed is definitely better than being alone and being depressed . Either way , if you fancy a chat , feel free to PM.
Who is the “hearer” of chat?

(1) There is no “hearer” for this event

(2) There is a “hearer”, but they aren’t mentioned in the text

(3) The speaker/writer of that sentence (“me”)

(4) The audience of that sentence (“you”)

(5) With friends

(6) Someone/something else mentioned in the text: ________

Figure 6.1: Examples of possible recoverability-based annotations

Other approaches for acquiring possible implicit roles would be to either transfer data from explicit coreference data (by actually deleting arguments), or to get implicit role information from projection or distance supervision.

Augmentation of explicit coreference data would follow the tradition of explicit SRL data for implicit semantic roles learning (Hermann **et al.** 2015; Cheng & Erk 2018a; Liu **et al.** 2016). At its simplest, one can take texts annotated with explicit coreference data and simply delete arguments such as pronouns, so that the explicit semantic roles filled by those mentions can then be treated as if they were implicit roles referring to the antecedents of that deleted mention. This is a noisy approach which must also be treated with caution, however, because of the potential to change meaning through that deletion. For example, in example 182a, the deletion of "I" in subject position can lead to a valid implicit role annotation, but example 182b may change this construal, implying that the speaker is the one that moved (since dropped subjects in English are often first-person deictic subjects). Because of this, such augmented data could not be treated as high-quality training data, but would need to instead go through later annotations for quality control.

(182) I called a friend, but he had just moved into a new home.

a. $\emptyset_I$ called a friend, but he had just moved into a new home.

b. I called a friend, but $\emptyset_{he/i}$ had just moved into a new home. (Weischedel **et al.** 2011:c2e file 47)

Distant supervision and annotation projection are other approaches which could provide noisy annotations of implicit roles, which might require additional quality control. Both approaches assume that one used different resources to find an instance of an explicitly instantiated semantic role, and link the predicate to an instance of the same predicate without such an explicit mention, attempting to then link that predicate to the referent. Distant supervision from cross-document argument alignment (Roth & Frank 2013) finds explicit instances of semantic role links within a larger set of documents (such as news documents about the same event), and then links instances of an event with explicit arguments to mentions of the same event with no arguments. This adds noise due to the argument alignment and cross-document event coreference, leading to noisy annotations. Sikos **et al.** (2016) explored projection over parallel text – starting with a sentence where a semantic role is explicitly stated in one language, but where the participant does not seem to be stated in the same sentence in the other language. Such projection approaches also have sources of noise, due to both translation shifts (Dorr 1994) and to alignment errors. It should also be noted that such projection approaches have only been applied between two languages which are limited in their use of implicit role constructions (English and German), and that the projection approaches may have their most utility when projecting from a language such as English to a language with far more implicit roles, such as Chinese, Japanese or Arabic.

6.2.3 Confirmation Tasks

All such approaches propose methods for finding implicit roles, but all such approaches present quite noisy implicit role annotations. The same issue of noise is also relevant to the current implicit role datasets, as implicit role labeling is a difficult task with relatively low IAA. We thus suggest that the other kind of annotation would need to focus, as in Sankepally **et al.** (2018), upon starting with pairwise comparisons (such as those provided by the above noisy annotation sources) and getting annotation scores, ideally upon a scale from absolutely true to absolutely false. Such annotations could help one to utilize the various ways of capturing possible implicit roles, and may provide a path forward in cheaply developing high-quality corpora for implicit roles.

6.3 Computational Roadmap forward

The corpora and models present herein may provide the first opportunity to do this task as a general, wide-coverage way of approaching implicit roles, allowing us to extend from the specific work within SemEval to general implicit role resolution over difficult data such as discussion forum text. However, the models proposed in this thesis do not provide the desired level of performance on the implicit role test sets.

We suggest that the current literature, and the negative impacts of preprocessing seen here, support the value of using end-to-end neural networks at this task. This is supported by the recent state-of-the-art performance of neural models transferring from SRL (Do **et al.** 2017; Cheng & Erk 2018b). However, it is still unclear how one should leverage information from these SRL and coreference annotations.

Following earlier works such as Palmer (1985) and Silberer and Frank (2012), as well as more recent systems, one possible way of approaching this goal would be to train a system to do SRL and coreference at the same time. We suggest that many issues discussed in the prior chapters help to define the desiderata that we would want out of such a model – most importantly, that such a model would need to be able to learn the pragmatic factors (such as salience or "narrative schema" information) necessary to make connections between semantics roles, and thus to learn implicit role resolution behaviors. We suggest that there are two plausible avenues for such a system to be developed, but that important challenges remain in either approach.

6.3.1 One Possible Joint SRL/Coreference model – Span Selection

Increasingly, neural SRL and coreference models have relatively similar architectures. This similarity is most obvious in the comparison of new span-based SRL systems (He **et al.** 2018; Ouchi **et al.** 2018) and span-pruning-based coreference systems (Lee **et al.** 2018), which have very similar architectures. Both approaches treat their task as being a comparison between two spans (with a span representing a coreference mention, an SRL argument, or an SRL predicate), where the underlying tokens in each span are represented by a method of representing words in context (Peters **et al.** 2018; Devlin **et al.** 2018), and where one attends over these spans to attain a single vector for each span. In both approaches, most theoretically possible spans are pruned by an initial mention scoring function, and then the remaining possible comparisons are done using some kind of comparison function. They differ only in that the coreference models make a single linking decision (whether two mentions are coreferent), whereas SRL models decide whether a predicate and argument are connected while also deciding upon the semantic role label itself. We suggest that these tasks are similar enough that it may be possible to actually merge these two related approaches into a single, unified task.

The path towards joining those two approaches would require one to treat each semantic role not as a link to the predicate, but as a separate kind of mention to be clustered in the same "clustering" approach used in coreference. This would mean that the simple sentence "John ate his bagel" would consider both a series of entity spans ("John", "his", "his bagel", "ate") but also some semantic roles ("arg0 of eat", "arg1 of eat"), and attempt to build a cluster over both – so that you have one cluster corresponding to ("John", "his", and "arg0 of eat") and another cluster being ("his bagel" and "arg1 of eat").

The specific motivation for such an approach is that such clustered semantic roles could learn to make cross-sentence links independently of the mentions they locally refer to – which is exactly the kind of knowledge necessary to learn implicit role resolution. In addition, such models could directly learn information that is inherent to combinations of semantic roles, such as rules for combining semantic roles defined by grammatical control and binding structures, or by directly learning how particular semantic roles should combine. For example, "he was sentenced to prison after pleading guilty" has a long-distance relationship between *he* and the agent of *plead*, but may be easier to model by first attempting to link the agent of *plead* with the defendant of the *sentence* event. This would also potentially allow the learning of narrative schema effects through such clusterings, especially if combined with higher-order inference models used for coreference (Lee **et al.** 2018). However, it would dramatically exacerbate the issues that are already quite present in coreference models such as Lee **et al.** (2018); it would introduce a huge amount of additional entities into a document, making it hard to scale something into an entire document.

6.3.2 A Second Route towards Joint SRL and Coreference – Document-level AMR

A second route towards a combination of SRL and coreference would be through the task of "multi-sentence AMR" parsing – parsing sentences with AMR representations while simultaneously making coreference decisions, so that the output is a graph of each document, as defined by the annotations in the MS-AMR corpus (O'Gorman **et al.** 2018b). In such approaches, one would have a graph representing many sentences, and predict the existence of a semantic role using that context – likely using ways of leveraging rich representation of graph meaning in context, such as graph convolutional networks or gated graph neural networks (Damonte & Cohen 2019; Beck **et al.** 2018).

One version of this approach would be to parse each AMR on a sentence-by-sentence basis, but to incrementally link AMR concepts to prior entries in the MS-AMR graph during parsing, taking an incremental approach to coreference as proposed in recent linguistically inspired works (Webster & Curran 2014; Tuggener 2016). In analogy to context-aware frame-semantic parsing (Roth & Lapata 2015) or to pointer network approaches to implicit role resolution (Cheng & Erk 2018b), one would attempt to get such a model to attend to prior mentions of a given referent. However, it should be noted that in order for a model to be incentivized to learn to predict implicit roles when parsing explicit text, one might need to do modifications (such as adding noise to the input string) so that a model might rely more upon the richer contextual information and less upon surface strings.

Other directions in using AMR-style graphs for implicit role resolution would be through *denoising* approaches over AMR graphs or over document-level MS-AMR graphs. Given a graph representation of a given sentence or sequence of sentences, a denoising approach would be to remove or add semantic roles to a graph, and then predict which roles are correct and which were part of the added noise. Such approaches might be viewed as similar to the kind of *denoising autoencoder* methods (Vincent **et al.** 2010) which have shown to be effective in language model pre-training (Devlin **et al.** 2018), and might allow one to more directly learn the rich interactions involved with narrative schema and selectional preference information, through learning to judge the plausibility of a given semantic role in context.

6.3.3 Interpretation Types as Explicit Forms

A third path forward for implicit roles – having some similarity to the notion of dropping arguments as a way of data augmentation – would be to predict the interpretation type of a possible implicit role, but to do so in the context of recovering "dropped" content – adding explicit referents into a text corresponding to that implicit role, and then feeding such modified texts into a normal coreference system trained to work on such explicit forms. Recent work in implicit role resolution for Chinese (Yang **et al.** 2019) has focused upon ways of detecting implicit roles and characterizing the types of their "dropped pronouns" in this manner, adding back in various kinds of pronouns before coreference resolution. It should be noted that such an approach may be very promising for certain specific kinds of contexts, wherein the implicit role clearly alternates with a particular kind of explicit referent (such as replacing null complements with *it* or *that*), and may be particularly useful for languages in which some grammatically core implicit roles have morphological information required for disambiguation, such as number, person and gender. In this manner, one might imagine using the prediction of interpretation types to map from a sentence such as "now I understand!" to "now I understand [**it**]", and then running such an augmented document through traditional coreference models.

6.4 On Endeavors Connecting Linguistic and Computational Implicit Role Research

As deep learning approaches become more and more capable of handling coreference and SRL phenomena, it will be more and more likely that such systems can actually learn to represent the kind of underlying latent features which may be necessary to model implicit roles. However, rich contextual word embeddings are also dependent upon the explicit mention of referents in text. We suggest that the implicit role resolution data developed here, and the implicit role types outlined, may help to provide insight into such systems – whether they are directly trained for the implicit role task, or when probing unsupervised systems for whether they can inherently learn such information. The datasets proposed herein, as well as the probing corpus presented for future probing-based tasks, may help provide a measure of linguistic insight into what such models are actually learning, and to inform future representations which may be more capable of handling implicit content.

Bibliography

ABE, NAOKI, & HANG LI. 1993. Learning word association norms using tree cut pair models. Proceedings of ICML 1993 .

ABNEY, STEVEN, & MARC LIGHT. 1999. Hiding a semantic hierarchy in a markov model. Unsupervised Learning in Natural Language Processing .

ADI, YOSSI, EINAT KERMANY, YONATAN BELINKOV, OFER LAVI, & YOAV GOLDBERG. 2016. Fine-grained analysis of sentence embeddings using auxiliary prediction tasks. arXiv preprint arXiv:1608.04207 .

ALFARAZ, GABRIELA G. 2015. Variation of overt and null subject pronouns in the spanish of santo domingo. Subject pronoun expression in Spanish: A cross-dialectal perspective 3–16.

ALLEN, S, MARY HUGHES, & BARBORA SKARABELA. 2015. The role of cognitive accessibility in children's referential choice. The acquisition of reference 123–155.

ALLEN, SHANLEY EM, BARBORA SKARABELA, & HUGHES. 2008. Using Corpora to examine discourse effects in syntax. Corpora in Language Acquisition Research: History, methods, perspectives .

ALLERTON, DAVID J. 1975. Deletion and proform reduction. Journal of Linguistics 11.213–237.

ANDERSON, JOHN R, MICHAEL MATESSA, & CHRISTIAN LEBIERE. 1997. Act-r: A theory of higher level cognition and its relation to visual attention. Human-Computer Interaction 12.439–462.

ANDERSON, LLOYD B. 1982. The 'perfect'as a universal and as a language-specific category. Tense-aspect: Between semantics and pragmatics 227–264.

ARIEL, MIRA. 1988. Referring and accessibility. Journal of linguistics 24.65–87.

——. 2004. Accessibility marking: Discourse functions, discourse profiles, and processing cues. Discourse processes 37.91–116.

——. 2013. Centering, accessibility and the next mention. Theoretical Linguistics 39.39–58.

ARNOLD, JENNIFER E., 1998. Reference form and discourse patterns. Stanford University Stanford, CA dissertation.

AUER, PETER. 2014. Syntactic structures and their symbiotic guests. Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA) 24.533–560.

Bach, Kent. 1994. Conversational impliciture. Mind & language 9.124–162.

Bahdanau, Dzmitry, Kyunghyun Cho, & Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 .

Baker, Collin F., Charles J. Fillmore, & John B. Lowe. 1998. The Berkeley FrameNet project. In Proceedings of the 17th International Conference on Computational Linguistics (COLING/ACL-98), 86–90, Montreal.

Banarescu, Laura, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, & Nathan Schneider. 2013. Abstract meaning mepresentation for sembanking. In Proceedings of the Linguistic Annotation Workshop.

Barbosa, Pilar P. 2011. Pro-drop and theories of pro in the minimalist program part 1: Consistent null subject languages and the pronominal-agr hypothesis. Language and Linguistics Compass 5.551–570.

Baroni, Marco, & Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. Computational Linguistics 36.673–721.

Beck, Daniel, Gholamreza Haffari, & Trevor Cohn. 2018. Graph-to-sequence learning using gated graph neural networks. arXiv preprint arXiv:1806.09835 .

Becker, Laura, 2018. Articles in the world's languages. University of Leipzig dissertation.

Bhatia, Archna, Mandy Simons, Lori Levin, Yulia Tsvetkov, Chris Dyer, & Jordan Bender. 2014. A Unified Annotation Scheme for the Semantic/Pragmatic Components of Definiteness. In Proceedings of LREC.

Björkelund, Anders, Bernd Bohnet, Love Hafdell, & Pierre Nugues. 2010. A high-performance syntactic and semantic dependency parser. In Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, 33–36. Association for Computational Linguistics.

Boas, Hans Christian, & Ivan A. Sag. 2012. Sign-based construction grammar. CSLI Publications/Center for the Study of Language and Information.

Bonial, Claire, Olga Babko-Malaya, Jinho D Choi, Jena Hwang, & Martha Palmer. 2010. Propbank annotation guidelines. Center for Computational Language and Education Research Institute of Cognitive Science University of Colorado at Boulder .

——, Bianca Badaru, Kira Griffitt, Ulf Hermjakob, & Kevin Knight. 2018. Abstract meaning representation of constructions: The more we include, the better the representation. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, & R Harald Baayen. 2007. Predicting the dative alternation. In Cognitive foundations of interpretation, 69–94. KNAW.

Brown, Penelope. 2008. Verb specificity and argument realization in tzeltal child language. Crosslinguistic perspectives on argument structure: Implications for learnability 167–189.

Carminati, Maria Nella, 2002. The processing of Italian subject pronouns. University of Massachusets at Amherst dissertation.

Carston, Robyn. 2002. Linguistic meaning, communicated meaning and cognitive pragmatics. Mind & Language 17.127–148.

Caruana, Rich. 1997. Multitask learning. Machine learning 28.41–75.

——, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, & Noemie Elhadad. 2015. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721–1730. ACM.

Caselli, Tommaso, & Piek Vossen. 2017. The event storyline corpus: A new benchmark for causal and temporal relation extraction. In Proceedings of the Events and Stories in the News Workshop, 77–86.

Chambers, Nathanael. 2017. Behind the scenes of an evolving event cloze test. In Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics, 41–45.

——, & Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, 602–610. Association for Computational Linguistics.

——, & ——. 2010. Improving the use of pseudo-words for evaluating selectional preferences. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 445–453. Association for Computational Linguistics.

——, & ——. 2011. Template-based information extraction without the templates. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 976–986. Association for Computational Linguistics.

Chang, Tao, Shaohe Lv, Xiaodong Wang, & Dong Wang. 2017. Zero pronoun identification in chinese language with deep neural networks. In 2017 2nd International Conference on Control, Automation and Artificial Intelligence (CAAI 2017). Atlantis Press.

Chen, Jinying, 2006. Towards high-performance word sense disambiguation by combining rich linguistic knowledge and machine learning approaches. University of Pennsylvania dissertation.

Chen, Wei-Te, & Will Styler. 2013. Anafora: A web-based general purpose annotation tool. In HLT-NAACL, 14–19. Citeseer.

Cheng, Pengxiang, & Katrin Erk. 2018a. Implicit argument prediction as reading comprehension. arXiv preprint arXiv:1811.03554 .

——, & ——. 2018b. Implicit argument prediction with event knowledge. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), volume 1, 831–840.

CHOMSKY, NOAM. 1981. Lectures on government and binding: The Pisa lectures. Walter de Gruyter.

CHRISTODOULOUPOULOS, CHRISTOS, & MARK STEEDMAN. 2015. A massively parallel corpus: the bible in 100 languages. Language resources and evaluation 49.375–395.

CHUNG, TAGYOUNG, & DANIEL GILDEA. 2010. Effects of empty categories on machine translation. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 636–645. Association for Computational Linguistics.

CLARK, HERBERT H. 1977. Bridging. In Thinking: Readings in cognitive science, ed. by Philip N Johnson-Laird & Peter C Wason, 411–420. CUP.

CLARK, STEPHEN, & DAVID WEIR. 2002. Class-based probability estimation using a semantic hierarchy. In Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies, 1–8. Association for Computational Linguistics.

COHEN, JACOB. 1960. A coefficient of agreement for nominal scales. Educational and psychological measurement 20.37–46.

COLE, MELVYN. 2010. Thematic null subjects and accessibility. Studia Linguistica 64.271–320.

CONDORAVDI, CLEO, & JEAN MARK GAWRON. 1996. The context dependency of implicit arguments. In Quantifiers, Deduction and Context, 1–32. CSLI.

CONVERSE, SUSAN P. 2005. Resolving pronominal references in chinese with the hobbs algorithm. In Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing.

COOK, ANNE E, & JEROME L MYERS. 2004. Processing discourse roles in scripted narratives: The influences of context and world knowledge. Journal of Memory and Language 50.268–288.

CRAWLEY, ROSALIND A, ROSEMARY J STEVENSON, & DAVID KLEINMAN. 1990. The use of heuristic strategies in the interpretation of pronouns. Journal of Psycholinguistic Research 19.245–264.

CRISTOFARO, SONIA. 2010. Semantic maps and mental representation. Linguistic Discovery 8.

CROFT, WILLIAM, PAVLINA PESKOVA, & MICHAEL REGAN. 2016. Annotation of causal and aspectual structure of events in red: a preliminary report. In Proceedings of the Fourth Workshop on Events, 8–17.

——, HAVA BAT-ZEEV SHYLDKROT, & SUZANNE KEMMER. 1987. Diachronic semantic processes in the middle voice. In 7th international conference on historical linguistics, volume 48, 179–192. Current Issues in Linguistic Theory.

CYSOUW, MICHAEL. 2007. Building semantic maps: the case of person marking. New challenges in typology: Broadening the horizons and redefining the foundations 225–247.

DAHL, DEBORAH. 1986. Focusing and reference resolution in PUNDIT. In Proceedings of the workshop on Strategic computing natural language.

——, Martha Palmer, & Rebecca Passonneau. 1987. Nominalizations in PUNDIT. In Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics, Stanford University, Stanford, CA.

Damonte, Marco, & Shay B Cohen. 2019. Structural neural encoders for amr-to-text generation. arXiv preprint arXiv:1903.11410 .

Daniluk, Michał, Tim Rocktäschel, Johannes Welbl, & Sebastian Riedel. 2017. Frustratingly short attention spans in neural language modeling. arXiv preprint arXiv:1702.04521 .

David, Oana, 2016. Metaphor in the Grammar of Argument Realization. UC Berkeley dissertation.

De Waal, David A, & Jan V Du Toit. 2011. Automation of generalized additive neural networks for predictive data mining. Applied Artificial Intelligence 25.380–425.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 .

Dice, Lee R. 1945. Measures of the amount of ecologic association between species. Ecology 26.297–302.

Do, Quynh Ngoc Thi, Steven Bethard, & Marie-Francine Moens. 2017. Improving implicit semantic role labeling by predicting semantic frame arguments. arXiv preprint arXiv:1704.02709 .

Dorr, Bonnie J. 1994. Machine translation divergences: A formal description and proposed solution. Computational Linguistics 20.597–633.

Dowty, D. 1991. Thematic proto-roles and argument selection. Language 67.547–619.

Doyle, Arthur Conan. 1884. The Hound of the Baskervilles. The Strand Magazine.

——. 1917. His last bow: a reminiscence of Sherlock Holmes. AL Burt.

Dozat, Timothy, & Christopher D Manning. 2016. Deep biaffine attention for neural dependency parsing. arXiv preprint arXiv:1611.01734 .

Dryer, Matthew S. 2014. Competing methods for uncovering linguistic diversity: The case of definite and indefinite articles (commentary on davis, gillon, and matthewson). Language 90.e232–e249.

Eid, Mushira. 1983. On the communicative function of subject pronouns in arabic. Journal of Linguistics 19.287–303.

Elbourne, Paul. 2008. The argument from binding. Philosophical perspectives 22.89–110.

Erk, Katrin. 2007. A simple, similarity-based model for selectional preferences. In ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, volume 45, 216–223.

Erker, Daniel, & Gregory R. Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. Language 88.526–557.

Ettinger, Allyson, Ahmed Elgohary, & Philip Resnik. 2016. Probing for semantic evidence of composition by means of simple classification tasks. In Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP, 134–139.

Fehri, Abdelkader Fassi. 2009. Arabic silent pronouns, person, and voice. Brill's Journal of Afroasiatic Languages and Linguistics 1.3–40.

Feizabadi, Parvin Sadat, & Sebastian Padó. 2014. Crowdsourcing annotation of non-local semantic roles. In EACL, 226–230.

——, & Sebastian Padó. 2015. Combining seemingly incompatible corpora for implicit semantic role labeling. In * SEM@ NAACL-HLT, 40–50.

Fellbaum, Christiane (ed.) 1998. Wordnet: An Electronic Lexical Database. Cambridge, Massachusetts: The MIT Press.

Ferraro, Francis, & Benjamin Van Durme. 2016. A unified bayesian model of scripts, frames and language. In AAAI, 2601–2607.

Fillmore, Charles J. 1986. Pragmatically Controlled Zero Anaphora. In Proceedings of the 12th Annual Meeting of the Berkeley Linguistics Society, 95–107, Berkeley, California.

——, & Paul Kay. 1995. Construction Grammar Coursebook. University of California, Berkeley .

Flores-Ferrán, Nydia. 2004. Spanish subject personal pronoun use in new york city puerto ricans: Can we rest the case of english contact? Language variation and change 16.49–73.

Fodor, J. 1975. The Language of Thought. Harvard University Press.

Fox, B. A, 1984. Discourse Structure and Anaphora in written and conversational english. University of California at Los Angeles dissertation.

Gärdenfors, Peter. 2014. The geometry of meaning: Semantics based on conceptual spaces. MIT Press.

Gardner, Matt, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, & Luke Zettlemoyer. 2018. Allennlp: A deep semantic natural language processing platform. arXiv preprint arXiv:1803.07640 .

Garrod, Simon, & Melody Terras. 2000. The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. Journal of memory and language 42.526–544.

Gerber, Matthew, & Joyce Y. Chai. 2010. Beyond NomBank: A study of implicit arguments for nominal predicates. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 1583–1592.

——, & ——. 2012a. Semantic role labeling of implicit arguments for nominal predicates. Computational Linguistics 38.755–798.

——, & Joyce Y Chai. 2012b. Semantic role labeling of implicit arguments for nominal predicates. Computational Linguistics 38.755–798.

Givón, Talmy. 1978. Definiteness and referentiality. In Universals of human language, volume 4, 291–330. Stanford University Press Stanford, CA.

——. 1983. Topic continuity in discourse: A quantitative cross-language study, volume 3. John Benjamins Publishing.

——. 2017. The Story of Zero. John Benjamins Publishing Company.

Glass, Leila. 2014. What Does It Mean for an Implicit Object to be Recoverable? University of Pennsylvania Working Papers in Linguistics 20.14.

Glorot, Xavier, Antoine Bordes, & Yoshua Bengio. 2011. Deep sparse rectifier neural networks. In Proceedings of the fourteenth international conference on artificial intelligence and statistics, 315–323.

Goldberg, Adele. 2001. Patient arguments of causative verbs can be omitted: The role of information structure in argument distribution. Language Sciences .

——. 2006. Constructions at work: The nature of generalization in language. Oxford University Press, USA.

Goldberg, Adele E. 2005. Constructions, lexical semantics and the correspondence principle: Accounting for generalizations and subregularities in the realization of arguments. The syntax of aspect 215–236.

Gorinski, Philip, Josef Ruppenhofer, & Caroline Sporleder. 2013. Towards weakly supervised resolution of null instantiations. In Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) - Long Papers, 119–130.

Govindarajan, Venkata Subrahmanyan, Benjamin Van Durme, & Aaron Steven White. 2019. Decomposing generalization: Models of generic, habitual, and episodic statements. arXiv preprint arXiv:1901.11429 .

Green, Georgia M. 1996. The structure of context: The representation of pragmatic restrictions in hpsg. In Proceedings of the 5th annual meeting of the Formal Linguistics Society of the Midwest, Studies in the Linguistic Sciences, volume 24, 215–232.

Grosz, B. J., A. K. Joshi, & S. Weinstein. 1994. Centering: A Framework for Modelling the Coherence of Discourse. Technical Reports (CIS) p. 232.

Grosz, Barbara J, Aravind K Joshi, & Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In Proceedings of the 21st annual meeting on Association for Computational Linguistics, 44–50. Association for Computational Linguistics.

Grüning, André, & Andrej A Kibrik. 2005. Modeling referential choice in discourse: A cognitive calculative approach and a neural network approach. In Anaphora Processing: Linguistics, Cognitive and Computational Modeling, ed. by A. H. Branco, T. McEnery, & R. Mitkov. John Benjamins.

Gundel, Jeanette, Mamadou Bassene, Bryan Gordon, Linda Humnick, & Amel Khalfaoui. 2010. Testing predictions of the givenness hierarchy framework: A crosslinguistic investigation. Journal of Pragmatics 42.1770–1785.

Gundel, Jeannette, Nancy Hedberg, & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. Language 274–307.

Guzmán Naranjo, Matías. 2015. Unifying everything: Integrating quantitative effects into formal models of grammar. In Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics, 301–306.

Haegeman, Liliane. 1997. Register variation, truncation, and subject omission in english and in french. English Language & Linguistics 1.233–270.

Hajičová, Eva, & Markéta Ceplová. 2000. Deletions and their reconstruction in tectogrammatical syntactic tagging of very large corpora. In Proceedings of the 18th conference on Computational linguistics-Volume 1, 278–284. Association for Computational Linguistics.

Hakulinen, Auli, & Lauri Karttunen. 1973. Missing persons: on generic sentences in finnish. In ninth regional meeting. Chicago Linguistic Society (CLS 9), 157–171.

Han, Na-Rae. 2004. Korean null pronouns: classification and annotation. In Proceedings of the 2004 ACL Workshop on Discourse Annotation, 33–40. Association for Computational Linguistics.

——, 2006. Korean zero pronouns: analysis and resolution. University of Pennsylvania dissertation.

Hanks, Patrick. 2006. The organization of the lexicon: Semantic types and lexical sets. In Atti del XII Congresso Internazionale di Lessicografia: Torino, 6–9.

——, & Elisabetta Jezek. 2008. Shimmering lexical sets. In Proceedings of the XIII EURALEX International Congress, 391–402.

Hartmann, Iren, Martin Haspelmath, & Michael Cysouw. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. Studies in Language 38.463–484.

Haspelmath, Martin. 1997. Indefinite pronouns. Clarendon Press Oxford.

Hastie, Trevor, & Robert Tibshirani. 1987. Generalized additive models: some applications. Journal of the American Statistical Association 82.371–386.

Hawkins, John A. 1978. Definiteness and indefiniteness: a study in reference and grammaticality prediction. Taylor and Francis.

He, Luheng, Kenton Lee, Omer Levy, & Luke Zettlemoyer. 2018. Jointly predicting predicates and arguments in neural semantic role labeling. arXiv preprint arXiv:1805.04787 .

——, Kenton Lee, Mike Lewis, & Luke Zettlemoyer. 2017. Deep semantic role labeling: What works and what's next. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume 1, 473–483.

Hermann, Karl Moritz, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, & Phil Blunsom. 2015. Teaching machines to read and comprehend. In Advances in Neural Information Processing Systems, 1693–1701.

HIMMELMANN, NIKOLAUS P. 1996. Demonstratives in narrative discourse: a taxonomy of universal uses. Studies in anaphora 33.205.

HOCHREITER, SEPP, & JÜRGEN SCHMIDHUBER. 1997. Long short-term memory. Neural computation 9.1735–1780.

HOJI, HAJIME. 1998. Null object and sloppy identity in japanese. Linguistic Inquiry 29.127–152.

HOLMBERG, ANDERS. 2005. Is there a little pro? evidence from finnish. Linguistic inquiry 36.533–564.

——, AARTI NAYUDU, & MICHELLE SHEEHAN. 2009. Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi*. Studia Linguistica 63.59–97.

HOVY, EDUARD, MITCHELL MARCUS, MARTHA PALMER, LANCE RAMSHAW, & RALPH WEISCHEDEL. 2006. OntoNotes: the 90% solution. In Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers, 57–60.

——, TERUKO MITAMURA, FELISA VERDEJO, JUN ARAKI, & ANDREW PHILPOT. 2013. Events are not simple: Identity, non-identity, and quasi-identity. In NAACL HLT, volume 2013, p. 21.

HUANG, C.T. 1984. On the distribution and reference of empty pronouns. Linguistic Inquiry 18.321–337.

HUANG, YAN. 2018. Unarticulated constituents and neo-gricean pragmatics. Language and Linguistics 19.1–31.

IIDA, MASAYO. 1996. Discourse coherence and shifting centers in Japanese texts. In Centering in Discourse, ed. by Walker, Joshi, & Prince. Oxford University Press.

IIDA, RYU, KENTARO INUI, & YUJI MATSUMOTO. 2006. Exploiting syntactic patterns as clues in zero-anaphora resolution. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, 625–632.

——, MAMORU KOMACHI, KENTARO INUI, & YUJI MATSUMOTO. 2007a. Annotating a japanese text corpus with predicate-argument and coreference relations. In Proceedings of the Linguistic Annotation Workshop, 132–139. Association for Computational Linguistics.

——, ——, ——, & ——. 2007b. Annotating a japanese text corpus with predicate-argument and coreference relations. In Proceedings of the Linguistic Annotation Workshop, 132–139. Association for Computational Linguistics.

JACKENDOFF, RAY. 1999. Parallel constraint-based generative theories of language. Trends in Cognitive Sciences 3.393–400.

JAEGGLI, OSVALDO, & KENNETH J. SAFIR. 1989. The null subject parameter and parametric theory. In The null subject parameter, 1–44. Springer.

JANS, BRAM, STEVEN BETHARD, IVAN VULIĆ, & MARIE FRANCINE MOENS. 2012. Skip n-grams and ranking functions for predicting script events. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 336–344. Association for Computational Linguistics.

JANTUNEN, TOMMI. 2013. Ellipsis in finnish sign language. Nordic Journal of Linguistics 36.303–332.

JEZEK, ELISABETTA, & PATRICK HANKS. 2010. What lexical sets tell us about conceptual categories. Lexis. Journal in English Lexicology .

JOHNSON-LAIRD, PHILIP NICHOLAS. 1980. Mental models in cognitive science. Cognitive science 4.71–115.

——. 1983. A computational analysis of consciousness. Cognition & Brain Theory .

KAISER, ELSI, & JOHN C. TRUESWELL. 2008. Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. Language and Cognitive Processes 23.709–748.

KARTTUNEN, LAURI. 1968. What do referential indices refer to?, volume 3854. Rand Corporation.

KASHER, ASA. 1998. Pragmatics: Indexicals and reference, volume 3. SUNY Press.

KATZ, JERROLD J, & JERRY A FODOR. 1963. The structure of a semantic theory. language 39.170–210.

KAY, PAUL. 2004. Null instantiation of nominal complements. Ms. Berkeley: University of California .

KAYAMA, YUHKO. 2003. L1 acquisition of japanese zero pronouns: The effect of discourse factors. In Proceedings of the 2003 Annual Conference of the Canadian Linguistic Society, 109–120.

KEENAN, EDWARD L, & MATTHEW S DRYER. 1981. Passive in the world's languages. Linguistic Agency, University of Trier.

KEYDANA, GÖTZ, & SILVIA LURAGHI. 2012. Definite referential null objects in vedic sanskrit and ancient greek. Acta Linguistica Hafniensia 44.116–128.

KHUDYAKOVA, MARIYA V, GRIGORY B DOBROV, ANDREJ A KIBRIK, & NATALIA V LOUKACHEVITCH. 2011. Computational modeling of referential choice: Major and minor referential options. In Proceedings of the cogsci 2011 workshop on the production of referring expressions. boston (july 2011).

KIBRIK, ANDREJ A, MARIYA V KHUDYAKOVA, GRIGORY B DOBROV, ANASTASIA LINNIK, & DMITRIJ A ZALMANOV. 2016. Referential choice: Predictability and its limits. Frontiers in psychology 7.

KINGMA, DIEDERIK P, & JIMMY BA. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

KIPPER, KARIN, HOA TRANG DANG, MARTHA PALMER, & OTHERS. 2000. Class-based construction of a verb lexicon. AAAI/IAAI 691.696.

KOENIG, JEAN-PIERRE, & GAIL MAUNER. 1999. A-definites and the discourse status of implicit arguments. Journal of Semantics 16.207–236.

Kwon, Nayoung, & Patrick Sturt. 2013. Null pronominal (pro) resolution in korean, a discourse-oriented language. Language and Cognitive Processes 28.377–387.

Lambrecht, K., & Lemoine. 2005. Definite null objects in (spoken) french. In Grammatical Constructions: Back to the Roots. John Benjamins Publishing Company.

Landau, Idan. 2018. Missing objects in hebrew: Argument ellipsis, not vp ellipsis. Glossa: a journal of general linguistics 3.

Laparra, Egoitz, & German Rigau. 2013. Impar: A deterministic algorithm for implicit semantic role labelling. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume 1, 1180–1189.

Larjavaara, Meri, 2000. Présence ou absence de l'objet. University of Helsinki dissertation.

Le, Minh, & Antske Fokkens. 2018. Neural models of selectional preferences for implicit semantic role labeling. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018).

Lee, Kenton, Luheng He, Mike Lewis, & Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. arXiv preprint arXiv:1707.07045 .

——, Luheng He, & Luke Zettlemoyer. 2018. Higher-order coreference resolution with coarse-to-fine inference. arXiv preprint arXiv:1804.05392 .

Levesque, Hector, Ernest Davis, & Leora Morgenstern. 2012. The winograd schema challenge. In Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning.

Li, Wendan. 2004. Topic Chains in Chinese Discourse. Discourse Processes 37.25–45.

Liu, Nelson F, Matt Gardner, Yonatan Belinkov, Matthew Peters, & Noah A Smith. 2019. Linguistic knowledge and transferability of contextual representations. arXiv preprint arXiv:1903.08855 .

Liu, Ting, Yiming Cui, Qingyu Yin, Weinan Zhang, Shijin Wang, & Guoping Hu. 2016. Generating and exploiting large-scale pseudo training data for zero pronoun resolution. arXiv preprint arXiv:1606.01603 .

Lou, Yin, Rich Caruana, & Johannes Gehrke. 2012. Intelligible models for classification and regression. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 150–158. ACM.

Luraghi, Silvia. 1997. Omission of the direct object in latin. Indogermanische Forschungen 239–257.

Lyngfelt, Benjamin. 2012. Re-thinking FNI: On null instantiation and control in Construction Grammar. Constructions and Frames 4.1–23.

Manning, Christopher, & Ivan A Sag. 1999. Dissociations between argument structure and grammatical relations. Lexical and Constructional Aspects of Linguistic Explanation. Stanford: CSLI Publications 63–78.

MANSOURI, AOUS, 2016. Stative and Stativizing Constructions in Arabic News Reports: A Corpus-Based Study. University of Colorado Boulder dissertation.

MASSAM, DIANE, & YVES ROBERGE. 1989. Recipe context null objects in english. Linguistic inquiry 20.134–139.

MAUNER, GAIL, & J. P KOENIG. 2000. Linguistic vs conceptual sources of implicit agents in sentence comprehension. Journal of Memory and Language 43.110–134.

——, ——, ALISSA MELINGER, & BRETON BIENVENUE. 2002. The lexical source of unexpressed participants and their role in sentence and discourse understanding. In The Lexical Basis of Sentence Processing: Formal, Computational and Experimental Issues. John Benjamins.

——, & M. K TANENHAUS, 1995. Implicit Arguments in Sentence Processing.

MELNIK, NURIT. 2007. Extending partial pro-drop in modern hebrew: A comprehensive analysis. Head-Driven Phrase Structure Grammar p. 173.

MEYERS, A., R. REEVES, C. MACLEOD, R. SZEKELY, V. ZIELINSKA, B. YOUNG, & R. GRISHMAN. 2004. Annotating noun argument structure for NomBank. In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004), 803–806.

MICHAELIS, LAURA A. 2009. Sign-based construction grammar. The Oxford handbook of linguistic analysis 155–176.

MIHAILA, CLAUDIU, IUSTINA ILISEI, & DIANA INKPEN. 2010. Romanian zero pronoun distribution: A comparative study. In LREC.

MITTWOCH, ANITA. 1971. Idioms and unspecified NP deletion. Linguistic Inquiry 255–259.

——. 1982. On the difference between eating and eating something: Activities versus accomplishments. Linguistic Inquiry 113–122.

MNIH, ANDRIY, & GEOFFREY E HINTON. 2009. A scalable hierarchical distributed language model. In Advances in neural information processing systems, 1081–1088.

MOOR, TATJANA, MICHAEL ROTH, & ANETTE FRANK. 2013a. Predicate-specific annotations for implicit role binding: Corpus annotation, data analysis and evaluation experiments. In Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Short Papers, 369–375.

——, MICHAEL ROTH, & ANNETTE FRANK. 2013b. Predicate-specific annotations for implicit role binding: Corpus annotation, data analysis and evaluation experiments. In Proceedings of the 10th International Conference on Computational Semantics.

MORIK, KATHARINA, PETER BROCKHAUSEN, & THORSTEN JOACHIMS. 1999. Combining statistical learning with a knowledge-based approach: a case study in intensive care monitoring. Technical report, Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten ....

MOSTAFAZADEH, NASRIN, NATHANAEL CHAMBERS, XIAODONG HE, DEVI PARIKH, DHRUV BATRA, LUCY VANDERWENDE, PUSHMEET KOHLI, & JAMES ALLEN. 2016. A corpus and evaluation framework for deeper understanding of commonsense stories. arXiv preprint arXiv:1604.01696 .

Nakaiwa, Hiromi, & Satoshi Shirai. 1996. Anaphora resolution of Japanese zero pronouns with deictic reference. In Proceedings of the 16th conference on Computational linguistics-Volume 2.

Nakipoğlu-Demiralp, Mine. 2001. The referential properties of the implicit arguments of impersonal passive constructions. The verb in Turkish 129–150.

Napoles, Courtney, Matthew Gormley, & Benjamin Van Durme. 2012. Annotated gigaword. In Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction, 95–100. Association for Computational Linguistics.

Neal, Radford M. 1998. Regression and classification using gaussian process priors. Bayesian statistics 6.475.

Nissim, Malvina, Shipra Dingare, Jean Carletta, & Mark Steedman. 2003. An annotation scheme for information status in dialogue. HCRC, University of Edinburgh. Unpub. ms .

O'Gorman, Tim, Sameer Pradhan, Martha Palmer, Julia Bonn, Kathryn Conger, & James Gung. 2018a. The new propbank: Aligning propbank with amr through pos unification. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018).

——, Michael Regan, Kira Griffitt, Ulf Hermjakob, Kevin Knight, & Martha Palmer. 2018b. Amr beyond the sentence: the multi-sentence amr corpus. In Proceedings of the 27th International Conference on Computational Linguistics, 3693–3702.

Oku, S. 1998. Lf copy analysis of japanese null arguments. In CLS(34), 299–314.

Olsen, M.B., & P. Resnik. 1997. Implicit object constructions and the (in) transitivity continuum. 33rd Proceedings of the Chicago Linguistic Society 327–336.

Ono, Tsuyoshi, & Sandra Thompson. 2000. Deconstructing "Zero Anaphora" in Japanese. In Proceedings of the Twenty-Third meeting of the Berkeley Linguistics Society.

Orita, Naho, Naomi Feldman, Jordan Boyd-Graber, & Eliana Vornov. 2014. Quantifying the role of discourse topicality in speakers' choices of referring expressions. In Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics, 63–70.

Otheguy, Ricardo, Ana Celia Zentella, & David Livert. 2007. Language and dialect contact in spanish in new york: Toward the formation of a speech community. Language 770–802.

Ouchi, Hiroki, Hiroyuki Shindo, & Yuji Matsumoto. 2018. A span selection model for semantic role labeling. arXiv preprint arXiv:1810.02245 .

Owens, Jonathan, Robin Dodsworth, & Trent Rockwood. 2009. Subject-verb order in spoken arabic: Morpholexical and event-based factors. Language Variation and Change 21.39–67.

Padó, Sebastian, Ulrike Padó, & Katrin Erk. 2007. Flexible, corpus-based modelling of human plausibility judgements. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL).

PALMER, MARTHA. 2009. Semlink: Linking propbank, verbnet and framenet. In Proceedings of the generative lexicon conference, 9–15.

——, DEBORAH A. DAHL, REBECCA J. SCHIFFMAN, LYNETTE HIRSCHMAN, MARCIA LINEBARGER, & JOHN DOWDING. 1986. Recovering implicit information. In Proceedings of the 24th annual meeting on Association for Computational Linguistics, 10–19. Association for Computational Linguistics.

——, P. KINGSBURY, & D. GILDEA. 2005. The proposition bank: An annotated corpus of semantic roles. Computational Linguistics 31.71–106.

PARKINSON, DILWORTH B. 1987. Constraints on the presence/absence of 'optional'subject pronouns in egyptian arabic. In 15th Annual Conference on New Ways of Analyzing Variation, 348–360.

PARTEE, BARBARA H. 1984. Nominal and temporal anaphora. Linguistics and philosophy 7.243–286.

PAUN, SILVIU, JON CHAMBERLAIN, UDO KRUSCHWITZ, JUNTAO YU, & MASSIMO POESIO. 2018. A probabilistic annotation model for crowdsourcing coreference. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 1926–1937.

PAVLICK, ELLIE, & ANI NENKOVA. 2015. Inducing lexical style properties for paraphrase and genre differentiation. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 218–224.

PENG, HAORUO, 2018. Understanding stories via event sequence modeling. University of Illinois at Urbana-Champaign dissertation.

PENNINGTON, JEFFREY, RICHARD SOCHER, & CHRISTOPHER MANNING. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532–1543.

PEREIRA, FERNANDO, NAFTALI TISHBY, & LILLIAN LEE. 1993. Distributional clustering of english words. In Proceedings of the 31st annual meeting on Association for Computational Linguistics, 183–190. Association for Computational Linguistics.

PEREIRA, SIMONE. 2009. Zac. pb: An annotated corpus for zero anaphora resolution in portuguese. In Proceedings of the Student Research Workshop, 53–59.

PEREZ-LEROUX, ANA TERESA, MIHAELA PIRVULESCU, & YVES ROBERGE. 2017. Direct Objects and Language Acquisition, volume 152. Cambridge University Press.

PERIS, AINA, & MARIONA TAULÉ. 2012. Annotating the argument structure of deverbal nominalizations in spanish. Language Resources and Evaluation 46.667–699.

——, MARIONA TAULÉ, & HORACIO RODRÍGUEZ. 2012. Iarg-ancora: Annotating the ancora corpus with implicit arguments. In Proceedings of ACL 2012.

——, MARIONA TAULÉ, HORACIO RODRÍGUEZ, & MANUEL BERTRAN IBARZ. 2013. Liar c: labeling implicit arguments in spanish deverbal nominalizations. In International Conference on Intelligent Text Processing and Computational Linguistics, 423–434. Springer.

Perlmutter, David. 1971. Deep and surface constraints in generative grammar. New York: Holt, Rinehart and Winston .

Perry, John, & Simon Blackburn. 1986. Thought without representation. Proceedings of the Aristotelian society, supplementary volumes 60.137–166.

Peters, Matthew E, Waleed Ammar, Chandra Bhagavatula, & Russell Power. 2017. Semi-supervised sequence tagging with bidirectional language models. arXiv preprint arXiv:1705.00108 .

——, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, & Luke Zettlemoyer. 2018. Deep contextualized word representations. arXiv preprint arXiv:1802.05365 .

Pichotta, Karl, 2017. Advances in statistical script learning. UT Austin dissertation.

——, & Raymond J Mooney. 2016. Learning statistical scripts with lstm recurrent neural networks. In AAAI, 2800–2806.

Poesio, Massimo, Renata Vieira, & Simone Teufel. 1997. Resolving bridging references in unrestricted text. In Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts, 1–6. Association for Computational Linguistics.

Pollard, Carl, & Ivan A Sag. 1994. Head-driven phrase structure grammar. University of Chicago Press.

Pourdamghani, Nima, Yang Gao, Ulf Hermjakob, & Kevin Knight. 2014. Aligning english strings with abstract meaning representation graphs. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 425–429.

Pradhan, Sameer, Lance Ramshaw, Mitch Marcus, Martha Palmer, Ralph Weischedel, & Nianwen Xue. 2011. Conll-2011 shared task: Modeling unrestricted coreference in ontonotes. In Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, 1–27.

Prince, Ellen. 1981. Toward a taxonomy of given-new information. In Radical Pragmatics, ed. by Peter Cole. Academic Press.

——. 1992. The ZPG letter: Subjects, definiteness, and information-status. Discourse description: Diverse linguistic analyses of a fund-raising text .

Pustejovsky, James, Nikhil Krishnaswamy, & Tuan Do. 2017. Object embodiment in a multimodal simulation. In AAAI Spring Symposium: Interactive Multisensory Object Perception for Embodied Agents.

Rao, Sudha, Allyson Ettinger, Hal Daumé III, & Philip Resnik. 2015. Dialogue focus tracking for zero pronoun resolution. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 494–503.

Rappaport Hovav, Malka, & Beth Levin. 1998. Building verb meanings. The projection of arguments: Lexical and compositional factors 97–134.

RECANATI, FRANÇOIS. 2002. Unarticulated constituents. Linguistics and Philosophy 25.299–345.

——. 2007. It is raining (somewhere). Linguistics and Philosophy 30.123–146.

RECASENS, MARTA, MARIE-CATHERINE DE MARNEFFE, & CHRISTOPHER POTTS. 2013. The life and death of discourse entities: Identifying singleton mentions. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 627–633.

REISINGER, DREW, RACHEL RUDINGER, FRANCIS FERRARO, CRAIG HARMAN, KYLE RAWLINS, & BENJAMIN VAN DURME. 2015. Semantic Proto-Roles. Transactions of the Association for Computational Linguistics 3.475–488.

RELLO, LUZ, & IUSTINA ILISEI. 2009. A comparative study of spanish zero pronoun distribution. In Proceedings of the International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages (ISMTCL), 209–214.

RESNIK, PHILIP. 1993. Selection and information: a class-based approach to lexical relationships. IRCS Technical Reports Series p. 200.

——. 1996. Selectional constraints: An information-theoretic model and its computational realization. Cognition 61.127–159.

RICE, SALLY. 1988. Unlikely Lexical Entries. In Proce dings of the 14th An ual Berkeley Linguistics Society, Berkeley.

RITTER, ALAN, OREN ETZIONI, & OTHERS. 2010. A latent dirichlet allocation method for selectional preferences. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 424–434. Association for Computational Linguistics.

RIZZI, LUIGI. 1982. Negation, wh-movement and the null subject parameter. Issues in Italian syntax 117.184.

RODRIGUEZ, KEPA JOSEBA, FRANCESCA DELOGU, YANNICK VERSLEY, EGON W STEMLE, & MASSIMO POESIO. 2010. Anaphoric annotation of wikipedia and blogs in the live memories corpus. In Proceedings of LREC, 157–163.

ROOTH, MATS, STEFAN RIEZLER, DETLEF PRESCHER, GLENN CARROLL, & FRANZ BEIL. 1999. Inducing a semantically annotated lexicon via em-based clustering. In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, 104–111. Association for Computational Linguistics.

ROTH, MICHAEL, & ANETTE FRANK. 2013. Automatically identifying implicit arguments to improve argument linking and coherence modeling. In Proceedings of the Joint Conference on Lexical and Computational Semantics, 306–316.

——, & MIRELLA LAPATA. 2015. Context-aware frame-semantic role labeling. Transactions of the Association for Computational Linguistics 3.449–460.

RUDA, MARTA. 2014. Missing objects in special registers: The syntax of null objects in english. Canadian Journal of Linguistics/Revue canadienne de linguistique 59.339–372.

Rudinger, Rachel, Pushpendre Rastogi, Francis Ferraro, & Benjamin Van Durme. 2015. Script induction as language modeling. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 1681–1686.

Rumelhart, David E. 1975. Notes on a schema for stories. In Representation and understanding, 211–236. Elsevier.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, C.R. Johnson, & J. Scheffcyzk. 2006. FrameNet II: Extended Theory and Practice. Technical report, ICSI.

——, & Laura Michaelis. 2010. A constructional account of genre-based argument omissions. Constructions and Frames 2.158–184.

——, & Laura A Michaelis. 2014. Frames and the interpretation of omitted arguments in english. Linguistic Perspectives on Structure and Context: Studies in Honor of Knud Lambrecht 57–86.

——, Caroline Sporleder, R. Morante, C. Baker, & M. Palmer. 2010a. Semeval-2010 task 10: Linking events and their participants in discourse. In Proceedings of the 5th International Workshop on Semantic Evaluation, 45–50. Association for Computational Linguistics.

——, Caroline Sporleder, Roser Morante, Collin Baker, & Martha Palmer. 2010b. Semeval-2010 task 10: Linking events and their participants in discourse. In Proceedings of the 5th International Workshop on Semantic Evaluation, 45–50. Association for Computational Linguistics.

Sag, Ivan A. 2012. Sign-based construction grammar: An informal synopsis. Sign-based construction grammar 193.69–202.

Sag, Ivan A., & Jorge Hankamer. 1976. Deep and surface anaphora. Linguistic inquiry 7.

Samek-Lodovici. 1996. Constraints on Subjects. Glot International 2.11–12.

Sankepally, Rashmi, Tongfei Chen, Benjamin Van Durme, & Douglas W Oard. 2018. A test collection for coreferent mention retrieval. In SIGIR, 1209–1212.

Schank, Roger C, & Robert P Abelson. 1977. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Psychology Press.

Schenk, Niko, & Christian Chiarcos. 2016. Unsupervised learning of prototypical fillers for implicit semantic role labeling. In Proceedings of NAACL-HLT, 1473–1479.

Séaghdha, Diarmuid Ó. 2010. Latent variable models of selectional preference. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 435–444. Association for Computational Linguistics.

——, & Anna Korhonen. 2012. Modelling selectional preferences in a lexical hierarchy. In Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation, 170–179. Association for Computational Linguistics.

Searle, John R. 1980. The background of meaning. In Speech act theory and pragmatics, 221–232. Springer.

Sellars, Wilfrid, & others. 1956. Empiricism and the philosophy of mind. Minnesota studies in the philosophy of science 1.253–329.

Şener, Serkan, & Daiko Takahashi. 2010. Ellipsis of arguments in japanese and turkish. Nanzan Linguistics 6.79–99.

Shopen, Timothy, 1972. A generative theory of ellipsis. UCLA dissertation.

Sidner, Candace Lee. 1979. Towards a computational theory of definite anaphora comprehension in english discourse. Technical report, Massachusetts Inst of Tech Cambridge Artificial Intelligence lab.

Sigurðsson, Halldór Ármann, & Verner Egerland. 2009. Impersonal null-subjects in icelandic and elsewhere. Studia Linguistica 63.158–185.

Sikos, Jennifer, Yannick Versley, & Anette Frank. 2016. Implicit semantic roles in a multilingual setting. c 2016 The* SEM 2016 Organizing Committee. All papers c 2016 their respective authors. This proceedings volume and all papers therein are licensed under a Creative Commons Attribution 4.0 International License. License details: http://creativecommons. org/licenses/by/4.0 p. 45.

Silberer, Carina, & Anette Frank. 2012. Casting implicit role linking as an anaphora resolution task. In Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation, 1–10. Association for Computational Linguistics.

Simpson, Andrew, Arunima Choudhury, & Mythili Menon. 2013. Argument ellipsis and the licensing of covert nominals in Bangla, Hindi and Malayalam. Lingua 134.103–128.

Skarabela, Barbora, & Shanley EM Allen. 2010. How newness and joint attention work together in child Inuktitut: Assessing discourse-pragmatic models of early argument realization. In Proceedings of the Annual Boston University Conference on Language Development, volume 34, 385–396.

Song, Sanghoun, & Emily M Bender. 2012. Individual constraints for information structure. In Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, 330–348.

Song, Zhiyi, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, & Xiaoyi Ma. 2015. From light to rich ere: annotation of entities, relations, and events. In Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, 89–98.

——, Stephanie Strassel, Haejoong Lee, Kevin Walker, Jonathan Wright, Jennifer Garland, Dana Fore, Brian Gainor, Preston Cabe, Thomas Thomas, Brendan Callahan, & Ann Sawyer. 2014. Collecting natural sms and chat conversations in multiple languages: The bolt phase 2 corpus. In LREC, 1699–1704. Citeseer.

Sperber, Dan, & Deirdre Wilson. 1986. Relevance: Communication and cognition, volume 142. Harvard University Press Cambridge, MA.

SRIVASTAVA, NITISH, GEOFFREY HINTON, ALEX KRIZHEVSKY, ILYA SUTSKEVER, & RUSLAN SALAKHUTDINOV. 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15.1929–1958.

STANLEY, JASON. 2002. Making it articulated. Mind & Language 17.149–168.

——, & ZOLTÁN GENDLER SZABÓ. 2000. On quantifier domain restriction. Mind & Language 15.219–261.

STERNAU, MARIT, MIRA ARIEL, RACHEL GIORA, & OFER FEIN. 2015. Levels of interpretation: New tools for characterizing intended meanings. Journal of Pragmatics 84.86–101.

STOLL, SABINE, & BALTHASAR BIKEL. 2009. How Deep are Differences in Referential Density? Psychology Press 543–555.

STRASSEL, STEPHANIE, & JENNIFER TRACEY. 2016. Lorelei language packs: Data, tools, and resources for technology development in low resource languages. In LREC.

TAO, L. 1996. Topic discontinuity and zero anaphora in Chinese discourse: Cognitive strategies in discourse processing. Studies in anaphora 487–513.

TAO, LIANG, 1993. Zero anaphora in Chinese: Cognitive strategies in discourse processing. University of Colorado Boulder dissertation.

——, & ALICE HEALY. 1996. Cognitive strategies in discourse processing: A comparison of Chinese and English speakers. Journal of Psycholinguistic Research 25.597–616.

——, & ALICE F. HEALY. 1998. Anaphora in Language Processing: Transfer of Cognitive Strategies by Native Chinese, Dutch, English, and Japanese speakers. In Foreign Language Learning: Psycholinguistic studies of training and retention, 193–211. Lawrence Erlbaum Associates, Inc.

——, & ALICE F HEALY. 2005. Zero anaphora: Transfer of reference tracking strategies from chinese to english. Journal of Psycholinguistic Research 34.99–131.

TARALDSEN, TARALD. 1978. On the nic, vacuous application and the that-trace filter. Ms., MIT, Cambridge, Mass.

TENNEY, IAN, DIPANJAN DAS, & ELLIE PAVLICK. 2019a. Bert rediscovers the classical nlp pipeline. arXiv preprint arXiv:1905.05950 .

——, PATRICK XIA, BERLIN CHEN, ALEX WANG, ADAM POLIAK, R THOMAS MCCOY, NAJOUNG KIM, BENJAMIN VAN DURME, SAM BOWMAN, DIPANJAN DAS, & ELLIE PAVLICK. 2019b. What do you learn from context? probing for sentence structure in contextualized word representations. In Proceedings of ICLR.

TETREAULT, JOEL. 2002. Implicit Role Reference. Ms., University of Rochester.

TONELLI, SARA, & RODOLFO DELMONTE. 2011. Desperately seeking implicit arguments in text. In Proceedings of the ACL 2011 Workshop on Relational Models of Semantics, 54–62.

TONHAUSER, JUDITH. 2017. The distribution of implicit arguments in paraguayan guaraní. In Guaraní Linguistics in the 21st Century, ed. by B. Estigarribia & J. Pinta, 194–230. Brill Publishing.

Torgerson, Warren S. 1952. Multidimensional scaling: I. theory and method. Psychometrika 17.401–419.

Traugott, Elizabeth Closs. 2016. Do semantic modal maps have a role in a constructionalization approach to modals? Constructions and Frames 8.97–124.

Tuggener, Don, 2016. Incremental coreference resolution for German. Universität Zürich dissertation.

Ueno, Mieko, & Andrew Kehler. 2010. The interpretation of null and overt pronouns in japanese: Grammatical and pragmatic factors. In Proceedings of the 32nd annual meeting of the Cognitive Science Society, 2057–2062.

——, & ——. 2016. Grammatical and pragmatic factors in the interpretation of japanese null and overt pronouns. Linguistics 54.1165–1221.

Van de Cruys, Tim. 2014. A neural network approach to selectional preference acquisition. In EMNLP, 26–35.

Van Der Auwera, J, & V Plungian. 1998. Modality's semantic map. Linguistic Typology 2.79–124.

Van Durme, Benjamin, Phillip Michalak, & Lenhart K Schubert. 2009. Deriving generalized knowledge from corpora using wordnet abstraction. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, 808–816. Association for Computational Linguistics.

Van Rij, Jacolien, Hedderik Van Rijn, & Petra Hendriks. 2013. How wm load influences linguistic processing in adults: a computational model of pronoun interpretation in discourse. Topics in cognitive science 5.564–580.

Vincent, Pascal, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, & Pierre-Antoine Manzagol. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of machine learning research 11.3371–3408.

Vincze, Veronika, Klára Hegedűs, Alex Sliz-Nagy, & Richárd Farkas. 2018. Szegedkoref: A hungarian coreference corpus. In Proceedings of LREC-2018.

Vogelzang, Margreet, Petra Hendriks, & Hedderik van Rijn. 2015. Processing overt and null subject pronouns in italian: a cognitive model. In CogSci.

Wälchli, Bernhard. 2010. Similarity semantics and building probabilistic semantic maps from parallel texts. Linguistic Discovery 8.331–371.

——, & Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. Linguistics 50.671–710.

Walker, Marilyn, Sharon Cote, & Masayo Iida. 1994. Japanese discourse and the process of centering. Computational linguistics 20.193–232.

Wang, Su, Greg Durrett, & Katrin Erk. 2018. Modeling semantic plausibility by injecting world knowledge. arXiv preprint arXiv:1804.00619 .

Webster, Kellie, & James R Curran. 2014. Limited memory incremental coreference resolution. In COLING, 2129–2139.

Weischedel, Ralph, Sameer Pradhan, Lance Ramshaw, Martha Palmer, Nianwen Xue, Mitchell Marcus, Ann Taylor, Craig Greenberg, Eduard Hovy, Robert Belvin, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, & Ann Houston. 2011. Ontonotes release 4.0. Technical report, Linguistic Data Consortium.

Wilks, Yorick. 1978. Making preferences more active. Artificial intelligence 11.197–223.

Wu, Shumin, & Martha Palmer. 2015. Can selectional preferences help automatic semantic role labeling? In Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics, 222–227.

Wu, Zhibiao, & Martha Palmer. 1994. Verbs semantics and lexical selection. In Proceedings of the 32nd annual meeting on Association for Computational Linguistics, 133–138. Association for Computational Linguistics.

Xiang, Bing, Xiaoqiang Luo, & Bowen Zhou. 2013. Enlisting the ghost: Modeling empty categories for machine translation. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 822–831.

Yang, Jingxuan, Jianzhuo Tong, Si Li, Sheng Gao, Jun Guo, & Nianwen Xue. 2019. Recovering dropped pronouns in chinese conversations via modeling their referents. In Proceedings of the Annual Meeting of the Association for Computational Linguistics.

Yatabe, Shûichi. 2001. The syntax and semantics of left-node raising in japanese. In Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar, 325–344.

Yin, Qingyu, Weinan Zhang, Yu Zhang, & Ting Liu. 2016. A deep neural network for chinese zero pronoun resolution. arXiv preprint arXiv:1604.05800 .

Zhang, Hongming, Yan Song, & Yangqiu Song. 2019a. Incorporating context and external knowledge for pronoun coreference resolution. In Proceedings of the Annual Meeting of the Association for Computational Linguistics.

Zhang, Sheng, Xutai Ma, Kevin Duh, & Benjamin Van Durme. 2019b. Amr parsing as sequence-to-graph transduction. In Proceedings of ACL 2019.

Zhou, Jie, & Wei Xu. 2015. End-to-end learning of semantic role labeling using recurrent neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), volume 1, 1127–1137.

Zifonun, Gisela, Ludger Hoffmann, Bruno Strecker, & Joachim Ballweg. 1997. Grammatik der deutschen Sprache, volume 1. Walter de Gruyter.

Zwarts, Joost. 2010. Semantic map geometry: Two approaches. Linguistic Discovery 8.377–395.

——. 2015. Conceptual spaces, features, and word meanings: The case of dutch shirts. In Applications of Conceptual Spaces, 57–78. Springer.

Appendix A

Model Details

This presents the requisite details for re-implementing models presented in Chapter 5.

A.1 Additional Details and Parameters for Feature Model

This section contains additional details for the re-implementation of the models presented in Chapter 5.

A.1.1 Selectional Preference Model Details

Representations of mentions were initialized with 50-dimension embeddings from GloVE vectors (Pennington **et al.** 2014). "Role" representations were represented as predicate specific representations, either as dependencies combined with a predicate (such as "prep_across with predicate=gaze") or as arguments with roleset information, using 50-dimension embeddings. These representations were concatenated and passed to a feedforward layer (100-dimension hidden layer), passed through a tanh nonlinearity, and passed through a 100x1 layer for a final prediction score, which then was used to minimize the following loss using pseudodisambiguation and a hinge loss:

$$loss = max(0, margin + Score(predicate + role, m_{negative}) - Score(predicate + role, m_{correct}))$$

As noted in section 5.2.2, this was trained with a mixture of PropBank semantic roles and dependency roles, using all gold PropBank labels from OntoNotes (Hovy **et al.** 2006) and sampling ten instances of dependency relations from Annotated Gigaword (Napoles **et al.** 2012) for each instance of a PropBank relation. The "similarity constraint" condition mentioned relied upon a mapping between each PropBank predicate-specific role and the dependency-predicate combination that most commonly expressed that role, such as between "arg1 of break.01" and "direct object of break". In that condition, a separate loss function was added for instances of PropBank roles, minimizing the cosine distance between the PropBank role and its correlated dependency link; this provided additional transfer of selectional preference constraints to PropBank roles with little explicit training data in Ontonotes. The model was trained with Adam (Kingma & Ba 2014), with batch size of 1000, patience of 10, and for a maximum of 100 epochs, using the AllenNLP toolkit.

A.1.2 Narrative Schema Model Details

As outlined in section 5.2.3, the narrative schema model processed a sequence of event-role pairs, and produced a 200-dimensional representation of that event sequence which represented its narrative schema behavior. Each event instance was characterized by an embedding for the numbered argument and an embedding for the roleset. These were learned as a combined embedding (D=200). Experiments to augment those representations with arguments of the events did so by embedding headwords and relation terms for two arguments (D=200 each) , concatenating each argument and mention with the role, and passing them through a feedforward network (output D=200) with tanh non-linearity, and then concatenating each argument with the main role, passed through another FFNN (D=200), and a tanh nonlinearity. Regardless of the use of event arguments, the sequence of D=200 outputs was passed through a single-layer unidirectional LSTM (hidden unit=200), traversing events in order of their mention in a text. The last state of the forward LSTM was taken as the representation of that event sequence, with dropout applied (p=0.1).

This method was applied to both the correct prior sequence of events and N randomly selected other event-role sequences and was applied to the target implicit role (which, by definition, was a sequence of length one). Similarity was determined by dot-product between that target event-role pair and all other sequences. Loss minimized the margin ranking loss – for each negative example narrative-event$_{negative}$, minimizing

$$max(0, margin + Score(narrative - event_{negative}) - Score(narrative - event_{positive}))$$

, with margin=0.1. The model was trained with the AllenNLP toolkit with Adam optimizer, with batch size of 10, using ten negative examples per true instance, for a maximum of 100 epochs, using patience of 20.

A.1.3 Anaphoricity / Mention Importance Model Details

As noted in section 5.2.5, the anaphoricity models used a set of discrete features: the part of speech of the headword, its dependency relation, the lemma, and a sequence of the headwords and dependency relations of each dependent of this mention, and the lemma of the mention's head. All features were embedded (embedding dimension=20) with Dropout of 0.2 (Srivastava **et al.** 2014) and then simply averaged (using the "bag of embeddings" encoder in AllenNLP). These were passed to a FFNN layer (d=100) with a tanh nonlinearity and Dropout (p=0.1) and then passed to a final layer with output dimension of 1 for a scalar prediction. The binary cross-entropy between that prediction and the true value (being a coreference cluster or not) was minimized using Adam, with base size 32 and 100 epochs with patience of 4.

A.1.4 Salience model Details

Salience model input started with embedded representations of the headword, the part of speech of the headword, the dependency relation of the headword, the number of mentions within that mention's coreference chain, and whether that grammatical relation matches that of the current pronominal mention. Each was embedded (D=32), dropout applied (p=0.2), and that set of 5 terms was concatenated into a vector (D=150). This was passed to a FFNN (hidden dimensions = 100), with tanh nonlinearity and Dropout (p=0.4). A final 100×1 feedforward layer was applied to produce the prediction. Loss minimized binary cross-entropy (whether any given mention was the referent of the pronoun). This was optimized with Adam, using batch size of 32 and 100 epochs (patience of 10).

A.1.5 Recoverability and Interpretation Type Model Details

Recoverability models, as outlined in section 5.2.6, utilized a sequence of features for characterizing the implicit role and its lexical information, and a separate set of features for characterizing the constructional context of that role.

The lexical and semantic features of the implicit role were all represented as discrete, categorical features which were embedded into a sequence (d=50), and then reduced into a vector (d=50) by averaging the embeddings(using AllenNLP bag of embeddings encoder). Numbered argument(NARG) and PropBank roleset (PB) were added directly, but other resources were derived from lexicons and corpora. The LEX features used mappings from a given PropBank roleset to FrameNet Frame Elements (FEs) and VerbNet thematic roles, as provided in the PropBank development branch. For each argument which the PropBank semantic role mapped to, a score was added to a list regarding whether that semantic role was labeled as "core"(1.0) or "peripheral"(-1.0) in that lexicon, and the PropBank role was represented as an average over all of those mappings. Those with no mappings were represented as 0. Explicit realization information(EXPL) was calculated over the set of all OntoNotes, EWT, and BOLT PropBank data (Palmer **et al.** 2005; O'Gorman **et al.** 2018a), counting the number of instances of that numbered argument for that roleset over all instances of that roleset in the data. DEF data was calculated over that same set of corpora, measuring a set of all instances where a mention has an explicit article, pronoun, or indefinite pronoun, and dividing those instances into definite and indefinite. For all instances, the DEF ratio was the number of explicitly definite NPs over all instances with a clear definiteness encoding.

The syntactic representation of terms used a list of discrete representations – the part of speech of the predicate, the dependency role of the predicate, and the list of all dependents, as defined by dependency roles concatenated with part of speech of each dependent. These were embedded (d=50) and passed through a forward LSTM (dropout 0.3). The ELMO representation was generated as described in Model 3, using the top layer from ELMO and running a bidirectional LSTM outputting 100 dimension representation. Word representation for the LSTM model (ALL) were represented using 100-dimensional GloVE vectors (Pennington **et al.** 2014) and a positional embedding (d=5) into a bidirectional LSTM (d=100), and calculated with key-value attention (Bahdanau **et al.** 2014; Daniluk **et al.** 2017) using the representation of the semantic role, with D=200 for both the query and key and for the value vectors.

Outputs from both of those representations were concatenated, passed through a feedforward layer (D=400; dropout p=0.5; ReLU nonlinearity) and then passed to the prediction layer for the FrameNet, interpretation types and recoverability layer, applying a mask to each so that cross-entropy would only apply to the relevant predictions.

A.2 Implicit Role Resolution Models

A.2.1 Model 1 – Simple Additive Model

Model 1, by design, had very few parameters (as outlined in section 5.3.2.1). Input for N candidate antecedents came from the five simple feature-based models (each input being $N \times 1$). Each input was passed through two feed-forward layers, 1×20 and 20×1, with a softplus nonlinearity (Glorot **et al.** 2011) and Dropout (Srivastava **et al.** 2014) (with p=0.3) after the first. These five features were concatenated into a $N \times 5$ matrix. Weights for constructions were learned using a D=5 embedding layer, with each dimension corresponding to a feature. Dimensions were fed through a softmax layer to produce 5×1 matrix.

The predictions for each candidate were directly summed for each candidate into a single score. In the condition with constructional weights, this was preceded by multiplying the candidate matrix with the softmax of the constructional weights. These were fed into a feedforward network (D=100) with softplus nonlinearity and dropout p=0.3. Output was passed into a second feedforward network (d=1) for prediction.

The estimate of whether the referent was recoverable was provided by predicted distributions over the FrameNet DNI/INI labels, interpretation type prediction, and recoverability predictions discussed in section 5.2.6. These were concatenated with additional scalars associated with predicates (rate of explicitness of the semantic role, and the downsampling rate used below). These were passed to a feedforward layer(d=100) with softplus nonlinearity and dropout (p=0.3) and fed into a second FFNN (100x1) to produce a single scalar judgment of recoverability.

The loss for correct recoverability minimized the binary cross-entropy between that prediction and the correct mention. As there were multiple correct candidates, loss for detecting the actual correct candidate minimized probability weight given to all incorrect candidates. For handling the unbalanced data, instances of non-recoverable mentions were given less weight, through a parameter to down-weight the loss of non-recoverable examples (1.0 for Beyond NomBank, 0.5 for SemEval-2010-10, 0.06 for MS-AMR). All model were built with AllenNLP 0.4 (Gardner **et al.** 2018), and optimization used Adam (Kingma & Ba 2014),

A.2.2 Model 2 – Dense Model

The feature-based model discussed in section 5.3.2.2 uses a more traditional syntactic and semantic features. These features were extracted from more traditional sources of linguistic features, using dependency parses (using the AllenNLP implementation of Dozat & Manning (2016)), coreference (using the AllenNLP implementation of Lee **et al.** (2017)) and SRL (using the AllenNLP implementation of He **et al.** (2017)) and NER (using the AllenNLP implementation of the Peters **et al.** (2017) baseline model). Table A.1 outlines the features extracted from these models, mostly copied or adapted from prior feature-based models of implicit semantic role resolution.

The first set of features used characterized the predicate and implicit role, independent of the candidate, were used both for resolving role to a particular candidate but also for predicting recoverability. These overlap with the features listed in Table A.1, and were hashed into a fixed-width array (D=5000) with dropout (p=0.1) and passed through two feedforward layers with 50 hidden units and a softplus non-linearity. Features marked with an asterisk are scalar features concatenated with this output, and then both used in the resolution model, and also passed into a feed-forward layer (hidden units 100) and another 100×1 layer for the actual recoverability prediction. As with Model 1, the loss for correct recoverability minimized the binary cross-entropy between that prediction and the correct mention. For handling the unbalanced data, as in Model 1, a parameter for down-weighting non-recoverable instances was used (1.0 for Beyond NomBank, 0.5 for SemEval-2010-10, 0.06 for MS-AMR).

For resolution of correct referent, the candidate-specific features describing features between a candidate and the implicit role were hashed into fixed-width array (D=10000) with dropout p=0.1 and fed in to a FFNN (D=200, dropout p=0.3) and a tanh nonlinearity, and concatenated with all scalar features marked with an asterisk, and with the representation of the implicit role mentions above. These were passed through another feed-forward neural network (D= 300) with ReLU nonlinearity, one more layer with 20 units, ReLU and Dropout=0.3, and passed through a final 20×1 feed-forward neural network for the final prediction regarding which candidate was correct. As there were multiple correct candidates, loss for detecting the actual correct candidate minimized probability weight given to all incorrect candidates. Losses for resolution and recoverability prediction were combined when mentions were recoverable, and only the recoverability loss was used when mentions were not recoverable.

A.2.3 Model 3 – ELMO-based Model Parameters

The EMLO based model represented each sentence using the top layer of ELMO for each sentence, using the "2x4096_512_2048cnn_2xhighway" model from Allennlp 0.4 . The ELMO weights were kept frozen, and each candidate span is passed to a bidirectional LSTM (hidden unit size 100).

A predicate representation was the concatenation of the Elmo representation of the predicate (selection of the LSTM layer for the predicate) and embeddings from discrete features: a set of six embeddings of characterizations of the implicit role (PropBank numbered argument, PropBank roleset, Propbank function tag, FrameNet (when available), VerbNet (when available), and the most common dependency relation used with the missing PropBank argument), with embedding

Features for implicit role and predicate	Source
Predicate-specific features	
Predicate-specific semantic role	-
Predicate part of speech	-
Predicate dependency relation	-
Numbered argument of implicit role	-
VerbNet, FrameNet, and PropBank function tag labels for implicit role	-
Frequency of the verbal form of predicate within the document.	Gerber & Chai 2012 #16
Whether predicate is in main clause	-
Hand-crafted construction type described in section 5.5.0.2	-
Number of mentions of predicate in document*	-
all the FrameNet, Interpretation types, and recoverability predictions discussed in section 5.2.6)*	-
Explicitness of the semantic role (rate of role occurrence for particular predicate)*	-
Features for particular candidates	**Source**
Implicit role and any semantic roles of candidate chain have the same integer argument position	Gerber & Chai 2012 #8
Implicit role and any semantic roles of candidate chain are identical	-
Whether candidate and predicate are arguments of the same predicate.	Gerber & Chai 2012 #6
Dependency relation of candidate headword	-
For core roles; whether argument was in main clause of sentence	-
Whether or not the dependency relation associated with implicit role is relation held by candidate	-
Whether candidate headword lemma is same as predicate lemma	-
Number of mentions between candidate and predicate	-
Every combination of word and implicit role	Gerber & Chai 2012 #1
Frequency of candidate part of speech in explicit mentions of semantic role*	-
The five scalar features provided by the interpretable model*	-
Number of subject mentions between candidate and predicate*	-
Sentence distance from candidate to predicate*	Gerber & Chai 2012 #2

Table A.1: Features used in Dense model

dimension of d=50, and the output of the recoverability and interpretation type predictions. This totaled a 411-length vector, which was passed through a single feedforward network (d=200), followed by ReLU nonlinearity and dropout(p=0.2), followed by a second feedforward network (d=50).

As noted in section 5.3.2.3, this used two representations of mentions. A first, slightly more lightweight representation was generated from the feature models described above and discrete features. The first mention representation was generated from the same set of five features described in the interpretable model above. The embeddings of discrete features were included for the NER tag, the headword, part of speech, dependency relation, and last semantic role of the mention (embedding dimension=50). These were concatenated and passed through two feedforward layers (d=50). The dot product of the mention and the predicate array was used as a pruning score, and only the top K instances were used for the next step, using K=15.

The second, richer, mention representation was generated using ELMO vectors. All the sentences for these mentions were passed into the ELMO model (and the LSTM on top, D=100) and the spans were converted into a fixed-width representation using a self-attention model (d=100). These were combined with original mention representation layers above, and passed through a feedforward layer with 200 hidden units, dropout(0.3) and a ReLU nonlinearity, and then passed through a second nonlinearity (hidden units 100). The dot product between that new representation and the representation of the event was used as the final score.

Determination of whether an entity was recoverable was determined using not only the implicit role, but the highest-scoring implicit role candidate. The implicit role features, the score of the highest-scoring referent, and the actual features of the highest-scoring referent from the second mention model were all concatenated together. These components were passed through two feedforward hidden layers (d=100 for the first, d=1 for the second prediction layer), with ReLU nonlinearity and dropout (p=0.3) on the first layer. Loss minimized the binary cross-entropy between that prediction and the correct value. As with Model 2, since there were multiple correct candidates, loss for detecting the actual correct candidate minimized probability weight given to all incorrect candidates. For handling the unbalanced data, as in Model 1, a parameter for down-weighting non-recoverable instances was used (1.0 for Beyond NomBank, 0.5 for SemEval-2010-10, 0.06 for MS-AMR). All model were built with AllenNLP 0.4 (Gardner **et al.** 2018), and optimization used Adam (Kingma & Ba 2014).

Appendix B

Guidelines for Implicit Role Interpretation Type Annotations

The following are a set of guidelines for *implicit role interpretation type* labeling. The goal of this annotation is to start with a PropBank numbered argument that is not explicitly realized not normal PropBank annotation, and to label roughly what kind of thing it refers to, and how you might expect to find it in the text.

You can get a sense of this task by first thinking about answering questions about such unstated referents:

- "I must [**admit**] that I was surprised" Q: Who is the hearer of "admit"? A: The person being talked to
- "John is out of money. [**Sold**] the family car, even." Q: who is the seller of "Sold"? A: John
- "I was [**reading**] in the backyard." Q: What is the thing being read? A: Some unstated book/article

These are all instances of implicit roles where you know something about the semantic role that was left unstated. However, they differ in the kind of referent, and how you might find it – the first is based upon knowledge of speech (that "I must admit" means admitting to the person you are talking to), the second could be paraphrased with a pronoun and clearly refers to a prior referent, and the third is this vague unspecified thing being read. The coarse-grained version of these distinctions can be broken up into four types, and will be familiar to any annotators familiar with the FrameNet annotation of DNI and INI categories:

- Definite: A reader/audience understands what the argument refers to.

- Indefinite: The reader/audience doesn't get to know what this refers to – it's a vague, unstated thing
- In between: You know some things about the referent, but not an exact identity
- Invalid: That numbered argument doesn't exactly exist in the semantics at all.

However, we will also use an inventory of more specific categories, outlined below. The rest of this document focuses on how to annotate that more specific inventory in a consistent manner. The rest of these guidelines will contain a simple flowchart to provide an initial, very abbreviated sense of which labels to use, and then a series of more specific descriptions of each implicit role. This set of guidelines is intended to be read alongside a set of examples of each category in order to get practice doing this annotation.

B.1 Interpretation Type Flowchart

This is a flowchart for getting annotation started, but is necessarily simplified to fit into a simple flowchart: confer with the examples and definitions which follow for more detailed guidelines. It may be useful, while starting, to find the two most likely options from this chart and to consult those examples and definitions before deciding upon a final answer.

Is this semantic role valid in context, and unstated?

- YES, it's valid **and** we would label this in PropBank (in scope) → LOCAL-MENTION
- YES, it is valid and not in scope: Do you know what it refers to?
  - YES: If that because this event has already been mentioned?
    - YES → REPEATED-EVENT
    - NO: Do you know specifically who this refers to, or is this a nonspecific/general referent?
      - YES: If specific entity/proposition: is the referent a speaker or addressee?
        - YES: Label it DEICTIC
        - NO: Is this paraphrasable with a pronoun (or "this" or "that")?
          - YES: → likely SALIENT/RECENT (but consult below)
          - NO: → likely SCRIPT-INFERRABLE (but consult below)
      - NO, it doesn't refer to a particular referent: What do you know about the referent?
        - A single instance of a kind of thing; you could refer to it as *a X* → TYPE-IDENTIFIABLE
        - This refers to a general group rather than an individual. What kind of group?
          - A whole class of things, such as people in general → GENERIC
          - A group of non-specific entities → ITERATED/REPEATED/SET
- NOT REALLY: Just that it exists, nothing more. Could/Will this come up again?
  - DEFINITELY: If this is clearly going to be relevant to later text→ ESTABLISHING/CATAPHORIC

POSSIBLY It could, in some contexts, but inferred→ ARBITRARY/NONSPECIFIC
NOT POSSIBLE: this only comes to mind if it's explicitly stated→ LOW-IMPORTANCE
NO: Not valid or existent in any way. Why isn't it valid?
PREDICATE ISN'T PREDICATIVE OR EVENTIVE: → NONPREDICATIVE!
PREDICATE IS EVENTIVE: Is the role invalid in a context like this?
YES: → NOT-A-ROLE-IN-THIS-CONTEXT
NO: → OTHER-INVALID

B.2 Types of Definite Implicit Roles

B.2.1 Repeated Event Implicit Roles

Roles where it's clear that the implicit role is inferrable not because of any fancy pragmatic inference, but because the event has already been mentioned. You should look for these instances when the actual implicit role is an argument of a nominal predicate (especially a definite one), as in the "sale" below. The various arguments "sale" are just knowable because you've seen the clause with "sold":

(183) They sold the company for $400, ... The sale still needs to be approved by the FCC

You should also label an event as a REMEMBERED-EVENT instances even if the event isn't in the text, but is clearly being referred to as an event in the common ground. This will occur for the arguments of named events in particular – for example, the attacker and target of "the 9/11 attacks" are definitely known, so you would just treat them as REPEATED-EVENT mentions.

B.2.2 Deictic (i.e. Speaker/Addressee)

Deictic implicit roles, for the most part, refer to the speaker or the hearer (to be more particular, they are roles which are identifiable precisely because they refer to the speaker or the hearer.). One may also treat label other deictic references (things which are left out because they are co-present with a speaker) with this label, but can expect them to be rare in the text. Examples of these often conversational phrases or expressions of someone's interests or beliefs, such as:

(184) Look out! (arg0=you)

(185) Any suggestions? (arg0 = you, arg2 = me)

(186) This feels scratchy (arg0 of feel = to me)

B.2.3 Salient/Recent (Pronoun-like)

These refer to an antecedent that is identifiable in the recent prior discourse and which must be referred to in the recent prior text. You can usually replace these with a pronoun (he/she/it) or a demonstrative (that/this). E.g.:

(187) I talked to Bill. (he) told me he was retiring.

You should also use these to refer to the proposition under discussion too – so when you are talking about things you know, believe, or claimed, and it's clearly referring to the last-mentioned things:

(188) "[you screwed up.]$_i$" "I know $\emptyset_i$!" *arg1 of know = last proposition (screwed up)*

(189) "[We'll win in November.]$_i$ I think $\emptyset_i$, at least." *thing thought = last proposition mentioned (we'll win in November)*

Consult the note below on how to differentiate these from script-inferrable instances. As a general rule, for such SALIENT/RECENT instances, you should often be able to rephrase them with a pronoun or demonstrative.

B.2.4 Script-inferrable

This label should be used for knowable referents that aren't identifiable because of specific kinds of interpretations above, but instead are understandable because of a general understanding of how a particular argument would be filled in context. E.g., If someone robs a bank and is arrested, and you mention a "trial" event, they are probably the defendant, because we know how that sequence of events tends to work. For English, this will encompass many of the implicit roles that are knowable.

If you label an implicit role with this type, you should often be able to make a (perhaps tenuous) generalization. E.g. for the link above, you might say "people who are arrested often go on trial".

B.2.5 Edge case tests

Salient vs. Script-inferrable test: Remember that discussion about how if someone robs a bank and is arrested, and you mention a trial event, they are probably the defendant? Such a link should generally be something you would infer even across long distances. In contrast, for example 187 above – if you were to add a few sentences of unrelated statements between those two clauses, it is likely that they would change the meaning of the referent.

Deictic vs. non-deictic test: If something implicitly refers to a speaker or addressee, but only does so because they happen to be the participant being tracked, you don't want to label them with DEICTIC. For example, a sentence like "I was arrested for robbery. The trial is next week", then the defendant is "I", but it is inferrable because of one's knowledge of sequences of events, rather than because it is a first or second-person reference. In such contexts, don't label it as DEICTIC, but use the appropriate label such as SCRIPT-INFERRABLE.

B.3 Somewhat Definite mentions

B.3.1 Generics

This label should be used when referring to a whole class of things, but is most commonly used to refer to *people in general.* E.g., "Smoking is not allowed in this establishment" means something like "people can't smoke in this establishment." In most contexts, you should be able to paraphrase a generic using the indefinite pronoun "one."

Ambiguities with deictics: There are situations where it will be unclear whether an argument means *people in general* or a particular speaker or addressee, as in "that's annoying!" or "this tastes awful (to me / in general)". For the sake of consistency, whenever you can read such an ambiguous referent, then you should default to that deictic reading if it is available.

B.3.2 Type-identifiable

Another class of somewhat knowable referent is implicit roles where you have a vague referent or a vague quantity, but the type is known. In such cases, when not referring to the whole class of things but a few instances of that type, simply label them as TYPE-IDENTIFIABLE:

(190) I was reading (some kind of book)

(191) We went drinking (alcoholic beverages)

You can also use this for unstated, but relatively prototypical, extents, degrees, and domains:

(192) We went to the restaurant (arg2 = some distance)

(193) That building is really high (in terms of vertical height)

(194) I bought a new car (with some amount of money)

B.3.3 Bridging/Copied

You will likely not see many instances of this class, but if you have a referent that clearly refers to a prior entity, but makes a new *copy* of that entity, then use the "bridging-or-copied" label. E.g., in the example below, the prize won (the medal) for the second clause isn't the same medal that John won – it makes a new referent (the medal) copied from John.

(195) "John won a medal in division, and Mary won in hers too"

Use this label for any other instance where an implicit referent seems to "copy" some other referent in the discourse.

B.4 Indefinites

B.4.1 Establishing/cataphoric

Sometimes you will get mentions that are unstated, but obviously going to be resolved, as in the victim of "There's been a murder!". Another example of this would be the topic of "buzz" in example 196 below.

(196) London has been buzzing all day. The Prime Minister is going to resign. (example modified from prior literature (Hawkins 1978:102))

These clearly referential, important implicit roles are going to be separated out as establishing / cataphoric in the data. One simple test for these is that immediately after their mention, you might be able to use a construction such as "it was X" to express who the referent is.

B.4.2 Iterated/repeated/set

If something refers to a vague collection of referents, but it's a vague set – often because the event is iterated, and therefore the set is simply defined as the set of entities that fill that role – then use this general label of ITERATED/REPEATED/SET. E.g., for something like the example below, the victim (arg1) of "kill.01" is not really defined – we know that it's a vague collection of referents, but those referents are really only defined by the semantic role itself (the set of people that were killed).

(197) He killed his way across Europe.

B.4.3 Arbitrary/Nonspecific

If you don't know *anything* about the referent other than the fact that it exists, and it's not clearly encoded as an ongoing participant (those establishing/cataphoric roles), just mark it as arbitrary/nonspecific.

B.5 Invalid Semantic Roles

Finally, there are situations where we basically want to say that these arguments aren't really inferred at all.

B.5.1 Local mention

If you think that this numbered argument would actually be labeled in PropBank – i.e., it has a valid local referent – then it's invalid that it's listed in this task at all! Mark this as a "local mention". These guidelines currently assume prior PropBank training, but one may wish to consult Bonial **et al.** (2010) for information about PropBank definitions of "in scope".

B.5.2 Contextually Invalid (not-a-role-in-this-context)

A sentence that says "the great plains are vast and open" might prompt an annotator for the agent of opening, because we use the same sense for stative open and causative verbal *to open*. The annotation may prompt you to label such an agent, but it is not an argument which makes any sense on context, as "open" in that situation is stative and can't possibly have an explicit causer. We want to label that this argument as NOT-A-ROLE-IN-THIS-CONTEXT

B.5.3 Non-predicative

Sometimes an argument is invalid because the predicate itself isn't really something that should have arguments. E.g., in a context such as "the sales associate will help you", one perhaps does not want to model that instance of "sales" as being a predicative event of selling. Even with abstract usages like this, use this only when the usage is so non-predicative that it doesn't seem appropriate to use something like GENERIC.

B.5.4 Low-importance

Finally, some arguments could be argued to be mentionable in a particular context, but are really not inferrable at all except when they are explicitly stated. For example, the PropBank lexicon representation of "work.01" has a "co-worker" role. This may be important in some contexts when explicitly stated (such as "I worked with Sam"), but it is unlikely that one would attempt to infer any kind of coworker argument from a mention of the verb work. In such cases where the implicit role is simply very unlikely to be inferred, feel free to use LOW-IMPORTANCE. In particular, you should often use this for *benefactive* roles unless they are very clearly implied by an event – although it is easy to attempt to stretch ones causal reasoning and discern a benefactive for any given mention, we should avoid doing so unless it is a clear argument of the event.

B.6 Importance of Reviewing Prior Annotations

As a final note, it should be emphasized again that these guidelines are intended to provide a basic understanding of the annotations, but the training for this task also requires that an annotator review a set of prior decisions of implicit role annotations in order to gain consistency with prior work.