

ABSTRACT

Title of Dissertation: ARE YOU ASKING ME OR TELLING ME?
LEARNING CLAUSE TYPES AND SPEECH
ACTS IN ENGLISH AND MANDARIN

Yu'an Yang
Doctor of Philosophy, 2022

Directed by: Professor Valentine Hacquard
Department of Linguistics

Professor Jeffrey Lidz
Department of Linguistics

Languages tend to have three major clause types (declaratives, interrogatives, imperatives), dedicated to three main speech acts (assertions, questions, commands). However, the particular forms that these clause types take differ from language to language, and have to be learned. Previous experimental results suggest that by 18 months old, children differentiate these clause types and associate them with their canonical speech act. This dissertation investigates how children learn to identify different clause types and speech acts.

To learn clause types, children need to identify the right categories of clauses (the “clustering problem”) and figure out what speech act they are canonically used for (the “labeling problem”). I investigate the extent to which learners need to rely on pragmatic information (i.e., knowing what speech act a given utterance of a sentence is conveying), to solve not just labeling, but the clustering itself. I examine the role of pragmatics computationally by

building two Bayesian clustering models. I find that morpho-syntactic and prosodic information are not enough for identifying the right clause type clustering, and that pragmatics is necessary. I applied the same model to a morphological impoverished language, Mandarin, and found that the model without pragmatics performs even worse. Speech act information is crucial for finding the right categories for both languages. Additionally, I find that a little pragmatics goes a long way. I simulate the learning process with noisy speech act information, and find that even when speech act information is noisy, the model hones in on the right clause type categories, when the model without fails.

But if speech act information is useful for clause type learning, how do children figure out speech act information? I explore what kind of non-clause type cues for speech act information are present in the input. Even if children must rely on clause type information to figure out speech acts, they could have access to additional information that is unrelated to clause typing, but informative for recognizing speech act type. When speakers perform speech acts, because of the conventional functions of these speech acts on the discourse, the performance might be associated with certain socio-pragmatic features. For example, because of questions' response-elicitation function, we might expect speakers to pause longer after questions. If children are equipped with some expectations about the functions of communication, and about what questions do, they might be able to use these socio-pragmatic cues to figure out speech act.

I explore two cues that could potentially differentiate questions from other speech acts: pauses, and direct eye gaze. I find that parents tend to pause longer after questions, and attend to the child more when asking questions. Therefore it is in principle plausible that there are some socio-pragmatic features that children can use, in addition to their growing knowledge of clause types to infer the speech act category of an utterance. This little bit of information about speech act could then be used to provide the information that the child needs in order to get the clause type clusters identified accurately.

ARE YOU ASKING ME OR TELLING ME? LEARNING CLAUSE
TYPES AND SPEECH ACTS IN ENGLISH AND MANDARIN

by

Yu'an Yang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:

Professor Valentine Hacquard, Co-chair/co-advisor

Professor Jeffrey Lidz, Co-chair/co-advisor

Professor Yi Ting Huang, Dean's Representative

Professor Naomi Feldman

Professor Alexander Williams

Dr. Daniel Goodhue

© Copyright by
Yu'an Yang
2022

Preface

The work reported in this dissertation is highly collaborative. Chapter 3 and Chapter 4 report on joint work with Naomi Feldman, Valentine Hacquard, and Jeffrey Lidz; Chapter 5 reports on joint work with Thomas Schatz, Naomi Feldman, Valentine Hacquard, and Jeffrey Lidz; Chapter 6 reports on joint work with Daniel Goodhue, Valentine Hacquard, and Jeffrey Lidz. This work has also been supported by the National Science Foundation (Doctoral Dissertation Research Improvement grant #2140764 and NRT award #1449815).

Dedication

To my parents, Chang Yang and Dr. Jun Liu, for their unwavering support.

Acknowledgments

To be added.

Table of Contents

Preface	ii
Dedication	iii
Acknowledgments	iv
Contents	xiii
List of Tables	xvi
List of Figures	xvii
List of Abbreviations	xx
Chapter 1 Introduction	1
1.1 Learning matrix clause types	5
1.1.1 The clustering and labeling problem	5
1.1.2 The usefulness of speech act information	8
1.1.3 The problem with speech act information	10
1.1.4 Learning clause types: Interim summary	11
1.2 Learning speech act categories	13
1.3 Summary of the learning problem	16
1.4 Roadmap	17
Chapter 2 Background	20
2.1 Theories of clause types, sentential force, and speech acts	20
2.1.1 Speech acts and force	20
2.1.2 Clause types	28
2.1.3 The role of prosody	30
2.2 Acquisition of speech acts and clause types	34
2.2.1 Early communicative abilities	34
2.2.2 Early knowledge of clause types and speech acts	37
2.3 Summary	41
Chapter 3 Learning to identify clause types in English	42
3.1 Background	44
3.1.1 Two learners	44
3.1.2 Clause types in English	46
3.1.3 Linguistic knowledge and capacities of 18-month-olds	53

3.2	Corpus study	55
3.2.1	Corpus and methods	56
3.2.2	Annotation schema	57
3.2.2.1	Clause Type	57
3.2.2.2	Speech Act	58
3.2.2.3	Formal features	59
3.2.3	Results	60
3.2.3.1	Overview	60
3.2.3.2	Morpho-syntactic cues	66
3.2.3.3	Informativeness of syntactic and pragmatic cues	68
3.3	Modeling the learning of clause types	72
3.3.1	Generative models	73
3.3.2	Inferences	75
3.3.3	Simulations with infant-directed speech	77
3.3.3.1	The SID model	78
3.3.3.2	The SPID model	81
3.3.3.3	Interim Summary	84
3.3.4	Simulating noisy pragmatic information	84
3.4	Discussion	88
3.4.1	Implications	89
Chapter 4 Learning to identify clause types in Mandarin		92
4.1	Background	93
4.1.1	Clause types in Mandarin	93
4.1.2	Mandarin-acquiring children’s knowledge of clause types	98
4.2	Corpus study	99
4.2.1	Corpus and methods	99
4.2.1.1	Clause Type	100
4.2.1.2	Morpho-syntactic features for clause typing	101
4.2.2	Results	103
4.2.2.1	Overview	103
4.2.2.2	Morpho-syntactic cues	109
4.3	Modeling the learning of clause type in Mandarin	109
4.3.1	Performance of the SID model	110
4.3.1.1	The conservative simulation	110
4.3.1.2	The knowledgeable simulation	112
4.3.2	Performance of the SPID model	114
4.3.2.1	The conservative simulation	114
4.3.2.2	The knowledgeable simulation	117
4.3.2.3	Compare all four simulations	120
4.3.3	Simulations with noisy pragmatic information	121
4.4	Discussion	122
Chapter 5 Prosody, clause types, and speech acts in English		125
5.1	Background	126

5.1.1	Prosody, clause types, and speech acts in English	126
5.1.2	Children’s knowledge of prosodic features	129
5.2	Corpus study	131
5.2.1	Methods	131
5.2.2	Results	133
5.3	Learning clause type categories with prosody	136
5.4	Discussion	141
Chapter 6	Learning speech acts	143
6.1	Background	146
6.1.1	Pauses in conversation	146
6.1.2	Attentional behavior	148
6.1.3	Interim Summary	148
6.2	Corpus study	149
6.2.1	Methods	149
6.2.2	The correlations we expect	150
6.2.3	Results	150
6.2.3.1	Pauses	150
6.2.3.2	Parents’ attentional behavior	152
6.3	Conclusion	153
Chapter 7	Conclusion and discussion	154
7.1	Summary of findings	154
7.2	The pragmatic syntactic bootstrapping hypothesis	157
7.3	Conclusion	159
Bibliography	161

List of Tables

2.1	Clause types and their conventionalized force; adapted from Portner (2004:238)	28
2.2	Languages that use morpho-syntactic and/or prosodic features to differentiate interrogatives and declaratives	31
2.3	Intonational markings used to differentiate polar interrogatives and declaratives (adapted from Butler, Frota, & Vigário 2012)	33
3.1	Morph-syntactic features and their associated clause types	52
3.2	Morpho-syntactic cues and their examples	60
3.3	Morpho-syntactic cues associated with interrogatives and imperatives compared to declaratives, according to the supervised SPID model	71
3.4	Variables in syntactically informed distributional learner model, their distribution, and explanation	74
3.5	Variables of the SPID model, their distributions, and explanations	74

List of Figures

1.1	F0 of declarative and interrogative sentences in Brazilian Portuguese (p.c. Jéssica Mendes)	3
2.1	Waveforms and pitch curves of a declarative (right) and an interrogative by a female Akan speaker (Figure 7 in Genzel & Kügler (2020))	32
3.1	Extended structure of CP proposed by Farkas & Roelofsen (2017)	49
3.2	The structure of rising declaratives, adapted from Farkas & Roelofsen (2017) . .	49
3.3	Distribution of clause types in the corpus	61
3.4	Subcategories of interrogatives	62
3.5	Distribution of speech acts in the corpus	62
3.6	The speech acts performed by each clause type in parents’ speech	63
3.7	The clause type used to express each speech act in parents’ speech	64
3.8	The speech act expressed by different subcategories of interrogatives	65
3.9	Number of sentences with/without various formal cues in each clause type; darker colors represent number of sentences with the cue, lighter colors, number of sentences without the cue	67
3.10	Comparing the two supervised models by rand score (performance over 10 iterations)	70
3.11	The syntactically informed distributional learner (left) and syntactically and pragmatically informed distributional learner (right)	73
3.12	Comparing all four learners (distributional, pragmatic distributional, supervised distributional, supervised pragmatic) by adjusted rand score	78
3.13	The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SID model	79
3.14	The proportion of actual declaratives, interrogatives, and imperatives clustered in one category	79
3.15	Number of sentences with/without certain formal features in each cluster (Cluster 1 ~ Interrogatives, Cluster 2 ~ Imperatives, Cluster 3 ~ Declaratives), darker colors represent the number of sentences with the feature. UFI stands for Unknown Functional Item (e.g. <i>wh</i>), see Table 3.2 for details.	80
3.16	The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SPID model	82
3.17	The proportion of actual declaratives, interrogatives, and imperatives clustered in one category	82

3.18	Number of sentences with/without a formal property in each cluster (Cluster 1 ~ Interrogatives, Cluster 2 ~ Declaratives, Cluster 3 ~ Imperatives), darker colors represent the number of sentences with the property, lighter colors represent sentences without the property.	83
3.19	Performance of the SPID model with different levels of noise in the speech act information; dotted line indicates the adjusted rand score of the SID learner . . .	85
3.20	Proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SPID model; with 70% (left) and 80% (right) noise in speech act information	86
3.21	Proportion of actual declaratives, interrogatives, and imperatives clustered in one category, with 70% (left) 80% (right) noise in speech act information	86
3.22	Morpho-syntactic profile of each cluster in simulations with 80% noise in speech act information (Cluster 1 ~ Imperatives, Cluster 2 ~ Interrogatives, Cluster 3 ~ Declaratives).	87
3.23	Mmorpho-syntactic profile of each cluster in simulations with 70% noise in speech act information (Cluster 1 ~ Interrogatives, Cluster 2 ~ Imperatives, Cluster 3 ~ Declaratives).	88
4.1	Pitch contour associated with a <i>wh</i> -interrogative sentence	96
4.2	Pitch contour associated with a <i>wh</i> -indefinite sentence	96
4.3	Distribution of clause types in the corpus	103
4.4	Subcategories of interrogatives	104
4.5	Subcategories of imperatives	105
4.6	Distribution of speech acts in the corpus	106
4.7	The speech acts performed by each clause type in parents' speech	107
4.8	The clause type used to express each speech act in parents' speech	107
4.9	Number of sentences with/without various formal cues in each clause type; darker colors represent number of sentences with the cue, lighter colors, number of sentences without the cue	109
4.10	The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SID model	110
4.11	The proportion of actual declaratives, interrogatives, and imperatives clustered in one category	111
4.12	The number of sentences with/without certain formal features in each cluster, darker colors represent the number of sentences with the feature.	112
4.13	The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SID model	113
4.14	The proportion of actual declaratives, interrogatives, and imperatives clustered in one category	113
4.15	The number of sentences with/without certain formal features in each cluster, darker colors represent the number of sentences with the feature.	114
4.16	The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SPID model	115
4.17	The proportion of actual declaratives, interrogatives, and imperatives clustered in one category	116

4.18	The number of sentences with/without a formal property in each cluster (Cluster 1 ~ imperatives, Cluster 2 ~ declaratives, Cluster 3 ~ interrogatives), darker colors represent the number of sentences with the property, lighter colors represent sentences without the property.	117
4.19	The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SPID model	118
4.20	The proportion of actual declaratives, interrogatives, and imperatives clustered in one category	118
4.21	The number of sentences with/without a formal property in each cluster (Cluster 1 ~ declaratives, Cluster 2 ~ imperatives+declaratives, Cluster 3 ~ interrogatives), darker colors represent the number of sentences with the property, lighter colors represent sentences without the property.	119
4.22	Performance of conservative and knowledgeable SID and SPID	120
4.23	Performance of the conservative SPID model with different levels of noise in the speech act information; dotted marks the rand score of the SID learner	121
4.24	Performance of the knowledgeable SPID model with different levels of noise in the speech act information; dotted marks the rand score of the SID learner	122
5.1	Two types of frequent rhetorical polar interrogatives produced by a female (up) and male (down) participant with contexts similar to (80) (Figure 11 from Dehé & Braun 2020)	128
5.2	“You think it’s a birdy?”	133
5.3	Proportion of utterances with final rise across different speech acts	134
5.4	Proportion of utterances with final rise across different clause types	135
5.5	Proportion of declaratives, <i>wh</i> and polar interrogatives with final rise	136
5.6	The syntactically informed distributional learner model with prosody (left) and syntactically and pragmatically informed distributional learner model with prosody (right)	137
5.7	Performance of all four models as measured by adjusted rand score	138
5.8	The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SID with prosody (left) and the SPID model with prosody (right).	139
5.9	The proportion of actual declaratives, interrogatives, and imperatives clustered in one category as identified by the SID+prosody model (right) and SPID+prosody model (left)	139
5.10	Morpho-syntactic profile of each cluster identified by the SID+prosody model (Cluster 1 ~Interrogatives , Cluster 2 ~ Imperatives, Cluster 3 ~ Declaratives).	140
5.11	Morpho-syntactic profile of each cluster identified by the SPID+prosody model (Cluster 1 ~ Declaratives , Cluster 2 ~ Interrogatives, Cluster 3 ~ Imperatives).	140
6.1	Upper face action units (AU) identified by participants as indicating the act of questioning, as reported by Domaneschi, Passarelli, & Chiorri (2017)	145
6.2	Duration of pause (ms) after each type of speech act	151
6.3	Proportion of parents’ looks to the child before, during, and after an utterance	152

List of Abbreviations

Chapter 1: Introduction

We can use language to do a great many things, such as warning, arguing, guessing, pleading, naming or declaring. But from the point of view of linguistics, three particular kinds of speech acts appear to stick out. These are the broad classes of assertions, questions and commands (or requests). They are special because languages tend to have dedicated three clause types to the performance of these three kinds of speech acts (Sadock & Zwicky 1985, König & Siemund 2007, Aikhenvald 2016, Portner 2018, see König 2020 for a recent review). Declaratives are typically used for assertions (1a) and (2a), interrogatives for questions (1b) and (2b), and imperatives for commands (1c) and (2c):

(1) English clause types:

- | | |
|------------------|-------------------------|
| a. That's Elmo. | Declarative, Assertion |
| b. Is that Elmo? | Interrogative, Question |
| c. Find Elmo! | Imperative, Request |

(2) Mandarin clause types:

- | | |
|--|------------------------|
| a. Zhe shi Elmo.
This is Elmo
“This is Elmo.” | Declarative, Assertion |
| b. Zhe shi Elmo ma ?
This is Elmo SFP
“Is that Elmo?” | Interrogative |
| c. Zhizhi Elmo!
Point Elmo
“Point at Elmo!” | Imperative |

While the functions stay constant, the forms vary from language to language. As the above examples in English and Mandarin show, both English and Mandarin have declaratives, interrogatives, and imperatives to perform the functions of asserting, questioning, and commanding, respectively. But the form of each clause type in these two languages differ. For example, if we compare the interrogative clauses in the two languages with the declarative clauses, we can see that the English interrogative clause has a different word order than the English declarative clause, with the subject and the auxiliary switching places in the interrogative clause. In Mandarin, the difference arises at the edge of the sentence, as the interrogative clause has an additional sentence final particle *ma* at the end of the sentence. In Mandarin, one can also use A-not-A constructions for polar interrogatives:

- (3) Zhe **shi-bu-shi** Elmo?
 This be-NEG-be Elmo
 “Is this Elmo?” A-not-A Interrogative

The canonical function of such sentences is to perform a questioning speech act, just like its English counterpart (1b), but interrogativity in this example is marked by the presence of the disjunctive negative structure (i.e. *shi-bu-shi*).

Besides using relative word order of constituents (English), particles (Mandarin *ma*), or disjunctive negative structures (Mandarin A-not-A), we also find languages like West Greenlandic that differentiate the two clause types by verb inflection:

- (4) West Greenlandic
- a. neri- vutit
 eat- IND.2SG.PST
 “You ate.” Declarative
- b. neri- vit
 eat- INT.2SG.PST
 “Did you eat?” Interrogative

König & Siemund 2007:18, ex (50)

As (4) shows, West Greenlandic uses the suffix *vutit* for declaratives and *vit* for interrogatives.

Morpho-syntactic features are not the only way to differentiate clauses in languages. In languages like Italian and Portuguese, declaratives and interrogatives share the same morpho-syntactic features and only differ in intonation, as shown in Figure 1.1:

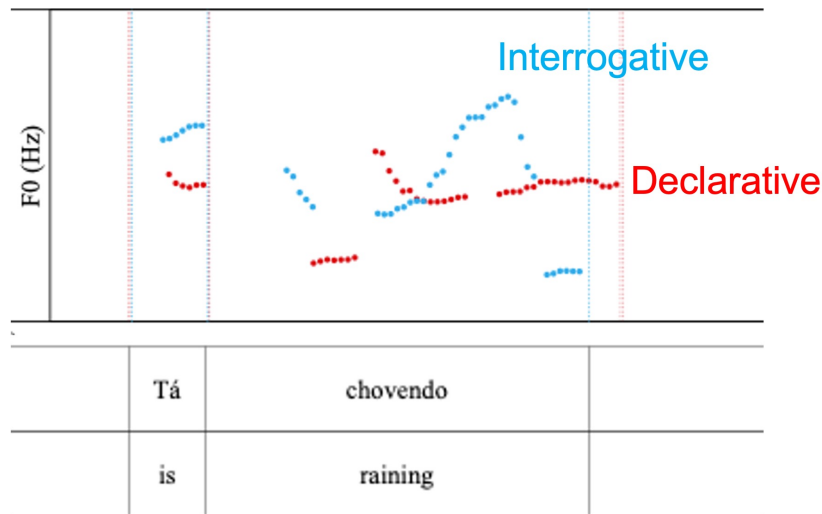


Figure 1.1: F0 of declarative and interrogative sentences in Brazilian Portuguese (p.c. Jéssica Mendes)

In Portuguese, interrogative and declarative clauses can be string-identical, and the only difference lies in their pitch contour (see Truckenbrodt 2009 for discussion on a detailed discussion of the role that intonation played in differentiating clause types and speech acts in Portuguese).

For *wh*-interrogatives, languages differ in whether the *wh*-phrase is fronted to the beginning of a clause, like English (5), or stays in situ, like the Mandarin sentence in (6):

(5) **What** did Elmo eat?

(6) Elmo chi-le **shenme**?
 Elmo eat-ASP what
 “What did Elmo eat?”

Despite all these cross-linguistic differences in how clause type information is encoded in the surface form, we can still see that in each language, we find these three major clause types (declaratives, interrogatives, and imperatives) that are canonically associated with the same three major speech acts (assertions, questions, and requests/commands).

People refer to this mapping between form and meaning as *sentential force* or *Mood* (Chierchia & McConnell-Ginet 1990, Portner 2018). This category is systematically related to clause types, i.e. the forms of sentences, on the one hand, and speech acts, i.e. the pragmatic functions that the sentences are used to perform, on the other. In this dissertation, I will mostly consider the learning of *matrix* clauses and their associated sentential force: how children come to associate matrix interrogatives with questioning force, matrix declaratives with assertive force, and matrix imperatives with requesting force.

While each of the three major clause types bears a canonical association with one of the major speech acts, this mapping is not inviolable. One can use a sentence with a particular clause type to perform a speech act other than the canonically associated one. For example, when we use the sentence *Can you pass the salt?* at the dinner table, then, even though the sentence is clearly an interrogative in English with subject-auxiliary inversion, the speech act performed by the speaker is best characterized as a requesting act. The same is true in Mandarin:

- (7) Keyi qing ni bang wo diyixia zhijin **ma**?
Can please you help me pass napkin SFP
“Can you please pass me the napkin?”

Sentences with the particle *ma*, as discussed above, are interrogatives in Mandarin. But just like its English counterpart, when you utter (7) to someone sitting next to the napkin box, you are making a request rather than asking a question.

Thus, although languages tend to have dedicated clause types that are typically associated with particular speech acts, these clause types can be used for other speech acts.

As adults, we effortlessly understand this canonical association between clause types and

speech acts – and when it is violated. When we hear someone utter “Is it raining?” we assume that they are asking a question, and when we hear someone says “It’s raining!” we assume that, all else equal, they are making an assertion. But for a child whose grammar is still developing, this might not be trivial, as they have to figure out the makeup of the declaratives, interrogatives, and imperatives in their language and associate them with their canonical speech acts.

Remarkably, children seem to have figured this out from a young age. By 18 months, they seem to be able differentiate interrogatives from declaratives, and understand that people use interrogatives to ask questions, all while their grammar and understanding of the world are still in development (Geffen & Mintz 2011, 2015, Casillas & Frank 2017, Perkins 2019, Goodhue, Hacquard, & Lidz 2022 among others). In this dissertation, I examine how children can figure out clause typing and how they learn to associate the three major clause types with the three major speech acts.

In the rest of this chapter, I first explain in detail the two learning problems related to clause type categories (Section 1.1.1). In Section 1.1.2, I explore the relevance of information about speech acts to the task of learning clause types. Then, in Section 1.1.4, I detail my plan for probing the question of how children might figure out the right clause types, and in Section 1.2, I address the remaining issue of learning speech acts. Finally, in Section 1.3 I summarize the different aspects of the problem and lay out the roadmap of this dissertation.

1.1 Learning matrix clause types

1.1.1 The clustering and labeling problem

As we have seen, languages tend to have three matrix clause types. Given the near-universality of these main clause types, it may be reasonable to assume that children expect that their language is likely to have three main clause types. But even if we assume that the knowledge of clause type categories is innate, learners still face two main problems.

First, input sentences do not wear their clause type categories on their sleeves. Rather, learners need to identify the specific signals in the surface form of the sentences in their language that are associated with the three clause types. That is, they need to identify the right formal properties of sentences that allow them to categorize a sentence with all other sentences of the same clause type. This is the **clustering problem**. To use English as an example, English-acquiring children have to figure out that (8) shares its clause type with (9b), even though they have different lexical items, because the subjects in both follow the auxiliaries; but also that (8) has a different clause type than (9b) even though both share the same lexical items, because the subject in the latter precedes the auxiliary. The learner needs to recognize that subject-auxiliary inversion is a formal feature that is relevant for clause typing.

(8) Do you want a cookie?

(9) a. Is that Elmo?

b. That's Elmo!

Second, after identifying the clusters, learners need to determine the canonical function of each cluster in the system. That is, after clustering sentences into three categories, children still need to learn which one of these clusters is the interrogatives (typically used to perform questions), which is the declaratives (typically used to perform assertions), and which is the imperatives (typically used to perform commands). This is the **labeling problem**.

Solving these two problems is neither trivial nor straightforward. For the clustering problem, each clause type category could be related to a variety of surface forms, none of which is obligatorily present and many of which can occur in sentences with a different clause-type category. For example, as we have discussed, in English, the hallmark of interrogativity is subject-auxiliary inversion, which can be seen in polar interrogatives (9a) and *wh*-interrogatives (10) below. However, this association of word order and interrogativity has many exceptions. Subject *wh*-interrogatives like (11) do not have this formal feature.

(10) Who *can* **Sue** hug? Object *wh*, Subject-Auxiliary Inversion

(11) **Who** *can* hug Ann? Subject *wh*, no Subject-Auxiliary Inversion

Conversely, some morpho-syntactic properties typically associated with interrogative clauses could also appear in other settings. For example, in English, some declaratives exhibit subject-auxiliary inversion:

- (12) a. Mary would never eat tripe in her life.
b. Never in her life would Mary eat tripe.

In English, fronting the negator *never* can cause the raising of an auxiliary, giving the sentence an appearance of subject-auxiliary inversion, but the clause type category is not interrogative.

Therefore, learners need to infer the right clause type category of sentences they hear in the input, but they might not see the crucial surface morpho-syntactic features for clause typing, or the surface features that they do see misalign with the actual clause type category of the sentence. A particular formal feature can, in principle, occur in sentences having different clause types, and sentences having a particular clause type can have different formal features. Thus, even a learner who is able to, for example, recognize subject-auxiliary inversion might have difficulty to recognize its role for determining interrogativity.

But this many-to-many mapping problem is not the only challenge for learners to face when solving the clustering problem. Learners also have to deal with cases where the relevant morpho-syntactic features are masked, for instance, by other syntactic operations. For example, left-edge-ellipsis is such an operation (Zwicky & Pullum 1983):

(13) Want to go out

The string in (13) could result from eliding the subject pronoun from a declarative clause like (14a), or result from eliding the subject and the auxiliary from a polar interrogative

like (14b). But the surface form of (13) itself does not have enough information to help us identify its clause type feature.¹

- (14) a. You want to go out?
b. Do you want to go out?

The problem is even more prominent in languages like Mandarin where the morphological cues are scarce. For example, [– subject] is a feature of imperatives in Mandarin, like in English, but since Mandarin is a pro-drop language, subjects are frequently elided, regardless of the clause type. As a result, it is likely that a learner observes more subject-less non-imperative sentences than imperative sentences.

Thus, even if we assume that learners come with the expectation that their language is likely to have three clause types, they still have to figure out the right clustering of the clauses (the clustering problem), and label the clusters with the canonical functions of the clauses (the labeling problem). They might face many challenges when solving the two problems, as the formal features for one clause type category might not show up in the surface string, or might appear in sentences of another category. Could children learn clause typing from surface features alone?

As already mentioned, cross-linguistically, we see a robust association of the three major clause types with the three major speech acts. Might this association itself be a useful source of information to the learner?

1.1.2 The usefulness of speech act information

At the beginning of this chapter, we saw that cross-linguistically, declaratives are canonically mapped to assertions, interrogatives to questions, and imperatives to commands. Could exploiting this mapping potentially help the child learn each side of the learning problem: i.e., what the clause types of their language is, and what speech acts speakers perform?

¹Note that the intonation would not help in this case either, because both (14a) and (14b) are likely to have a final rising contour L* H-H% (Gunlogson 2008, Jeong 2018, Goodhue 2021).

Let's consider this hypothesis and evaluate how pragmatics might help a learner to solve the clustering and labeling problems for clause types.

Clearly, having access to some speech act information is necessary to solve the labeling problem: to understand that a particular clause type is declarative (i.e. typically used for asserting) and another is an interrogative (i.e. typically used for questioning), it is not just useful but indeed *necessary* for a learner to be able to recognize some speech act information. Absent such information, there would be no way to associate a particular cluster with a particular speech act type.

As for the clustering problem, surface formal features alone may suffice to allow a learner to cluster sentences into three distinct formal categories. However, as we discussed in the last section, the formal features for one clause type category might not show up in the surface string, or might appear in sentences of another category. The question then is, is the surface formal features of the sentences that the learners observe in their input sufficient for identifying the three clause types? If not, what information might bridge this gap between learners' input and the abstract clause type categories they need to acquire? That is, what information might help learners bootstrap into the abstract clause type and sentential force categories (cf. Pinker 1984, Gleitman 1990, Hacquard & Lidz 2018)?

The obvious candidate for a source of information that goes beyond formal features is again information about the speech act being performed. We have seen that cross-linguistically, the three clause type categories are systematically related to three speech acts. If learners know that there are three clause types that are associated with three speech acts and are able to observe that sentences with subject-auxiliary inversion predominantly appear in sentences that are used for questioning, this might help them to recognize the relevance of subject-auxiliary inversion for clause typing. The same can be said about other formal features. Thus, despite the mapping between the three major clause types and the three major speech acts being many-to-many, learners could take advantage of pragmatic information related to speech act type to fill in the gaps left by surface formal features in

the input.

1.1.3 The problem with speech act information

Tracking speech acts to infer clause types could also hinder children’s learning clause types, again because this mapping too is many-to-many. In some contexts, it is possible that the conventionalized speech act associated with a clause type is not the actual speech act performed by uttering it. **Indirect speech acts** are these mismatching cases where the primary, “non-literal” speech act of an utterance is different from the conventionalized, “literal” speech act of a sentence associated with its clause type (Searle 1975a, 1976, Bach & Harnish 1979, Searle & Vanderveken 1985, Portner 2004, Starr 2014, Portner 2018, Murray & Starr 2020 a.o.). As we have discussed briefly at the beginning of the chapter, when you utter *Can you pass the salt?* at the dinner table, it is likely that you intent this utterance to be taken as a request. As an interrogative clause with subject-auxiliary inversion, its conventionalized (and “literal”) act is questioning, but the primary act performed is requesting. As a result, some speech act categories can be expressed by more than one clause types, and vice versa. For example, interrogatives can be used to express assertions, questions, requests/commands. You can use the interrogative sentence in (15b) to assert that you are a nerd;² or you can use the interrogative sentence in (15c) to request your friend to pass the salt to you.

(15) Speech acts expressed by interrogatives

- | | |
|--------------------------------|-----------|
| a. Is it snowing? | Question |
| b. Am I a nerd or am I a nerd? | Assertion |
| c. Can you pass the salt? | Request |

²This example is reported in Maria & Kyle 2017, ex. (2a). While some like Maria & Kyle (2017) argue that rhetorical questions still preserve discourse functions as questions, but even so, they agree that rhetorical questions share similarities with assertions. But in this particular case, even the discourse function of questions (i.e. as response elicitation device) seems to be absent: it does not require any responses, and giving one seems extremely marked.

Conversely, questions can be expressed by declaratives, interrogatives, and imperatives. You can utter any of the sentences in (16) to perform a questioning act.

(16) Clause types expressing questions

- | | |
|------------------------------|---------------|
| a. Is it raining? | Interrogative |
| b. I wonder if it's raining. | Declarative |
| c. Tell me if it's raining! | Imperative |

If such mismatching cases are prevalent in children's input, speech act information might not be helpful for children to figure out clause types. Just like the problem with formal features was that they might not appear to be uniquely associated with a single clause type, the problem with speech act types is that they are also not uniquely associated with a single clause type.

1.1.4 Learning clause types: Interim summary

To summarize our discussion so far, we have seen that languages tend to have three clause types dedicated to three speech acts, and as we will see in the next chapter, by 18 months old, children seem to be able to differentiate these clause types and associate them with their canonical speech act. To gain this ability, they need to identify the right categories of clauses (the clustering problem) and figure out what speech act they are canonically used for (the labeling problem). To solve the labeling problem, some speech act information must be available, but if there are too many mismatching cases between speech acts and clause types, this information might not be useful after all. To solve the clustering problem, children need to pay attention to the surface morpho-syntactic features of each sentence in their input. But in the input, the surface features might be absent or misleading. Again, it seems plausible that speech act information can help, but this information too is potentially misleading.

This dissertation therefore investigates these questions: first, are the surface formal features of the sentences in the input sufficient for children to figure out the clustering of

clause types? Second, if not, for both the clustering and the labeling problem, is speech act information helpful or hurtful?

I compare two learners, a syntactically informed distributional learner (SID), and a syntactically and pragmatically informed distributional learner (SPID). Both learners use the surface morpho-syntactic features of the input sentences to attempt to cluster sentences into three categories, i.e. to learn the clause type categories. But the SPID additionally has access to information about which speech act is performed by a sentence. In Chapter 3, I simulate these two learners with Bayesian clustering models on English data. By comparing the performance of the two models, I show that morpho-syntactic features alone are not sufficient to learn the three clause types, but adding information about speech act type suffices to learn the clause types, even if the speech act information is noisy. I conclude that speech act information is helpful, indeed crucial, to solve the clustering problem. Chapter 5 shows that adding prosodic information does not change the two models performance, pragmatic information is still needed for finding the right clustering.

In Chapter 4, I test the models' performance for learning Mandarin clause typing, where the learners need to use a different set of surface features for clause typing. Moreover, due to its impoverished morphological system, the surface formal features for clause typing might be even more likely to be absent or misaligned. The results from the two models suggest that the surface formal features in Mandarin are even less informative for clause typing than in English; without speech act information, the learner might not be able to identify any clause types. In Chapter 5, I add prosodic features to the cues that the models have to learn clause typing, and find that the results remain the same.

This insufficiency of syntax leads me to the following **pragmatic bootstrapping hypothesis**:

Infants use speech act information, in addition to observations of morpho-syntactic and prosodic features in the surface form of sentences, to cluster and label input sentences into the three major clause types.

This hypothesis states that to compensate for the insufficiency of surface formal features, children need to use speech act information to bootstrap into clause type categories.

1.2 Learning speech act categories

At first glance, it appears that for the pragmatic bootstrapping hypothesis to work, we have to assume that in order to have figured out the clause type categories, children need to have obtained speech act information. But this might catch us in a chicken-and-egg problem.

While some evidence suggests that 18-month-olds can infer speech acts, especially questions (Casillas & Frank 2017, Goodhue, Hacquard, & Lidz 2022), their inferences might not be perfect, so they might misidentify some or many speech acts. That is, children might have only limited access to speech act information. Moreover, there is the problem of how children can infer speech act categories in the first place – and it is undeniable that clause type information is useful for solving this problem. As adults, the primary way we identify the speech act performed by a given sentence is through its clause type. But this is precisely the problem that the child is trying to solve (i.e. identifying the clause type). If children need speech act information to learn to identify clause type categories, but they also need clause type information to identify speech act categories, it seems that we have a chicken-and-egg problem.

This way of putting the problem makes it seem like children have to either first learn to recognize speech acts and, having done this, move on to learn to recognize clause types, or first learn to recognize clause types and, having done this, move on to learning to recognize speech acts. But it is not a given that learning clause types and learning speech acts happen sequentially like that. Here I explore the hypothesis that it is likely that speech act and clause type are jointly learned: there is a semantic category namely sentence force that is systematically related to (a) syntactic distribution (i.e. the distribution of formal features associated with clause types) and (b) pragmatic functions (speech acts). Observations of

only the formal distribution or only the pragmatic function had the potential to mislead, and so the learners need to jointly learn both to mutually constrain each other. They learn to identify clause types by tracking formal regularities in conjunction with their growing knowledge of speech act and its associated social-pragmatic cues; simultaneously, they learn to identify speech acts by tracking socio-pragmatic cues in conjunction with their growing understanding of the formal features of various clause types.

To get one step closer from the pragmatic bootstrapping hypothesis to the **pragmatic syntactic bootstrapping hypothesis**, I first ask how much speech act information children need to identify clause types. If children do not need *perfect* speech act information to figure out clause types, then we do not need to assume that by the time children are learning to figure out clause types, they have already fully figured out speech acts.

In Chapter 3.3.4 and Chapter 4.3.3, I simulate the learning of clause type with various degrees of noise in the speech act information, so that we can see how much pragmatics a learner needs to succeed at the clustering problem. The results suggest that even if learners only have limited access to pragmatics, they could still benefit from this source of information. This allows us to abandon this misleading way of formulating the problem, that the learning of clause types has to happen after the learning of speech acts.

I then ask what kind of non-clause type cues for speech act information are present in the input. Even if children must rely on clause type information to figure out the speech acts, they could have access to additional information that is unrelated to clause typing, but informative for recognizing speech act type. For example, in conversations, we use questions to elicit responses and information, which leads to behaviors like pausing after questions, or looking directly at our interlocutor, to nudge them to answer our questions. This in turn means that we can expect different kinds of behavior to be correlated with each speech act. Thus, for instance, armed with an innate category for questions (e.g. Carruthers 2018³), and a theory of what questions do in conversations, the child can expect certain kinds of

³Carruthers (2018) argues that what is innate is the questioning *attitude* (“desire to know,” to be more precise), which for him is what gives us the act of questioning.

nonlinguistic behavior to be somewhat correlated with the act of asking a question.

Some candidates for cues that could potentially differentiate questions from other speech acts are prosody, pauses, and direct eye gaze. While we have seen that prosodic cues might not be helpful for clause typing, it could be that it is a useful cue for learning speech acts. Cross-linguistically, pitch rises tend to signal questions and pitch falls signal assertions; and some argue that this universality reflects the innate knowledge that high rising pitch connects to the speech act of questioning (Ohala 1984, Gussenhoven & Chen 2000, Gussenhoven 2002 among others). If children are armed with this knowledge that questions are associated with rising contours, they might expect rising contours to be correlated with the act of asking a question. If the input is such that more questions are uttered with rising contour than assertions, then this could be helpful for the child to figure out what utterances are questions in their input. But just as not all interrogatives have subject-auxiliary inversion and not all cases of subject-aux inversion are interrogatives, not all questions have final rises (Bartels 1999, Gussenhoven 2000, Hedberg, Sosa, & Fadden 2004, Ladd 2008), and not all final rises are questions (e.g. Bartels 1999, Gordon 1999, Ladd 1981, Ward & Hirschberg 1985, Goodhue et al. 2016). In English, *wh*-questions are usually produced with final fall (Hedberg, Sosa, & Fadden 2004, Ladd 2008), and the rise-fall-rise contour could be associated with assertions (Ladd 1981, Ward & Hirschberg 1985, Goodhue et al. 2016). Nevertheless, it is still possible that more questions than assertions are produced with rises in the input, which could still be informative for the child. It could also be that rises reliably track certain interrogatives – polar interrogatives – and that learners would first acquire these interrogatives, and then bootstrap other interrogatives from there, by using correlations in other morphosyntactic or pragmatic cues. We therefore need to verify empirically whether prosodic information is potentially informative. I find that parents do not use final rises more often with questions, but polar interrogatives have more final rises than other types of speech acts and clause types, including *wh*-interrogatives and declaratives.

The canonical function of questions is to solicit responses or seek information (Searle

1975a, Levinson 1983, Stivers 2010, see Krifka 2011 for a recent overview). When we use questions, it is likely that we pause after questions, or look directly at our interlocutor, as a way to signal to them that they need to take over the conversational turn. If children have innate expectation that certain speech act is used for response-elicitation, and if they know that pauses and direct eye gaze are how humans signal their expectations of a reciprocation of communication, then they may expect questions to be correlated to some degree with longer pauses and direct eye gaze. Again, this is something that needs to be investigated empirically.

If children have such expectations, would they find any socio-pragmatic cues to speech act in the input? In Chapter 5, I conduct a corpus study examining the prosody of English-speaking parents' utterances, and in Chapter 6 I examine the length of pause after utterances (Section 6.2.3.1), and proportion of eye gaze around the time of an utterance (Section 6.2.3.2).

I find that parents tend to pause longer after questions, and attend the child more when asking questions. To the extent that they are weakly correlated with the questioning act, it is in principle plausible that (a) a child could use these features, in addition to their growing knowledge of clause types to infer the speech act category of an utterance; and (b) this little bit of information about speech act could then be used to provide the noisy pragmatic information that the child needs in order to get the clause type clusters identified accurately.

1.3 Summary of the learning problem

This dissertation is about how English- and Mandarin-acquiring children figure out the make-up of the three major clause types and their associated sentential force in their language and how they link them to their canonical speech acts. Languages tend to have three major clause types (declaratives, interrogatives, imperatives), dedicated to three main speech acts (assertions, questions, commands, Sadock & Zwicky 1985 among others). However, the particular forms that these clause types take differ from language to language, and have to be

learned. Previous experimental results suggest that by 18 months old, children differentiate these clause types and associate them with their canonical speech act (Geffen & Mintz 2011, 2015, Casillas & Frank 2017, Perkins 2019, Goodhue, Hacquard, & Lidz 2022). To gain this ability, children need to identify the right categories of clauses (the **clustering problem**) and figure out what speech act they are canonically used for (the **labeling problem**).

This dissertation investigates whether the surface formal features are sufficient for learning the right clause types, and if not, how much the learners need to rely on the speech act information. I address these questions computationally by building two Bayesian clustering models simulating the learning processes of English- and Mandarin-acquiring children. I find that morpho-syntactic and prosodic features are not sufficient for acquiring clause type categories. A learner, especially a Mandarin learner, must have access to some pragmatic information in order to find the right clause types. I also show that even if the learner cannot perceive the speech act that is being performed correctly all the time, they can still benefit from taking speech act information into account (but to various degrees). I also demonstrate that length of pauses between utterances, and direct eye gaze could to some extent help children identify the speech act information.

1.4 Roadmap

This dissertation is organized as follows. Chapter 2 examines the developmental trajectory of speech acts and clause types, especially questions and interrogatives. As we will see, English-acquiring infants as early as 18 months seem to have already sensitive to the distinctions between different clause types and speech acts, and seem to understand the mapping between questions and interrogatives. The same holds for infants acquiring other languages as well, even though we have less evidence. Our question then is, how do 18-month-olds learn to figure out clause types?

Chapter 3 looks at how English-acquiring 18-month-olds could solve the problem. Specif-

ically, is information from syntax enough for children to find the right three clause type categories, or do they need pragmatic information like the speech act of the sentence to find the right clustering? I build two computational models to address this question, a syntactically informed distributional learner (SID), and a syntactically and pragmatically informed distributional learner (SPID). These two learners both need to infer the abstract clause type, but SID draws inferences from syntactic information alone while SPID uses both syntactic and pragmatic information. I use a corpus study to first provide a quantitative description of the type of input that infants get, and use the resulting annotated dataset as input for the computational models. I find that pragmatic information is indeed important for solving the clustering problem: without the speech act information, SID cannot find the right clause types. Additionally, a little pragmatics goes a long way, as even if 70% of the pragmatic information is noise, taking it into account still improves the learner’s performance.

In Chapter 4, I apply the same methodology to another language, Mandarin. Mandarin-acquiring infants figure out the clause types of their language around the same age as English-acquiring infants, but the two languages employ very different morpho-syntactic features for clause typing. How do Mandarin-acquiring infants solve the problem compared to English-acquiring infants? Do they also need pragmatic information? I compare the same two Bayesian clustering models simulating a syntactically informed distributional learner (the SID model) and a syntactically and pragmatically informed distributional learner (the SPID model), and found that learners might not be able to identify any of the clause types correctly without pragmatic information; even with pragmatic information, the learner might still have some difficulty identifying the imperative clause type. I also show that this improvement brought by pragmatics could be achieved with noisy speech act information, same as for English learners, but the threshold for the level noise allowed is 40%, which is lower than for English learners.

The role of prosody for clause typing and speech act identification is quite complicated. In Chapter 5, I turn to how this feature could be utilized for clause type and speech act

identification. I find that parents tend to use final rises more often when using polar interrogatives than when using declaratives and *wh*-interrogatives, but there is no difference among clause types and speech acts with how often parents use final rises. I also find that without separating polar and *wh*-interrogative, prosodic features do not improve the performance of the SID model, pragmatics is still needed to find the right clustering. I do not find the prosody of parents' questions to distinguish questions from the other types of speech acts, but nevertheless identify a prosodic distinction between polar interrogatives and declaratives. The other two cues,

In these simulation studies, I have assumed that infants can access information about the speech act types. How do they obtain such information about the speech acts of their parents' utterances? In Chapter 6, I explore potential cues from parents' behavior that might help English-acquiring infants identify questions. I find that both length of pauses between utterances and direct eye gaze could help a learner to distinguish assertions from questions/requests: parents pause longer after questions, and look at the child longer after questions. Even though these cues cannot perfectly predict the type of speech act that is being performed, they might still be crucially useful to learners. As my simulation with noisy speech act information suggests, learners can still benefit from very noisy, imperfect information about speech act type when learning to cluster sentences into clause types. This, I conclude, may be how children escape the chicken-and-egg problem of clause typing and speech act identification. Chapter 7 concludes the dissertation with some related discussion.

Chapter 2: Background

This dissertation investigates how children figure out the clause types of their language, on the basis of their input. In this chapter, I review what we already know from the existing literature on (1) the target knowledge, namely the formal properties of clause types and speech acts and their analyses in the formal literature; (2) children’s understanding of clause types and speech acts, and the linguistic and pragmatic capacities that they can draw from early in development, and what we know about the kinds of speech acts and clause types children are exposed to. I will come back to English and Mandarin-acquiring children’s morpho-syntactic knowledge in Chapter 3 and Chapter 4.

2.1 Theories of clause types, sentential force, and speech acts

2.1.1 Speech acts and force

Frege observes that it is possible to express a thought without judging it to be true – that considering a particular claim and judging it to be true are different things altogether. In asking a polar question like *Is it raining?* one is not judging anything to be true. But in answering *Yes* or asserting the declarative sentence like *It is raining*, one does. Nevertheless, *Is it raining?* and *It is raining* have something in common: they have the same content, or as Frege puts it, they contain the same *thought*. From this observation, he derives the influential distinction between *content* and *force*. The following passage from *Der Gedanke* makes this distinction clear:

An interrogative sentence and an assertoric one contain the same thought; but the assertoric sentence contains something else as well, namely assertion. The interrogative sentence contains something more too, namely a request. Therefore two things must be distinguished in an assertoric sentence: the content, which it has in common with the corresponding propositional question; and assertion. The former is the thought or at least contains the thought. So it is possible to express a thought without laying it down as true. The two things are so closely joined in an assertoric sentence that it is easy to overlook their separability.

Frege 1918:62, Translated by Peter Geach and R. H. Stoothoff

That is, the polar question *Is it raining?* and the assertion *It is raining* both contain the same thought, namely *whether it is raining*. But *Is it raining?* presents this thought with questioning force and *It is raining* presents it with assertoric force. By presenting a thought with assertoric force, one expresses that one judges the thought to be true. But there are other ways of presenting the same thought, for example with questioning force.¹

The idea of “force” is further developed by Austin (1975) in his work on speech acts. The core observation is that, utterances may have a variety of forces, not just asserting and questioning, but also promising, declaring, warning, marrying, naming and so on. According to Austin, the same sentence can in principle be used with a broad variety of forces:

We may be quite clear what *Shut the door* means, but not yet at all clear on the further point as to whether as uttered at a certain time it was an order, an entreaty or whatnot. What we need besides the old doctrine about meanings is a new doctrine about all the possible forces of utterances.

Austin 1975:251

¹However, Frege argues that the question of truth does not arise for sentences expressing commands, requests, wishes etc., so even though a sentence expressing request have sense, its sense is not a *thought*.

This new doctrine is to categorize utterances primarily by what one *does* with them, rather than what they *mean*. His analysis of what one does with an utterance happens on three levels: the locutionary act (the act of speaking), the illocutionary act (the act done through speaking, e.g. informing, warning, requesting), and the perlocutionary act (the by-product of producing the utterance). Let's walk through an example with these three levels. If we utter a sentence like *The campus is closed tomorrow*, the locutionary act is that the speaker utters these very words; the illocutionary act is that the speaker informs the addressee of the news about campus shutdown, and (at least one of) the perlocutionary acts here is that the speaker is changing the addressee's plans about coming to campus tomorrow. The illocutionary act is of particular interest, as it is the speaker's intent to inform by producing the utterance. To be more precise, according to Austin, the speaker and hearer are engaging in a conventional procedure of information sharing in which performing a certain locutionary act *is* to perform a certain illocutionary act. The locutionary act has its illocutionary force in virtue of being a part of this conventional procedure. The utterance has informing force by convention.

Austin is in particular interested in developing a taxonomy of speech acts. He focuses his attention on categorizing *speech act verbs*: the verbs that can be explicitly used to invoke a particular convention when appearing in the same sentence as the word *hereby*. For example, one invokes the procedure of information sharing (i.e. one makes explicit that one is performing the illocutionary act of informing) by saying *I hereby inform you*. Note that not all verbs with a communicative meaning are speech act verbs. For example, it is odd to say *I hereby surprise you*. So, in Austin's taxonomy, *surprising* is not an illocutionary force, but *informing* is.

Searle (1969) points out that there is no reason to believe that the verbs of English (or any other language, for that matter) exhaust the illocutionary forces. There could be illocutionary forces that do not correspond to a verb. However, Searle agrees with Austin that speech acts, and in particular their illocutionary forces, involve certain social conventions and

constitutive rules. Searle moreover follows Grice (1957) in stressing the role of the speaker's intention to perform a certain act and to be recognized as performing this act. Specifically,

In the performance of an illocutionary act, the speaker intends to produce a certain effect by means of getting the hearer to recognize his intention to produce that effect, and furthermore, if he is using words literally, he intends this recognition to be achieved in virtue of the fact that the rules for using the expressions he utters associate the expressions with the production of that effect.

Searle 1969:259

Searle (1976) then gives the following broad categorization of speech acts by what they are used to achieve:

- (17)
- a. assertives = speech acts that commit a speaker to the truth of the expressed proposition
 - b. directives = speech acts that are to cause the hearer to take a particular action, e.g. requests, commands and advice
 - c. commissives = speech acts that commit a speaker to some future action, e.g. promises and oaths
 - d. expressives = speech acts that express on the speaker's attitudes and emotions towards the proposition, e.g. congratulations, excuses and thanks
 - e. declarations = speech acts that change the reality in accord with the proposition of the declaration, e.g. baptisms, pronouncing someone guilty or pronouncing someone husband and wife

Like Frege, Searle also draws a strict distinction between *force* and *content*. However, there are some significant differences between Frege's and Searle's views on the distinction. For Frege, content is what the sentence radical expresses (e.g. what's in common between a declarative and a polar interrogative). But for Searle, expressing a content is an act, and

he does not see how “sentences could perform acts” (Searle 1976:257). Rather, acts are performed by speakers. Thus, speakers perform illocutionary acts, and (most of) illocutions involve the expression of a content.² The content of an illocutionary act is the proposition that the speaker expresses by uttering a particular sentence. Nevertheless, of course, which sentence the speaker utters (and how they utter it) influences what the content and the force of the illocutionary act are.

According to Searle, sentences have a proposition-indicating element and a function-indicating element that in later work he calls the *illocutionary force indicating device* (IFID) of a sentence (Searle 1976). For him, IFIDs are formal features of a sentence. For example, English IFIDs include word order, stress, intonational contour, punctuation, the mood of the verb, and certain lexical elements (notably, Austin’s speech act verbs). However, he also notes that in many circumstances the context alone is sufficient to indicate force, so IFIDs do not always need to be explicit (Searle 1976:257).

This way of associating the formal features of a sentence with illocutionary force leads Searle to the discussion of **indirect speech acts**. These are sentences that “contain the illocutionary force indicators for one kind of illocutionary act can be uttered to perform, in addition, another type of illocutionary act” (Searle 1975b:168). For example, subject-auxiliary inversion is an IFID indicating the questioning force, but a dinner guest uttering *Can you pass the salt?* is using the sentence to perform a requesting act. Searle argues that the speaker performs the questioning act by way of performing the requesting act; the questioning force is the “secondary” and “indirect” illocutionary act and the “literal” meaning of the utterance, whereas and the requesting force is the “primary” act and “nonliteral” meaning of the utterance (Searle 1975b:170).

This tension between the form of the sentence and the intention of the speaker seems to be a problem for Searle. On the one hand, Searle concedes that “the utterance has two illocutionary forces,” questioning and requesting, because “the speaker intends to produce in

²Searle explicitly mentions that there are illocutionary acts without contents, such as *Ouch!* or *Hurrah!*

the hearer the knowledge that a request has been made to him, and he intends to produce this knowledge by means of getting the hearer to recognize his intention to produce it” (p.168); on the other hand, he argues that the sentence itself, and sentences like this one, “do not have an imperative force as part of their meaning” (p. 172), because there is no IFIDs in the sentence related to the imperative force, and the same sentence could be used to not convey the imperative intention. Searle’s solution is to argue that even though there is no IFID for the imperative force, the hearer can identify an ulterior illocutionary point beyond the illocutionary point contained in the meaning of the sentence (Searle 1975b), and use recognizing the existence of this ulterior illocutionary point to infer the requesting force. An “illocutionary point” refers to “the point or purpose of a type of illocution” (Searle 1975a), which the hearer can infer with principles of conversations.

Searle & Vanderveken (1985) then pursue the project of developing a taxonomy for all possible illocutionary forces. They suggest that each possible illocutionary force can be defined as a septuple of values, each of which is a “setting” of a value within one of the seven characteristics of illocutionary acts: the illocutionary point, its degree of strength, the mode of achievement, the content conditions, the preparatory conditions, the sincerity condition, and the degree of strength of the sincerity condition. It follows, according to this suggestion, that two illocutionary forces F1 and F2 are identical just in case they are characterized by the same setting of these seven values.

While it is true that there are problems with this specific set of criteria that Searle & Vanderveken (1985) proposed (e.g. it seems that “illocutionary point” is just another name for illocutionary force), the broader problem with the Austianian-Searlean approach seems to be that there is a general pattern that speakers can use sentences to perform all kinds of acts, and it is very hard to pinpoint the exact act that is performed in a specific case, no matter how many criteria we propose. This problem already manifests in some way in our previous illustration of Austin’s theory with the utterance *The campus is closed*. When talking about the example, it seems natural to say that the illocutionary act is informing,

but couldn't we say just as well that the speaker intends to produce an effect of warning in the hearer with the sentence? Or an effect of lamenting? Or all of the above? This problem makes it extremely difficult to implement their taxonomy of speech acts. As many examples (such as *The campus is closed*) do not come with an explicit speech act verb like *inform* or *warn* (e.g. in *I'm warning you that ...*), it is almost impossible to discern the intended effect to the level of fine-grainedness given by an Austinian taxonomy of speech act verbs or a Searlean taxonomy of possible acts. As any group of annotators attempted to label corpus data using the Austinian-Searlean methods would tell you, applying these definitions to specific examples is extremely difficult and achieving any sort of inter-annotator agreement in the task of classifying speech acts according to such a taxonomy is next to impossible.³

At the same time, discarding the link between IFIDs and force because of the existence of indirect speech acts (e.g. as suggested by the radical pragmatics approach, Atlas & Levinson 1981, Levinson 1983) seems to miss this important association between form and function. As we have seen in Chapter 1, languages tend to have declaratives, interrogatives, and imperatives dedicated to assertions, questions, and requests/commands. Ignoring this universality seems to miss an important generalization in language. While it is true that there is a mismatch between the conventional effects linked to polar interrogatives and the speech act performed by the speaker in the indirect speech act *Can you pass the salt?* example, the very fact that we intuitively interpret the sentence as a question suggests that we probably do not want to take form out of the equation.

At this point, it appears fruitful to attend not to the fine distinctions of Austinian speech act verbs, nor to the distinctions of Searle and Vanderveken's metaphysics of action, but rather to how utterances affect the context they are made in (Hamblin 1971, Stalnaker 1978b, Lewis 1979, Gazdar 1981). As Chierchia & McConnell-Ginet (1990) points out, while a sentence like (18) can be uttered to perform many different Austinian speech acts, such as

³An unfortunate consequence of the multitude of illocutionary forces that one may attribute to the same utterance is that, in order to boost inter-annotator agreement, annotation schemes sometimes blend formal categories with functional categories. For example, in the schemas discussed in Ninio et al. (1994) and Stolcke et al. (2000), *wh*-interrogative and polar interrogatives are included in the labeling.

claiming, guessing, reminding, warning, or threatening, there is a similarity across all these cases. Namely, that the declarative sentence places the proposition expressed by this sentence “in the common ground and discards any possibilities rendered no longer live because of their inconsistency with that (possibly new) information” (Chierchia & McConnell-Ginet 1990:171).

(18) The bull is in the field.

We can therefore distinguish the illocutionary force of the whole utterance (i.e. the **utterance force**, Murray & Starr 2018, crediting Chierchia & McConnell-Ginet 2000; cf. Portner 2018) from the force of the sentence. Chierchia & McConnell-Ginet (1990) refers to the latter, namely the link between grammar and meaning, as **sentential force**. As they put it, the sentential force is “what the grammar assigns to the sentence to indicate how that content is conventionally presented” (p. 164).⁴ For example, the conventional effect of an assertion on the discourse could be described as proposing to add a proposition to the common ground (CG, Stalnaker 1978a, 2002). Similarly, the questioning speech act can be described as adding a question to a stack of questions, referred to as the questions-under-discussion stack (QUD, Roberts 1996, Ginzburg 1995), and requests/commands could be adding an item to the To-do-list (Portner 2004). Table 2.1 summarizes the clause types and their canonical functions, and their conventionalized effects on the discourse.

⁴Some use the term *sentence mood* (e.g. Portner 2018). But looking at the definition given by Portner (2018), it seems that these two refer to the same thing:

Sentence mood is an aspect of linguistic form conventionally linked to the fundamental conversational functions within semantic/pragmatic theory.

Portner 2018:122

Portner 2018 also uses the term *sentential force*, but to refer to the conventional effects that an utterance have on the discourse, which roughly corresponds to what Chierchia & McConnell-Ginet (2000) means by utterance force.

	Canonical Function	Conventionalized effect on the discourse
Declarative	Assertion	Proposing to add proposition to the common ground
Interrogative	Question	Add current question to the Question Under Discussion stack
Imperative	Request	Add (or propose to add) certain property to the To-Do List of the addressee

Table 2.1: Clause types and their conventionalized force; adapted from Portner (2004:238)

Regardless of theoretical assumptions, what can be agreed on is that languages tend to have three major clause types (declaratives, interrogatives, and imperatives), and they are linked to three major speech acts (assertions, questions, and requests/commands). In this dissertation, I will follow Chierchia & McConnell-Ginet (1990) to refer to this link between grammar and meaning as the **sentential force**. I will also loosely use the term **speech act** to refer to the conventional effects that the whole utterance has on the discourse, which could be different from the sentential force of the sentence. For example, I consider the indirect speech act example *can you pass the salt* an interrogative clause, that has interrogative sentential force, but the whole utterance expresses a requesting speech act.

2.1.2 Clause types

As we have discussed, languages tend to have dedicated three clause types for kinds of speech acts (Sadock & Zwicky 1985, König & Siemund 2007, Aikhenvald 2016, Portner 2018, see König 2020 for a recent review).

Traditionally, clause types are viewed as form-function pairs: “[w]hen there’s a regular association of form and the speaker’s use of sentences, we will speak of the form-use pair as a sentence type” (Sadock & Zwicky 1985:156). Conceiving clause types as a form-use mapping runs into problems with embedded clauses like the underlined clause in (19), as we cannot tease apart the “use” of this portion of the sentence alone:

- (19) Mary knows who Ann can hug.

As formal categories, clause type categories participate in syntactic relations like syntactic selection. Certain verbs select one clause type and not the other. For example, *wonder* selects interrogatives but not declaratives, and *think* selects declaratives but not interrogatives:

- (20) a. Mary wonders who Ann can hug.
 b. *Mary wonders Ann can hug Sue.

- (21) a. Mary knows who Ann can hug.
 b. Mary knows Ann can hug Sue.

As shown by the contrast between (20) and (21), *wonder* is grammatical with embedded interrogatives, but not declaratives, while *know* is grammatical with both.

This formal category is often observed to have two typological universals. First, languages tend to have three major clause types (declaratives, interrogatives, and imperatives) associated with three major speech acts (assertions, questions, and requests/commands), as we have seen in (1), even though the expression for each type

Second, among these three clause types, declaratives are usually considered the default clause type, whereas interrogatives and imperatives are the results of some operations on declaratives (Sadock & Zwicky 1985, Chomsky 1957, 1995, Akmajian 1984, Platzack & Rosengren 1997, Rizzi 1997 among many others).

To capture these two language universals, the clause type information is often analyzed as related to an abstract feature occupying C^0 (Chomsky 1995, Cheng 1991, Rizzi 1997, 2001, Chomsky & Lasnik 1977, Platzack & Rosengren 1997, Akmajian 1984, C.-H. Han 1998). An interrogative clause has the [+int] value, imperative [imp], and declarative [-int].⁵

In this dissertation I use the term *clause type* to specifically refer to the way of grouping sentences grammatically that are associated with the *sentential force* in matrix context.

⁵In many analyses, the value for imperative C^0 is [imp] instead of [+imp], partly due to there is no [+int, +imp] clause type (Platzack & Rosengren 1997, C.-H. Han 1998 among others). In this dissertation, I do not make any commitment on the specific analysis of imperatives. Since we are always comparing imperatives with declaratives, I will sometimes use [+imp] as the feature for imperatives and [-int, -imp] for declaratives. But this is not to make a theoretical claim about the hypothesis space of C^0 , as it is unclear which clause type corresponds to [+imp, +int].

Clause types are distinguished strictly based on formal criteria, and clause type categories are formal categories. I additionally assume that at the matrix level, the semantics of [\pm int, \pm imp] is the sentential force of the sentence. As this dissertation only deals with the learning of clause types in matrix contexts, the difference between clause type and sentential force is negligible for the learner: they are both abstract categories that correlate with some surface features on the one hand and speech act on the other. I will therefore use clause type and sentential force interchangeably when discussing the learning problems.

2.1.3 The role of prosody

Searle counted prosodic features among the IFIDs, the illocutionary force indicating devices, but the relationship between prosodic features, clause type, and speech act is by no means straightforward. Cross-linguistically, rising contour is frequently associated with the questioning act (Bolinger 1978, Ladd 1981, Gussenhoven & Chen 2000, Ladd 2001, a.o.). Some have argued that this universality reflects the innate knowledge that high rising pitch connects to the speech act of questioning (dubbed the “Strong Universalist Hypothesis”, Ladd 1981). When participants are given prosodic contours in a language they do not speak, they tend to recognize the ones with higher end rises as being used to perform a question act (Gussenhoven & Chen 2000); when hearing low-pass-filtered sound files with no recognizable lexical information, participants tend to infer the ones with final rises as indicators for turn transitioning, the function typically associated with questions in a conversation (Bögels & Torreira 2015).

Speech act could directly influence the use of prosodic features without clause type information. For example, in a scenario where parents are looking for their missing children, they might call their children’s name with a rising intonation (e.g. *John?*). This fragment is understood to be used to perform a questioning act.⁶

⁶Although one could argue that these are elided interrogative clauses; but there are other cases where fragments with rising intonation could be understood as elided declaratives:

- i. A: Who has a sister?

In discussions of intonational typology, it seems that intonation should be considered as a surface feature for clause typing too. In some languages such as Italian and Portuguese, prosody is the only feature that distinguishes polar interrogatives and declaratives (König & Siemund 2007, Truckenbrodt 2009, Frota 2002 among many others), and some languages use a combination of morpho-syntactic and prosodic cues (e.g. French, Sanuma, König & Siemund 2007, Aikhenvald 2016). There are also languages that might not use intonation for clause typing. For example, Vata (a Kru language) uses unspecified open vowels, and no prosodic features, to mark interrogatives (Rialland 2007).⁷ Table 2.1.3 lists some of the languages that use morpho-syntactic and/or prosodic features for clause typing.

	Languages
Morpho-syntax and Prosody	English (Ladd 2008, but see detailed discussion in Chapter 5), French, Catalan (Prieto & Rigau 2007)
Prosody alone	Italian (König & Siemund 2007), Portuguese (Truckenbrodt 2009, Frota 2002)
Morpho-syntax alone	Shekgalagari (Hyman & Monaka 2011), Vata (Rialland 2007), Cantonese (Wong, Chan, & Beckman 2005; cf. Ma, Ciocca, & Whitehill 2011)

Table 2.2: Languages that use morpho-syntactic and/or prosodic features to differentiate interrogatives and declaratives

For languages that do use prosody for clause typing, the specific features associated with the difference between interrogatives and declaratives differ from language to language. One frequent pattern seems to be to use some form of rising intonation for interrogatives (especially polar interrogatives), and falling intonation with declaratives, but exceptions

B: John?

B’s answer could be an elided declarative: “John has a sister.” See Chapter 5 for more discussion on the relation between speech acts and clause types.

⁷One caveat is that in many of the field works that the various paper cited, the authors do not make a distinction between *question* the speech act and *interrogative* the clause type.

do exist (notably, falling for interrogatives in Roermond Dutch (Gussenhoven 2002), and rising for declaratives in Chickasaw (Gordon 1999)). But there are other strategies too. For example, Akan (a Kwa language) uses breathy termination to indicate interrogatives (Genzel & K ugler 2020):

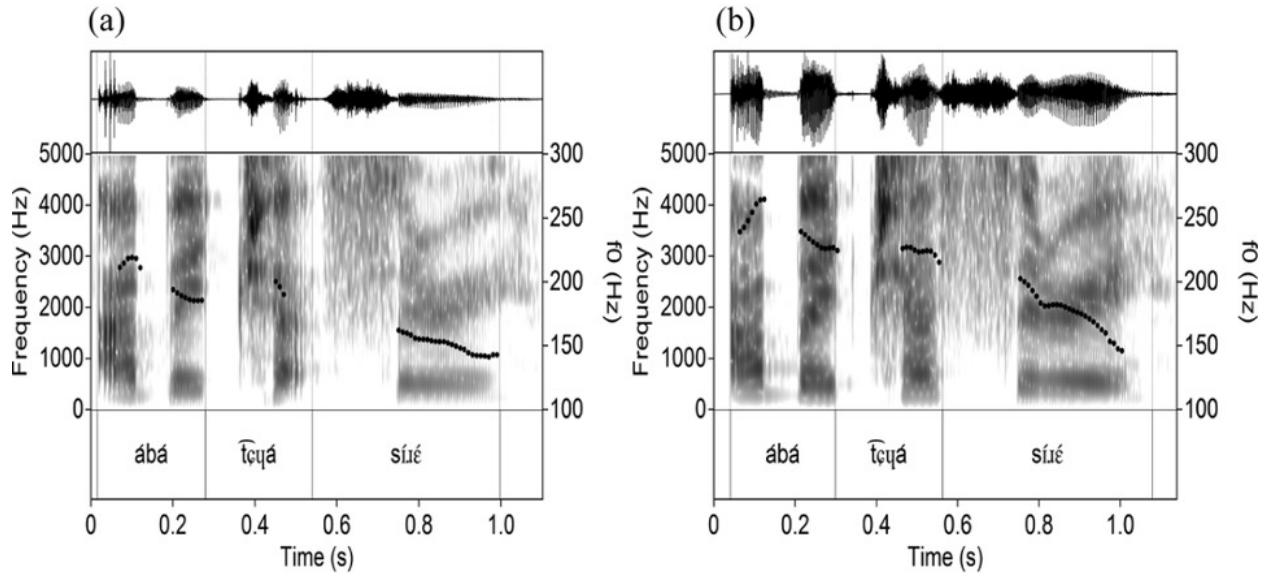


Figure 2.1: Waveforms and pitch curves of a declarative (right) and an interrogative by a female Akan speaker (Figure 7 in Genzel & K ugler (2020))

As shown in Figure 2.1, the interrogative sentence is produced with a breathy voice quality at the end, as compared to the declarative sentence (F0 lowering here is a by-product of the laryngeal setting of producing the breathy voice). Table 2.1.3 lists some of the prosodic features that languages use to mark the differences between polar interrogatives and declaratives:

Prosodic feature	Polar interrogative feature (language)
Boundary tones:	H% or LH% (Dutch, French, Japanese) HL% (Roermond Dutch) L% (Catalan, Chickasaw, Bininj Gun-wok)
Peak alignment	late peak alignment (Neapolitan Italian, Russian)
Peak height	Higher (Russian, Japanese)
Register expansion	Less downdrift (Wollof, Danish)
Final lengthening	Lengthening (Nateni)
Final voice quality	Breathy termination (Akan, Moba)

Table 2.3: Intonational markings used to differentiate polar interrogatives and declaratives (adapted from Butler, Frota, & Vigário 2012)

Considering the fact that some languages do not use prosodic features for clause typing,⁸ it is possible that children do not have a priori knowledge that prosodic features are informative for clause type categorization. But what about speech acts? We do not have enough data from Vata to determine whether the question act could still carry a different set of prosodic features despite the fact that interrogatives are not marked by prosody, but given Gussenhoven and Chen’s (2000) experiments with non-native prosodic contours, it is likely that speakers of Vata might be able to use rising intonation to identify questions.

If children do not assume that prosody informs clause typing, Italian- and Portuguese-acquiring children might need to learn the connection between prosody and clause typing, in addition to learning clause type and speech act categories. But it is also possible that children do not need to learn this connection at all. If children assume that prosody informs speech acts, and that speech acts are informative of clause typing, prosodic information could indirectly help the learning of clause type categories.

It is also possible that children come to the learning task assuming that prosody *does* inform clause typing. For children learning languages like Vata, where prosody is irrelevant for clause typing in their language, the input might be such that this connection is simply

⁸However, we do not have data on whether Vata or Shekgalagari uses prosody to distinguish imperatives from declaratives. Given the fact that languages tend not to have prosodic distinctions between declaratives and imperatives, our generalization has a good chance to hold.

uninformative. Thus, assuming a connection between clause typing and prosody would not make any difference for the learning the clause type categories. To test this possibility, we need to look at the input to Vata-acquiring children, and see if assuming the link between prosody and clause type would hinder children from learning the right clause type clustering.

Another complication related to prosody is that languages use different prosodic features for clause typing, and many of them are not correlated with a rising F0 at the end. For example, Akan’s breathy termination in fact results in the lowering of F0. Therefore, instead of having a built-in knowledge of the form “final rise \sim questionhood,” children might need to learn which prosodic features are associated with which speech act/clause type clusters, the same way that they need to learn which morpho-syntactic features go with which clause type cluster.

Both English and Mandarin have dedicated morpho-syntactic features for interrogatives (e.g. subject-auxiliary inversion in English). But in both languages, it seems that there are also prosodic differences associated with interrogatives. We will discuss the prosodic features of these two languages in more details in Chapter 4 and Chapter 5.

2.2 Acquisition of speech acts and clause types

2.2.1 Early communicative abilities

As mentioned in the previous chapter, it would be very useful for learners if they could rely on speech act information to learn how clause types are expressed in their language. Speech acts are, unfortunately, not directly observable, but there may be cues accessible to learners that are broadly indicative of speech act. One good candidate are cues for communicative intent. Previous studies have shown that infants are sensitive to communicative intentions even in their first months of life, and they can distinguish signals for communication, such as eye contact, intentional pointing, and speech, from non-communicative signals. Infants tend to follow the gaze of their interactive partner. Newborn babies already show rudi-

mentary form of gaze following (Farroni et al. 2004); as young as 3 months of age, infants seem to automatically orient to gaze cues, as they could shift gaze fast with adult’s gaze direction (Hood, Willen, & Driver 1998); 6-month-olds follow gazes that are communicative, such as adults’ directly gazing at an object (Gredebäck et al. 2008), or being greeted (e.g. “Hello!”) in infant-directed speech (Senju & Csibra 2008). Human infants, but not chimpanzees, understand pointing as intentional (Pika & Liebal 2006, Povinelli et al. 1997, Morissette, Ricard, & Décarie 1995); 9-month-olds remember different aspect about an object depending on whether the object is presented in a communicative (such as pointing) vs. non-communicative scenario (Yoon, Johnson, & Csibra 2008); 12-month-olds not only follow the direction of pointing, but also infer that the pointing is to indicate information such as where a toy is hidden in hide-and-seek games (Behne, Carpenter, & Tomasello 2005, Behne et al. 2012). Starting from six months old, infants can readily interpret speech, but not other vocalizations like coughing, as indicating communicative intentions (Vouloumanos, Martin, & Onishi 2014).

Of course, merely recognizing an act *as communicative* is not the same as distinguishing speech acts – beyond the recognition that there is communication, one needs to infer the function of the communication to approach speech act recognition. Indeed, besides recognizing a signal as communicative, infants can attribute simple goals and desires to an agent, and use shared experience to infer their goals. By 9 months old, infants can recognize an agent’s goal of reaching for an object (Woodward 1998, Baldwin et al. 2001);⁹ 14-month-olds can distinguish actions that are intentional and ones that are accidental, and only imitate actions that are intentional (Carpenter, Akhtar, & Tomasello 1998, Sakkalou et al. 2013), and if things go wrong, they offer to help achieve others’ goals (Warneken & Tomasello 2007); they could use shared experience such as cleaning up toys to interpret an agent’s pointing as putting away the pointed object away, and would not attribute this goal to an agent who don’t have the shared experience (Liebal et al. 2009).

⁹But studies show that still have difficulty with the avoidance type of goals until 14 months old (Feiman, Carey, & Cushman 2015).

Relatedly, it seems that infants have some understanding of other people’s desires and beliefs, attitudes that are frequently associated with the speech acts of assertions, questions, and requests. For example, 9-months-old can distinguish whether someone is unwilling or unable to hand them a toy (Behne et al. 2005); 12-month-olds can use eye gaze and positive affect to infer which object an agent is going to reach for (Phillips, Wellman, & Spelke 2002), and by 18 months old, infants use positive affect alone to infer which object the agent prefers, even if the preference is different from their own (Repacholi & Gopnik 1997). Around 12 months old, infants can use pointing not to merely direct others’ attention, but to request others to share information (Kovács et al. 2014); 13-month-olds seek information to clarify ambiguous situations (Vaish, Demir, & Baldwin 2011); by 18 months old, infants are shown to be able to attribute beliefs to other people (Onishi & Baillargeon 2005, Surian, Caldi, & Sperber 2007, Song et al. 2008, Song & Baillargeon 2008, Scott & Baillargeon 2009, Perner & Roessler 2012, see Scott & Baillargeon 2017 for a recent review). 18-month-olds could use common ground knowledge shared with an agent to infer whether the agent is requesting an action with an object (e.g. open the door with the key) or simply playing with the object (Schulze & Tomasello 2015).

Infants also have some understanding of the norms of conversations. For example, from early on, they understand that conversations take turns. Longitudinal studies with 3 to 5-month-olds show that the overlap between infants’ vocalization and parents’ speech gradually decrease (Hilbrink, Gattis, & Levinson 2013), and the rhythm of their vocalization starts to mimic the structure of a conversation, as they wait for parents to finish their turn (Hilbrink et al. 2013, Hilbrink, Gattis, & Levinson 2015, Casillas, Bobb, & Clark 2016); they also use turn transition points (e.g. questions) to cast predictive eye gaze to the next speaker (Casillas & Frank 2017).

In sum, this body of work suggests that by 18 months old, the communicative abilities needed to acquire the distinctions between speech acts and map them to clause types are in place. In the next section, we will see that indeed by this age, infants seem to figured out

the major speech acts and are able to link them to their canonical clause types.

2.2.2 Early knowledge of clause types and speech acts

Previous studies show that children seem to have knowledge about clause types and the associated speech acts (in particular interrogatives and questions) from as early as 18 months old.

Using corpus data, many studies show that infants can use one- and two-word utterances to express a variety of intentions that can be interpreted as assertions, questions, and requests/commands (Bateson 1975, Bates 1976, Ninio et al. 1994 among others); English-speaking children start producing interrogatives around 20 months starting with *wh*-interrogatives (Tyack & Ingram 1977, Stromswold 1995, Rowland et al. 2003). From the comprehension side, results from corpus studies on parent-child interactions show that children begin to respond to parents' questions, especially *who*, *what*, *where* questions, appropriately around one and a half years old (Ervin-Tripp 1978, Steffensen 1978, Shatz 1978a,b, Berninger & Garvey 1981, Shatz & McCloskey 1984, Clark & Lindsey 2015, Moradlou et al. 2020 among others).

Results from experimental studies show that around 12 months, English-speaking infants show sensitivity to differences in word order (Geffen & Mintz 2015) associated with declaratives and polar interrogatives. Data from preferential looking tasks show that already at 15 months old, infants look at the objects corresponding to the answers of *wh*-interrogatives (Seidl, Hollich, & Jusczyk 2003, Gagliardi, Mease, & Lidz 2016, Perkins & Lidz 2020), though their success might not necessarily reflect knowledge of the syntax of *wh*-interrogatives, as they might be relying on their knowledge of verb argument structure in the task (Perkins 2019).

By 18 months old, infants seem to be able to distinguish questions from assertions, and use this distinction to infer the epistemic state of the speaker. Luchkina, Sobel, & Morgan (2018) test how well 18-month-olds' learn labels of novel objects from speakers who made

statements about familiar objects (e.g., *This is a star*) and from speakers who asked questions (e.g., *Is this a star?*). During the test phase, both types of speakers use statements to label a novel objects (*Look, a lif! Lif!*). After repeating the statements several times, children are asked by an unfamiliar novel voice to look at one of the objects (e.g. *Look, a lif! Where's the lif?*). They find that 18-month-olds only learn labels of objects from the speaker who made statements during the familiarization phase, but not from the question-askers. This result suggests that 18-month-olds can differentiate questions from statements, and furthermore, they can use this distinction to infer that question-askers might be a less reliable information source than statement-makers.

Goodhue, Hacquard, & Lidz (2022) demonstrate that 18-month-olds can associate interrogative syntax with the speech act of questioning with the preferential looking paradigm. In the experiment, infants watch a video of two puppets cheering a mechanical arm for delivering cookies into a box. During the test phase, one of the puppets leaves the scene while the cookie is being delivered, and is thus ignorant of whether there is a cookie in the box. Then participants either hear a polar interrogative *is there a cookie in the box?* or a declarative *there's a cookie in the box!* that could be uttered by either puppet. They find that at 18 months old, infants look more at the ignorant puppet when hearing the polar interrogative, suggesting that they understand the polar interrogative as a question uttered by the ignorant puppet.

Casillas & Frank (2017) tap into children's knowledge of turn-taking to see if children differentiate questions from other types of speech acts. They measure how often children switch their gaze to track the upcoming speaker. Since questions are turn-transitioning points where a switch in speaker is mostly likely to happen, if children can identify questions, they should switch gazes to the upcoming speaker more often after questions. In the experiment, children are shown videos of a conversation between two puppets in one of four conditions: with normal speech, with words only (filtering out prosodic cues), with prosody, or with no speech. They find that even at 1 years old, children are faster at switching their gaze to the

addressee after questions than non-questions in general. Breaking down the different cues to questions, they find that in the word-only condition, even 1-year-olds switch their gaze to the upcoming speaker more often after questions, but in the prosody-only condition, 5-year-olds still are at chance distinguishing questions from non-questions. These results suggest that 1-year-olds might be able to use the morpho-syntactic properties of interrogatives to infer questions. Similar results are replicated with 2.5-year-olds by Lammertink et al. (2015).

As for imperatives and commands, Shipley, Smith, & Gleitman (1969) show that even in the “telegraphic speech” phase (15-30 months old), children respond appropriately to imperatives as commands. Orfitelli & Hyams (2012) show that 3-year-olds (and some 2.6-year-olds) can use the presence/absence of verb morphology to tell whether the sentence is a pro-drop declarative or an imperative, but it’s possible that children in the experiments are using *please* as a cue for command interpretation.

Previous studies on questions in speech to children suggest that parents use questions in many different ways (Holzman 1972, Shatz 1979, Tamir 1980, Yu, Bonawitz, & Shafto 2019). In particular, compared to questions in adult-directed speech (Stivers 2010), parents tend to use question as a pedagogical strategy. Zaitsev et al. 2020 investigates the frequencies of the three basic clause types and the corresponding speech acts in speech to children between the age 1 to 3 from the Providence corpus (Demuth, Culbertson, & Alter 2006) of CHILDES (MacWhinney 2000). They find that although the three clause types generally are associated with their canonical speech acts, there are cases of mismatches (indirect speech acts). However these mismatches are often marked, and tend to form a systematic subcategory. For examples, when interrogatives are used as requests, we often see the presence of modals and attitude verbs in the sentence; when declaratives are used to perform questions, we often see rising prosody¹⁰. This is relevant because these systematic marking might alert the learners that these uses of the clause types may be somehow special. In particular, questions are often asked via rising declaratives, and requests made with interrogatives.

¹⁰Although they did not explicitly annotate the prosodic information; rising prosody is inferred from the use of question marks in the transcript.

Going beyond English and turning to Mandarin, we have less evidence for children’s early knowledge of speech acts and clause types. Most study with younger children are documentations of their production data. Particularly, Mandarin-speaking children are observed to produce *ma*-interrogatives, A-not-A interrogatives, and *wh*-interrogatives before they turn 2 years old (Miao 1986, Miao & Zhu 1992, T. H.-T. Lee 1989, Li & Tang 1991, Li & Chen 1997b,a, Fan 2012, Li, Jing, & Wong 2017).

On the comprehension side, experimental studies testing children’s comprehension suggest that 18-30 month-olds can correctly link *wh*-interrogatives to the question speech act, but are less likely to respond to A-not-A polar interrogatives (Moradlou et al. (2020)). However, this result might be due to the way the experiment is set up. During the experiment, children are asked polar interrogatives (e.g. *Is this a duck?*) or *wh*-interrogatives (e.g. *What does a duck say?*) while looking at pictures of animals. But in this setting, particularly if the questioner is a person with authority, children might consider the polar question as redundant, and therefore tend to not respond to these questions.

By 3 years old, Mandarin-children have adult-like interpretation of *wh*-interrogatives (Fahn 2003), and that they can use prosodic information to determine whether the sentence is a *wh*-interrogative or a declarative (Yang et al. 2022). In Mandarin sentences like (22) is ambiguous between a declarative and a *wh*-interrogative.

- (22) Xiaoyang mei fang shenme shuiguo zai xiangzili
 Lamb NEG put what fruit in box
 a. “What fruit didn’t Lamb put in the box?”
 b. “Lamb didn’t put any fruits in the box.”

In this case, the two interpretations are disambiguated by prosodic cues: assigning prosodic prominence on *shenme* gives rise to a *wh*-interrogative interpretation, and without prosodic prominence, the sentence is a declarative. Yang et al. (2022) show that 3-year-olds, like adults, can access both interpretations of the sentence.

2.3 Summary

In this chapter, I reviewed what we already know from the existing literature on the formal properties of clause types and speech acts and their analyses in the formal literature. Regardless of the theoretical approach, people agree on the three major clause types and their corresponding three major speech acts.

Additionally, I reviewed literature on children's understanding of clause types and speech acts, and the linguistic and pragmatic capacities that they can draw from early in development, and what we know about the kinds of speech acts and clause types children are exposed to. The results from these studies suggest that by 18 months old, infants seem to have figured out the major clause types and speech acts, and the link in-between. These studies give us a developmental window for modeling the learning of clause types, namely that we need to take into account what infants know around 18 months old, to simulate how they figured out clause types and speech acts at this age. In the next chapter, I delve into the problem of how 18-month-olds figure out clause types.

Chapter 3: Learning to identify clause types in English

As discussed in the previous chapters, cross-linguistically we see three major clause types, declaratives, interrogatives, and imperatives, corresponding to three main speech acts, assertions, questions, and commands/requests. We also saw in Chapter 2 that by 18 months old, infants seem to have figured out clause types and their canonical functions, i.e., the type of speech acts they are canonically used for. As discussed in Chapter 1, the acquisition of clause types can be decomposed into two challenges: 1) the **clustering problem**, i.e. clustering clauses in the right three categories, and 2) the **labeling problem**, i.e. linking each category to its function.

This chapter investigates how learners solve these problems, especially the clustering problem, by probing the extent to which children need to rely on pragmatic information (i.e., inferring what speech act a given utterance of sentence is conveying). As discussed in Chapter 1, pragmatic information is essential for solving the labeling problem. The question I ask here, then, is whether pragmatic information is also crucial for solving the clustering problem, or, conversely, if it might hurt. Will a learner strictly tracking morpho-syntactic regularities home in on the right three clusters of clauses? Will a learner privy to some speech act information fare better? Does the speech act information help or hurt the learner? It can help by highlighting which formal distinctions matter for clause typing; it could hurt if clause types are systematically used in indirect speech acts. If pragmatic information like speech act helps, how much pragmatic information is required?

To answer these questions, I compare the performance of two computational models, a

Syntactically Informed Distributional Model (SID) and a Syntactically and Pragmatically Informed Distributional Model (SPID). Since the success of the learning models should be commensurate with infants' early successes at identifying clause types, I test the models using data from an annotated dataset created with parental sentences in the Providence Corpus (Demuth, Culbertson, & Alter 2006). I will show that the SID fail to identify all three clause types, suggesting that the morpho-syntactic information alone is not sufficient for learning clause types, and pragmatic information, i.e. the speech act that a sentence is used to perform, is crucial.

But how much pragmatics is needed, I manipulate the ratio of noise in the pragmatic information. I will show that learners need very little pragmatic information to help them figure out clause type. Even with 70% noise, models with speech act information still outperform the one without.

In this and the next Chapter, I assume that learners do not have access to prosodic information, and we will add this information in Chapter 5. I also assume that learners has access to pragmatic information. Chapter 6 will address the question of how learners might be able to glean speech act information other than relying on clause types.

This rest of the chapter is organized as follows: In Section 3.1, I first briefly introduce the two computational models mentioned above, and how they can help us answer our questions (Section 3.1.1). I then review the formal features of English declaratives, interrogatives, and imperatives (Section 3.1.2), setting the stage for discussing the learning of these clause types. While these are features that English employs for clause typing, infants at 18 months old might not be able to perceive all of these features. To allow our models to best capture 18-month-olds' learning process, we need to look at 18-month-olds' linguistic capacity and knowledge with regard to these features. So in Section 3.1.3 I briefly go through the linguistic capacities and knowledge of infants before 18 months old with regard to the formal features used for clause typing, which in turn will guide the way these features are coded in the corpus.

In Section 3.2, I report results from a corpus study on parents’ use of different clause types and speech acts, to give a quantitative description of the information present in the input. The annotated dataset that results from this corpus study will be used in our modeling experiments.

Section 3.3 details the two computational models for the two learners, SID and SPID, to see if pragmatic information is needed for solving the clustering problem. I then manipulate the ratio of noise in the pragmatic information to see how much pragmatics is needed (Section 3.3.4). One feature not addressed in chapter is prosody. As we have discussed in Chapter 2, the role of prosody for clause typing and speech act identification might be complicated. I will come back to prosodic features in Chapter 5, report results from a corpus study examining the prosody of parents’ sentences, and results from a modeling study investigating whether

3.1 Background

3.1.1 Two learners

As mentioned above, the question we set out to answer is whether morpho-syntactic features alone are sufficient for learners to solve the clustering problem. To this end, I build two computational models simulating a syntactically informed distributional learner (SID) and a syntactically and pragmatically informed distributional learner (SPID). Both learners are forms of distributional learning: they track distributions of certain features in their input, and use these observations to infer the underlying categories that gives rise these distributions (cf. Feldman et al. 2013, Gagliardi, Feldman, & Lidz 2017, Perkins, Feldman, & Lidz 2022, Perkins 2019, Nguyen & Wilson 2021; see Pearl 2020 for a recent review). Specifically, these two learners share the same goal of discovering the underlying clustering of sentences in parents’ speech, i.e. to infer the abstract clause type categories in English. The difference between them lie in the sources of information they take as input.

The **syntactically informed distributional learner** (SID) assumes that learners track the statistical distributions of *morpho-syntactic features* present in the surface forms of sentences to infer the abstract clause type category that gives rise to these distributions. This learner serves as our baseline; by looking at its performance, we will be able to see whether syntactic information by itself is sufficient for identifying clause types.

The **syntactically and pragmatically informed distributional learner** (SPID) also tracks the distributions of morpho-syntactic features, but at the same time, it keeps track of the speech acts expressed by these sentences as well. Thus, it infers clause type categories with information from both syntax and pragmatics.

Thus, we would be able to answer the question of whether pragmatics is crucial for the clustering problem by comparing the performance of these two learners: if pragmatic information helps the learner, not only with labeling but also with clustering, we would see SPID outperforms SID. Additionally, if SPID performs better, we can manipulate the ratio of noise in the pragmatic information that SPID receives, to see how much pragmatic is needed for learners to solve the clustering problem.

The success of these two learners depend on the specific features we feed into the models. In Chapter 2, we reviewed different taxonomies for speech act information and we may abstract away from language-specific ways of indicating force by having the SPID keep track of pragmatic information in terms of knowing whether a sentence is an assertion, question or request. But the morpho-syntactic features for clause typing differ from language to language and to test whether the SID can acquire clause typing from morphosyntax alone, we must test it on the actual features of the language we test it on. So, in the next section, we will go through the specific morpho-syntactic features that English employs for clause typing, to set the stage for our discussion of the learning of clause types.

3.1.2 Clause types in English

Generally, clauses in English are marked by the presence of verbs (23a). Sentences without verbs are often classified as fragments (Sadock & Zwicky 1985), as in (23b):

- (23) a. Mary hugged Ann.
b. Mary!

In this section, we will go over the properties of the three major types of clauses, declaratives, interrogatives, and imperatives, in English. It is generally assumed that the declarative is the default clause type in English, and thus has the formal features [-int, -imp]. The hallmark of the [+int] value of English *C* is subject-auxiliary inversion, which can be seen in polar interrogatives like (24a) and *wh*-interrogatives like (24b).

- (24) a. Can Mary hug Ann? polar interrogative
b. Who can Mary hug? *wh*-interrogative

However, this association of word order and interrogativity has many exceptions. For example, in subject *wh*-interrogative sentences like (25a) and sentences with embedded interrogatives (25b-c), the interrogative takes the same word order as a declarative.

- (25) a. Who can hug Ann? subject *wh*-interrogative
b. Mary wonders who Ann can hug. embedded *wh*-interrogative
c. Mary wonders whether Ann can hug Sue. embedded polar interrogative

As discussed in Chapter 1, these cases might be a problem for learners, as they obscure the mapping between the subject-auxiliary inversion rule and [+int]. While the presence of *wh*-phrases in (25a-b), and the complementizer *whether* (25c) in these sentences are also cues for [+int], these features suffer from a similar problem. For example, free relative sentences like (26) also appears with a clause-initial *wh*-phrase, but the *C* head this *wh*-clause has the feature [-int] (Bresnan & Grimshaw 1978, Caponigro 2003):

(26) Mary ate what Ann cooked.

Free relative sentences

These won't be an immediate problem, since our primary concern is with children's acquisition of sentential force, and thus I restrict myself to cases of matrix clauses, not embedded ones.

At the same time, some declaratives exhibit subject-auxiliary inversion as well, such as Negative Inversion sentences like (27): these sentence are generally considered declaratives, but the auxiliary *would* precedes the subjects in both.

- (27) a. Never in her life would Mary eat tripe.
b. Under no condition would Mary eat tripe.

These are cases where the speech act information might be particularly useful to a learner, as identifying the utterance as making an assertion would help the learner to avoid making the generalization that [-int] also triggers subject-auxiliary inversion.

Another tricky case is what is the clause type category for rising declaratives. When declaratives are associated with a L* H-H% rising intonation, they tend to be interpreted as questions (i.e. rising declaratives, Ladd 1981, Gunlogson 2003, 2008, Jeong 2018, Rudin 2018, Goodhue 2021). But are sentences like (28) declaratives (I will use ↗ to indicate the sentence has a final rising intonation, and ↘ for final fall)?

(28) It's raining ↗

(29) It's raining ↘

In Gunlogson's 2008 analysis, sentences like (28) are declarative clauses with a rising intonation, implying that intonation does not affect the clause type feature of the sentence. Meaning-wise, clause type and intonation contribute compositionally to the discourse function of the utterance.

Many analyses of rising declaratives in the dynamic discourse tradition adopt the same assumption. As these cases are rarely discussed in syntactic literature, it is unclear whether

syntacticians would classify them as having [-int] or [+int]. But we can go through some typical syntactic tests for [+int], to see if it's possible to assign [+int] to rising declaratives. One test is whether rogative verbs that only select interrogative clauses would embed (28). If (28) bears [+int], then the sentence with (28) embedded under *wonder* should be grammatical.

(30) *Mary wonders it's raining↗.

As shown in (30), this prediction is not borne out; (28) cannot be embedded under rogative verbs like *wonder*. This suggests that (28) does not have the feature [+int]. But it could be that (28) has a silent [+int], but this silent [+int] is realized as *whether* in embedded context. So instead of (30), embedded (28) should be:

(31) Mary wonders whether it's raining.

As the rising intonation associated with (28) would not manifest in embedded clauses, it's hard to distinguish the two possibilities syntactically. But Farkas & Roelofsen (2017) observe that if the embedded clause is preposed, the intonation difference is preserved, and rising declaratives pattern with polar interrogatives in these cases:

- (32) a. 'It's raining↘,' it appears/*she wondered.
 b. 'Is it raining?,' *it appears/she wondered.
 c. 'It's raining↗,' *it appears/she wondered.

Farkas & Roelofsen (2017) use this evidence to argue that rising declaratives are similar to interrogatives and that this similarity should be captured in part by syntax. They propose that the syntactic representation of sentence forms have two kinds of clause type markers, CLOSED/OPEN and DEC/INT, as in (3.1). Although they didn't specify where these markers reside in the syntactic structure, but they seem to adopt a split-CP approach following Rizzi (1997), and thus the two clause markers are both part of CP.

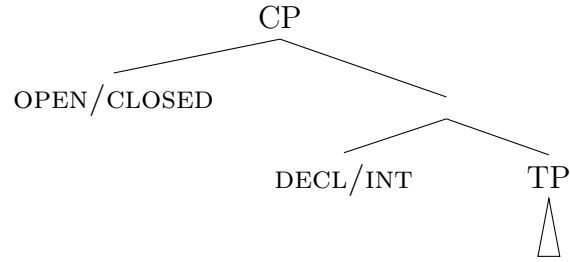


Figure 3.1: Extended structure of CP proposed by Farkas & Roelofsen (2017)

While the morpho-syntax of a sentence is linked to DECL/INT, the intonation (rise vs. fall) signals whether C bears the feature [OPEN] or [CLOSED]. Rising declaratives like (28) would be an “open declarative,” with the following structure:

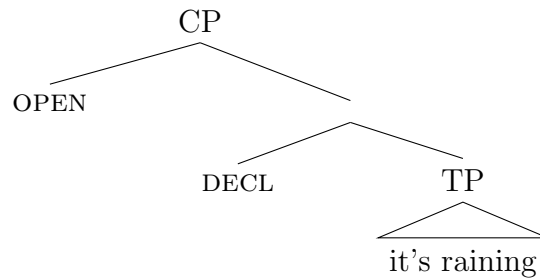


Figure 3.2: The structure of rising declaratives, adapted from Farkas & Roelofsen (2017)

So for Farkas & Roelofsen (2017), the clause type category of rising declaratives is not declarative with [−int] (but rather [OPEN,DEC]). Both OPEN and [+int] are equivalent to a Q-morpheme, so the semantics of rising declaratives is similar to that of polar interrogatives.

However, as many have noted, there might be two types of rising declaratives. The one discussed in Farkas & Roelofsen (2017) is the inquisitive one, but there might an assertive one (Jeong 2018, Goodhue 2021), as shown in (33b):

- (33) a. *S and A are on their way to a birthday party for the daughter of A’s friend. They stop at a store to get a birthday card. As they are both scanning the display for a card for the correct age, S is trying to remember how old the girl has just turned,*

and he thinks he remembers A telling him that she just turned nine, but he wants to confirm it.

S: She's nine ↗

- b. *S is enrolling his daughter in a summer camp program with the camp organizer A.*

S: I want to sign her up for Spanish classes in the mornings, and rock climbing in the afternoons.

A: Okay, there are limited places in each activity based on age group, and some of the age groups have already filled up for rock climbing. How old is your daughter?

S: She's nine ↗

ex. (12-13), Goodhue 2021:955

In (33a), the speaker uses the utterance to elicit a confirmation from the addressee regarding the age of the birthday girl, so the main goal is to solicit responses, similar to that of questions. But in (33b), the speaker uses the rising declarative to answer the question raised by the addressee, proposing to add the proposition *she's nine* to the common ground of the conversation. Even though there is an additional effect associated with the utterance (something to the effect of “Is there still room in the 9-year-old’s rock climbing group?”), this additional effect is not the main goal of the utterance.¹ Thus, we might not need to assume an interrogative semantics for rising declaratives.

Moreover, the pre-posing diagnostics that Farkas & Roelofsen 2017 use might not be conclusive. As pointed out by Rudin 2019, these cases of embedded rising declaratives can only be quotatives:

(34) *Context: Alvin (A) is talking to Bertha (B) about a conversation he had with Cynthia*

- a. A: Cynthia asked me if you're married.

You = A ✗; You = B ✓

¹There is another incredulous use of rising declarative, but the two uses mentioned here are more relevant for the current discussion on the speech act of rising declaratives, as most agree that incredulous rising declarative is also “inquisitive”, same as (33a).

b. A: Are you married ↗, Cynthia asked me.

You = A ✓; You = B ✗

For embedded clauses that do not have rising intonation like (34a), the embedded subject second person pronoun can only refer to the addressee. But when the embedded clause is associated with a rising intonation like (34b), the embedded subject second person pronoun can only refer to the speaker, and cannot refer to the addressee. This suggests that the conclusion drawn from the test with pre-posed embedded clause might need to be re-examined.

König & Siemund (2007) additionally note that interrogative clauses tend to be able to license polarity items like *ever*, *any*, but rising declaratives cannot license these items:

- (35) a. Have you ever met him?
b. *You have ever met him ↗

While these pieces of evidence suggest that rising declaratives might not be interrogatives, it technically does not eliminate the possibility that rising declaratives are their own clause type. But as Goodhue 2021 demonstrate, it is possible to derive the right semantics for all types of rising declaratives without assuming an additional OPEN/CLOSED layer in the syntax, the motivation for proposing a separate clause type for rising declaratives is also eliminated. Therefore, in this dissertation, I assume that rising declaratives are declaratives with [–int] in C^0 . I will return to prosody in Chapter 5.

Imperatives (sometimes analyzed as C having the feature [imp], Platzack & Rosengren 1997) in English typically use a bare verb stem. In most cases, subjects are missing (36a), but sometimes there are subjects expressing the addressee (36b).

- (36) a. Be quiet!
b. You be quiet!

In sum, declaratives ([–int] feature in C) in English are the unmarked clause type; interrogatives ([+int] in C) are associated with subject-auxiliary inversion, the presence of clause-initial *wh*-phrases, and the presence of the complementizer *whether*; imperatives

([imp]² in C) are marked by using verb stems, and by the absence of sentential subjects or having second person pronouns as subjects. So, to successfully infer the clause type categories, learners have to pay attention to morpho-syntactic features that include the presence or absence of the sentential subject, the form of the verb, the position of the auxiliary, the presence or absence of *wh*-phrases in clause-initial position, and the choice of complementizers in the surface form of sentences. Table 3.1 summaries the morpho-syntactic features and their associated clause types:

Feature	Examples
\pm VERB	(+) Find Elmo! Clause (-) Elmo! Fragment
\pm SUBJECT	(+) I 'll take it. [-imp] (-) Take it. [+imp]
\pm VERB SUFFIX	(+) Nobody feels good huh? [-imp] (-) Find Elmo! [+imp]
\pm SUBJ-AUX INVERSION	(+) Can you find the ladybug? [+int] (-) I can take it. [-int]
\pm SENTENCE-INITIAL <i>wh</i>	(+) What did you find? [+int] (-) I found it. [-int]
COMPLEMENTIZER	(+) I know whether it's wrong. [+int] (-) I know that it's wrong. [-int]

Table 3.1: Morph-syntactic features and their associated clause types

Of course, as we have discussed, many features do not have a one-to-one mapping with the abstract clause type categories. So apart from using the right features to cluster clauses, learners also need to avoid making certain generalizations about a feature in some cases (e.g. avoid associating subject-auxiliary inversion with declaratives upon seeing Negative Inversion sentences).

While these features are significant in the grammar, it's likely that not all of these features (or the full-fledged version of these features) can be perceived by 18-month-olds. If we want to model how 18-month-olds learn clause types, the way we code these formal features in

²Since we are always comparing imperatives with declaratives, I will sometimes use [+imp] as the feature for imperatives and [-int, -imp] for declaratives. But this is not to make a theoretical claim about the hypothesis space of C^0 , as it is unclear which clause type corresponds to [+imp, +int].

our corpus (which serves as the input to our models) has to be sensitive to their linguistic knowledge, and what they are able to perceive at the relevant age. In the next section, I will review what linguistic knowledge and capacities children have around 18 months with regard to these formal features.

3.1.3 Linguistic knowledge and capacities of 18-month-olds

Generally, infants have been shown to track distributional properties of various kinds, and 18-month-olds can perceive many grammatical features. How much do they know about the ones that are associated with English declaratives, interrogatives, and imperatives (Table 3.1)? In this section, I'll briefly review 18-month-olds' capability regarding the features reviewed in the last section, and along the way explain how we code these features in our corpus.

± VERBS: By 18 months, infants can recognize whether there is a verb in a sentence. They can use the frequent frames (e.g. adjacency to function words like auxiliaries, pronouns, or have affixes) associated with verbs to categorize a novel word as a verb (Echols & Marti 2004, Mintz 2006, Peterson-Hicks 2006, Soderstrom et al. 2007, Lidz, Omaki, & Orita 2012, Shi 2014, He & Lidz 2017 among many others), suggesting that they have knowledge of verbs, and frequent verb frames.

± VERB SUFFIXES: Around this age, infants can recognize some morphological markings on the verb. As early as 6 months old, English-learning infants seem to be able to segment *-s*, *-ed*, *-ing* from a nonce verb (Kim & Sundara 2021). 15-month-olds can segment English verbal suffix *-ing* from a word, but do not do so with pseudo-suffixes (Mintz 2013); 18-month-olds can distinguish well-formed auxiliary-affix dependencies (e.g. *is ...-ing*) vs. an ill-formed ones (e.g. *can ...-ing*, Santelmann & Jusczyk 1998). Thus, we can assume that 18-month-olds can tell whether verb suffixes are present in a sentence.

± SUBJECT: Some studies have shown that infants have some knowledge about subjecthood by 18 months old. They show sensitivity to subject-verb agreement in English, as they prefer grammatical sentences over ungrammatical sentences with agreement violation

(Soderstrom, Wexler, & Jusczyk 2002, Soderstrom et al. 2007, Nazzi et al. 2011 among others), even if the subject is not immediately adjacent to the verb. They can also use the frame [subject pronoun + verb] to categorize novel words as verbs (Babineau, Shi, & Christophe 2020, Peterson-Hicks 2006, Mintz 2006, Shi 2014 among others). Moreover, and perhaps more importantly for our study, 12-month-olds are sensitive to the change in word order when the position of the subject and auxiliary is switched (Geffen & Mintz 2015). While we do not have evidence that 18-month-olds can use the presence/absence of subjects to tell whether a sentence is imperative or not, we can assume that they can detect whether the subject is present or not.

± SUBJ-AUX INVERSION: By 18 months, infants can recognize whether there is an auxiliary in the sentence, and are sensitive to the relative word order between the auxiliary and the sentential subject. As mentioned above, well before they turn 18 months old, infants can already use preceding auxiliaries to categorize a novel word into the verb category (e.g. Peterson-Hicks 2006, Mintz 2006), suggesting that they understand the relation between auxiliaries and verbs. Additionally, as mentioned above, 12-month-olds can already detect subject-auxiliary inversion (Geffen & Mintz 2015, cf. Erreich 1984, Ambridge et al. 2006); with both word order and prosodic cues, even 7-month-olds are able to detect the difference between polar interrogatives and declaratives (Geffen & Mintz 2011). We therefore assume that infants are able to detect whether there is an auxiliary in the sentence, and further that they are able to detect whether the subject is preceding or following the auxiliary.

± SENTENCE-INITIAL *wh*: Perkins & Lidz (2021) shows that 18-month-olds, but not younger infants, begin to represent the sentence-initial *which NP* as the head of a long-distance dependency. These results are suggestive, but we do not yet have evidence for the full range of *wh*-phrases. In order to be conservative, we do not assume that infants have full-fledged knowledge of *wh*. However, infants at this age have been shown to have the ability to tease apart functional from content words based on their acoustic and phonological properties (Shi, Werker, & Morgan 1999, Shi 2014) and it is reasonable to assume that they

treat *whs* as functional items (Perkins 2019). Therefore, we will consider *wh* to be unknown to 18-month-olds, but still treat them as unknown *functional* items (UFI), following Perkins (2019).³ We can also assume that 18-month-olds can keep track of the position of these UFIs in a sentence, so we can code *wh*-items as sentence-initial UFIs.⁴

COMPLEMENTIZER: Complementizers indicate the clause type of the embedded clause instead of the matrix clause. Thus, if we focus on matrix clauses to investigate the acquisition of sentential force, complementizers might not be relevant for learner. Nonetheless, I annotated these as UFIs as well. Since infants can identify verbs, we assume that they can track that certain UFIs occur post-verbally.

Finally, infants can perceive some features of prosody. We will come back to this feature in Chapter 5.

In summary, apart from clause-initial *wh*-phrases and complementizers (which we will assume to be perceived as clause-initial UFIs), 18-month-olds have the formal features relevant for clause typing in English. But how do parents use clause types and speech acts? How serious is the many-to-many mapping problem, both for the mapping between clause type and speech act, and for the mapping between formal features and clause type? We will address these questions with a corpus study.

3.2 Corpus study

To understand how infants figure out clause types and speech acts, we need to not only know which features they can in principle understand, but also we need to establish what kind of

³Note that Perkins (2019) models what happens before 18 months, as her experiment narrows down the developmental window to 18 months old. Moreover, while her experiment only tests *which*, results from productions studies (e.g. Clark & Lindsey 2015) suggest that 18-month-olds might also understand *what*, *who*, and *where*. Thus, assuming that all *wh*-items are unknown functional items might in fact be too conservative. In the future, I plan to see if loosening this assumption (e.g. assume that some *wh*-items are known) will boost the performance of the SID.

⁴Other functional items for which we lack evidence for 18-month-olds knowing them include quantifiers, focus particles, complementizers (e.g. *whether*), and certain connectives such as *because*. I assume that these are UFIs as well.

information is even *available* to them in the input. In this section, I'll report findings from the corpus study on parental input to English-speaking infants.

3.2.1 Corpus and methods

For this study, we used data from the Providence sub-corpora (Demuth, Culbertson, & Alter 2006) from CHILDES (MacWhinney 2000), which consists of parent-child interactions from 6 families recorded between 2002-2005 in Providence, RI. We selected this particular corpus because it covers children's interaction with parents at the critical age that we are interested in (≤ 18 months old). Additionally, it contains transcripts, audio, and video data, providing us with an opportunity to not only look at the morpho-syntax of parents' sentences, but also their prosodic information, and even parents' behavior accompanying each utterance. Parent-child interaction data from five typically developing children in this corpus were included in this study: Alex, Lily, Naima, Violet, and William.

We sampled 500 conversational turns from each session, and annotated the resulting dataset with the annotation schema detailed in the next section.⁵ For the clause type and speech act information, each annotator completed a subset of transcripts (20% were double annotated, mean Cohen's $\kappa = 0.84$, range 0.8-0.91, 'almost perfect' per Landis and Koch's (1977) descriptive division). To annotate the speech act information, annotators were asked to look at 20 utterances before and 2 utterances after the current utterance in the conversation, as well as consulting the videos for contextual information. If the annotator could not decide on an annotation despite the contextual information, a second person would be consulted, and if still unclear, the utterance would be eliminated. 16 utterances were eliminated in this way. For the morpho-syntactic features, initial annotation was generated by a script using the morphological tagging provided by CHILDES, and then manually corrected.

In total, 9047 utterances were annotated. As we are primarily interested in the form and function of parents' sentences, we excluded children's utterances and uninterpretable

⁵The full annotation schema can be accessed here: https://github.com/Yu-an/annotation_tool/blob/main/schema.pdf

utterances from the dataset. In total, 7039 utterances were analyzed. We will return to the video and audio part of the annotation process in Chapter 6.

3.2.2 Annotation schema

3.2.2.1 Clause Type

All sentences in our sample were annotated with clause type information: declarative (37a), interrogative (37b), or imperative (37c). Minor clause types like exclamation were annotated as Other (37d), as our primary interest is on the three major clause types (and we do not have evidence for 18-month-olds' understanding of minor clause types). Two other categories were also included: Fragments and Ambiguous cases. In cases where the utterance only contains a noun or an interjective without verbs like (37e), the utterance was annotated as a Fragment:

(37) Clause type

a. It's all twisted.	Declarative
b. What happened?	Interrogative
c. Throw it.	Imperative
d. What a nice day!	Other
e. Elmo!	Fragment
f. Wanna get down?	Ambiguous

Ambiguous cases are sentences that do not contain enough information to decide the clause type categories. For example, sentences like (37f) could either be a case of ellipsis from the declarative sentence *you want to get down*, or ellipsis from the polar interrogative *do you want to get down*. In both cases, there is no overt morphological marking on the verb *want*, so after going through the ellipsis operation, the two sentences would end up with the same surface form. If the verb has an overt suffix, however, the two possibilities would give rise to different surface forms. For example, (38a) has to be a case of left-edge ellipsis from

polar interrogative *are you coming*, and cannot be from a declarative sentence like *you come*, as ellipsis the latter would not give us the *-ing* morpheme on the verb. For the same reason, (38b) cannot be elided from polar interrogative *Did you go to the store yesterday*. Thus, the two sentences in (38) were annotated as interrogative and declarative respectively, but (37f) was annotated as Ambiguous.

- (38) a. Coming?
 ✓ *Are you coming?*
 ✗ *You come?*
- b. Went to the store yesterday?
 ✓ *You went to the store yesterday?*
 ✗ *Did you go to the store yesterday?*

Interrogatives were further divided into subcategories as polar (39a), *wh* (39b), and disjunctive interrogatives (39c):

- (39) Sub-types of interrogatives
- | | |
|--------------------------------|---------------------------|
| a. Is that a big bird shovel? | Polar interrogative |
| b. Who is that? | <i>wh</i> -interrogative |
| c. Do you want water or juice? | Disjunctive interrogative |

3.2.2.2 Speech Act

Three major speech act categories, assertions, questions and requests/commands, were labeled; minor speech act categories such as exclamations and greetings were labeled as ‘Other:’

- (40) a. It’s all twisted! Assertion
- b. Is that the postman? Question
- c. Throw it! Request/command
- d. Good morning! Other

As discussed in the last chapter, while the three major clause types are associated with the three major speech acts in most cases, there are some mismatching cases. For example, (41) is a polar interrogative, but the non-literal (i.e. indirect speech act in Searle’s (1975b) terminology) act of the utterance is to make a request. In such cases, we annotated the utterance with the non-literal (indirect) speech act, i.e. (41) was annotated as a request.

(41) Can you put that down?

As discussed in Section 3.1.2, there is a live debate on whether to label these cases as declaratives, interrogatives or some more sophisticated category (Gunlogson 2008, Farkas & Roelofsen 2017).

(42) it is raining ↗

We followed Gunlogson (2008) and Goodhue (2021) in classifying the form of (42) as declarative, and labeled the speech act as either question or assertion depending on the contextual information, following Jeong (2018),⁶ Goodhue (2021). We will return to prosody in Chapter 6.

3.2.2.3 Formal features

As reviewed in the last two sections, infants at 18 months old can perceive many morpho-syntactic cues related to clause typing, as recorded in Table 3.1. In particular, infants at this age can detect whether there is a verb in the sentence; whether the verb has suffixes; whether there is a subject in the sentence, and its position in the sentence (whether in canonical pre-verbal position or inverted with an auxiliary); and whether there is an auxiliary in the sentence. Infants might not be able to identify all the *wh*-items or complementizers at this age, but they might be able to classify them as functional elements, as they may know the distinction between functional and content elements. We therefore put *wh*-items, quantifiers,

⁶Note that she follows Farkas & Roelofsen (2017) in assuming that the clause type of rising declaratives is [open, –int], different from my assumption.

connectives (except for *and*), and focus particles in one category “unknown functional item (UFI),” and annotated its position in a sentence: sentence initial, sentence-medial but before the verb, or after the verb. Each sentence was annotated with whether or not a specific cue is present. Table 3.2 summarizes the features we annotated and their examples.

	Examples
\pm Verb	(+) Find Elmo! (-) Elmo!
\pm Subject	(+) I ’ll take it. (-) \emptyset Take it.
\pm Verb Suffix	(+) Nobody feels good huh? (-) Find \emptyset Elmo!
\pm Auxiliary	(+) Can you find it? (-) I found it!
\pm Subj-aux inversion	(+) Can you find the ladybug? (-) I can take it.
\pm Sentence-initial UFI	(+) What did you find? (-) I can take it.
\pm Pre-verbal UFI	(+) Raccoon only comes out at night. (-) I can take it.
\pm Post-verbal UFI	(+) I know what ’s wrong. (-) I know you can do it.

Table 3.2: Morpho-syntactic cues and their examples

3.2.3 Results

3.2.3.1 Overview

In total, 7039 utterances were analyzed. Figure 3.3 shows the distribution of clause types in the dataset. Declarative clauses are the most frequent clause type, followed by interrogatives and imperatives.

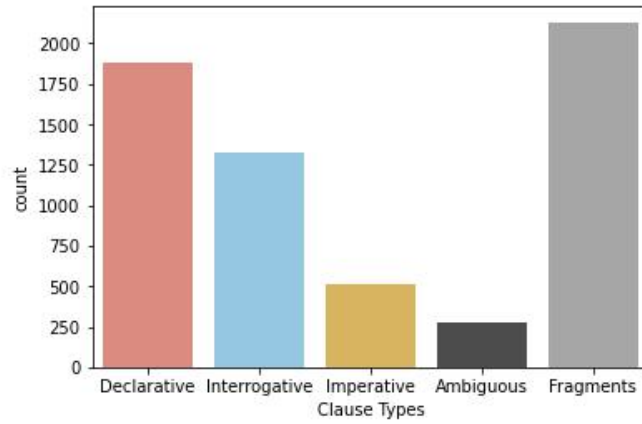


Figure 3.3: Distribution of clause types in the corpus

Zooming in on interrogatives, *wh*-interrogatives are more frequent than polar interrogatives; only 2 cases of disjunctive interrogatives were found in the dataset. Figure 3.4 shows the distribution of the subcategories of interrogatives.

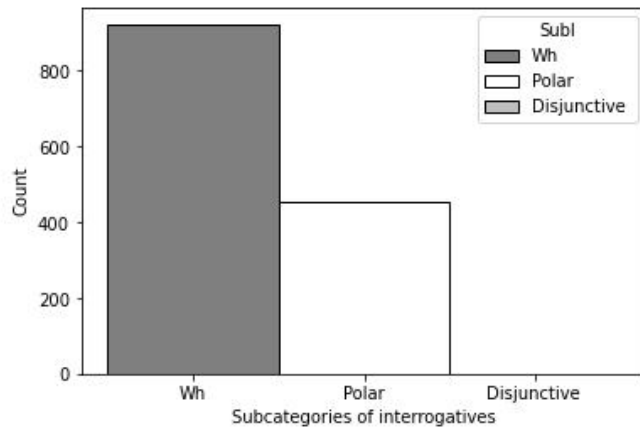


Figure 3.4: Subcategories of interrogatives

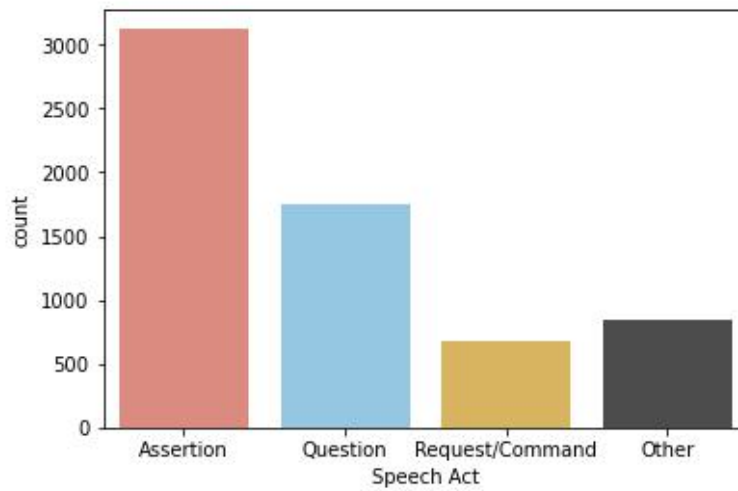


Figure 3.5: Distribution of speech acts in the corpus

As shown in Figure 3.6 and 3.7, the three major clause types are mostly used for their canonical functions. The majority of declaratives are used to express assertive force; the

majority of interrogatives are used to express question force; and the majority of imperatives are used to express command/request force.

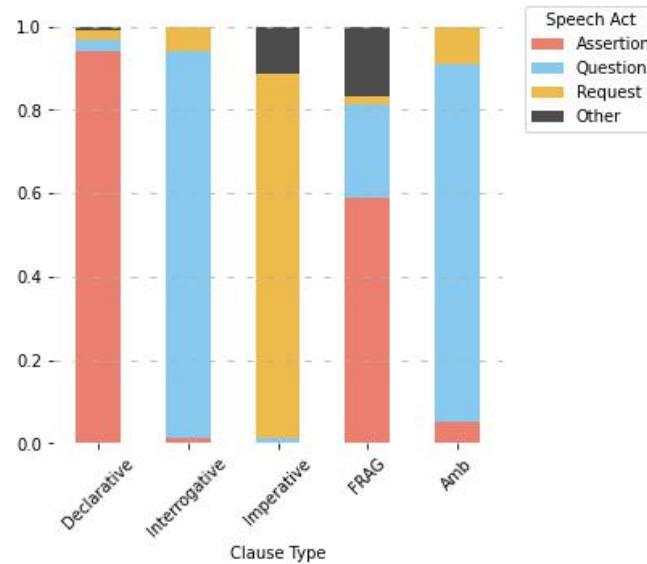


Figure 3.6: The speech acts performed by each clause type in parents' speech

Conversely, as shown in Figure 3.7, the majority of assertions are expressed by declaratives, the majority of questions by interrogatives, and the majority of commands/requests by imperatives.

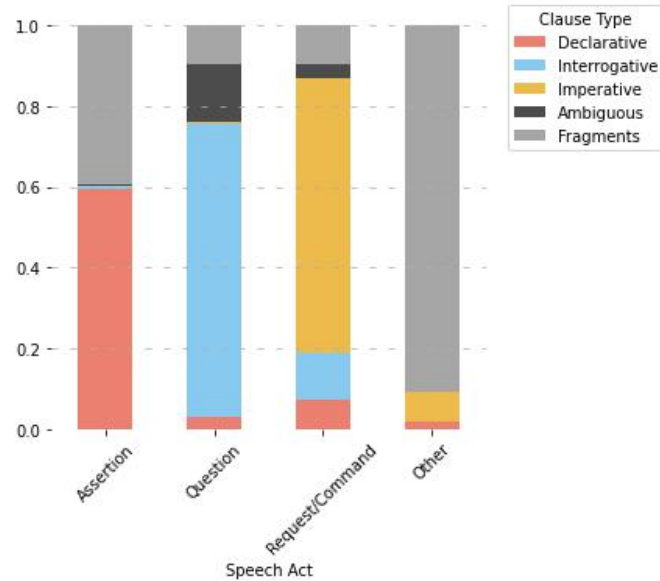


Figure 3.7: The clause type used to express each speech act in parents' speech

The few cases of mismatches between clause type and force appear to show that such mismatches are systematically marked by morpho-syntactic features, similar to what Zaitsev et al. (2020) have found. For example, declarative and interrogative sentences used for requests/commands tend to have modals, attitude verbs, or future morphology in the sentence:

- (43) a. I need you to help me. Mother of William, Session 010605
 b. Can you say hi? Mother of Lily, Session: 010117
 c. (previous utterance: I'm gonna do some work.)
 And you're gonna do some coloring. Mother of Violet, Session: 010407
 d. Are you gonna read to Mommy? Mother of Lily, Session 010102

Declarative and imperative sentences used as questions are predominantly marked with final rise intonation. We will return to the prosodic features of each clause type in Chapter 6.

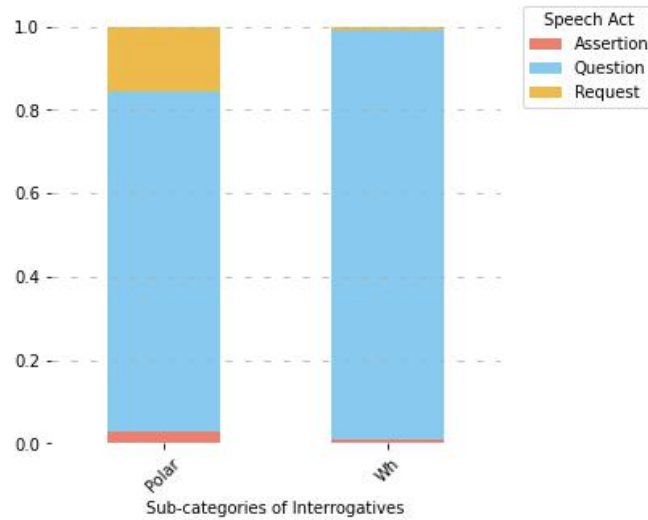


Figure 3.8: The speech act expressed by different subcategories of interrogatives

Turning now to the speech acts expressed by interrogatives, Figure 3.8 shows the proportion of speech acts that polar and *wh*-interrogatives express. The only two examples of disjunctive interrogatives were used both as questions, and are thus not graphed in this figure. Polar interrogatives are used to express more types of speech acts than *wh*-interrogatives: around 18% of polar interrogatives were used for indirect requests like in (44) or assertions like in (45):

(44) Can you give Mommy the ball? Mother of William, Session: 010412

(45) Doesn't he have sharp teeth? Mother of Lily, Session: 010611

Few *wh*-interrogatives are used as non-question. We found 4 cases of *wh*-interrogatives used to indirectly make a request (46), and indirectly making assertions by way of asking rhetorical questions (47):⁷

⁷We did also find sentences with *how about/what about* (Rawlins and Bledin 2021), but the majority of

- (46) Why don't you turn around. Mother of William, Session 010619
- (47) *wh*-interrogatives as assertions
- a. Who doesn't love cheese? Mother of Lily, Session: 010117
- b. Why are there no pens in this family? Mother of Violet, Session: 010407

In sum, clause types are typically used for their canonical function in the input to children. This suggests that having speech act information could indeed be more helpful than hurtful for learning to categorize clauses into clause types.

3.2.3.2 Morpho-syntactic cues

We now turn to the formal cues of each clause type, as reviewed in Section 3.1.2. In this analysis, we excluded fragment utterances that only contain a noun (*Birdie!*) or an injective (*Oh*). In total, 3923 sentences were included in the analysis. The distribution of the morpho-syntactic cues listed in Table 3.2 across different clause types is shown in Figure 3.9. Darker colors represent the number of sentences with the morpho-syntactic property, and lighter colors represent sentences without the property.

these sentences do not come with a verb and were considered fragments (i). But we also found one case of *how about* interrogative with a verb (ii), and the primary intention of the utterance is to make a request.

- i. How about this one? Mother of Alex Session 010512
- ii. How about we do these babies. Mother of Lily, Session 010102

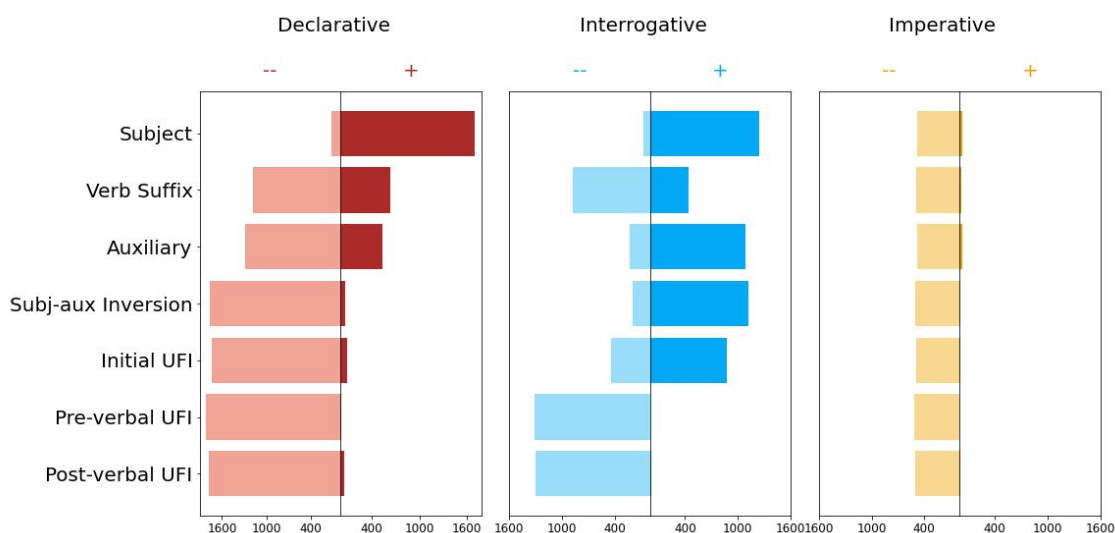


Figure 3.9: Number of sentences with/without various formal cues in each clause type; darker colors represent number of sentences with the cue, lighter colors, number of sentences without the cue

We can see that the three main clause types have different formal profiles in the input, but the general pattern aligns with the properties of $[\pm\text{int}]$ and $[\pm\text{imp}]$ discussed in Section 3.1.2: interrogativity is associated with subject-auxiliary inversion and sentence-initial *wh* (i.e. an unknown functional item for infants); $[\pm\text{imp}]$ is associated with the lack of verb suffixes (i.e. the use of bare verb stems) and subjects; declaratives, the default clause type, are associated with the presence of subjects and verb suffixes, and the lack of subject-auxiliary inversion and clause-initial UFIs.

From these results, we can see that the formal signatures of each clause type is consistently present in the input: auxiliary-inversion for interrogatives and the lack of subjects and verb suffixes for imperatives. But how informative are these cues to clause typing? And what can or does speech act information add to these cues? In the next section, I train two logistic regression models with actual clause type labels (i.e. using a supervised learning method in machine learning terms) to see how well the models find statistical regularities related to

clause typing in the data.

3.2.3.3 Informativeness of syntactic and pragmatic cues

In the previous section, we saw that in parents’ speech to 18-month-olds, the three clause types are mostly mapped to their canonical functions, and that they have different morpho-syntactic profiles as expected. In this section, I evaluate the informativeness of syntactic and pragmatic cues by comparing two supervised learning models. Unlike the models simulating the SID learner and the SPID learner I will discuss later, these two supervised models were trained on a subset of our annotated dataset with clause type labels given as input. That is, the models are given correctly annotated data for training (hence the term “supervised”) and attempt to determine the significant statistical regularities associated with each clause type. After the training phase, these models were tested on the part of data they have not seen, to see how well they could use the statistical regularities discovered in the training data to predict clause type labels.

The supervised SID model uses morpho-syntactic cues as predictors for clause typing, while the supervised SPID model uses both morpho-syntactic and speech act information. Both are multinomial logistic regression models trained on 90% of the annotated dataset, with clause type categories as dependent variable (declarative as the baseline), and a set of morpho-syntactic cues as predictors. As discussed in Section 3.1.2, [+int] is associated with the presence of cues (e.g. auxiliary inversion, clause-initial *wh*, complementizer), while [+imp] is associated with the absence of cues (e.g. verb suffix, subject). So the set of syntactic predictors in our models is [-subject, -verb suffix, +auxiliary, +subj-aux inversion, +clause-initial UFI, +pre-verbal UFI, +post-verbal UFI]. We then test the performance of both on the unseen 10% of the data to see how well they predict the clause type categories. We calculated the adjusted rand score of both models to evaluate their performance. The rand index is normally used to measure the similarity between two data clusterings (Rand 1971), calculated by taking all pairs of samples and counting ones that are assigned to the same or

different clusters in the predicted and true clusterings:

$$\text{RI} = \frac{\text{number of agreeing pairs}}{\text{number of pairs}} \quad (3.1)$$

The adjusted Rand index (ARI) is the corrected-for-chance version of the Rand index (cf. Hubert & Arabie 1985, Steinley 2004):

$$\text{ARI} = \frac{\text{RI} - \text{Expected RI}}{\max(\text{RI}) - \text{Expected RI}} \quad (3.2)$$

The performance of the two models over 10 iterations is shown in Figure 3.10. Overall, the supervised syntactically and pragmatically informed distributional learner outperforms the supervised syntactically informed distributional learner, suggesting that the speech act information provides additional information for clause typing.

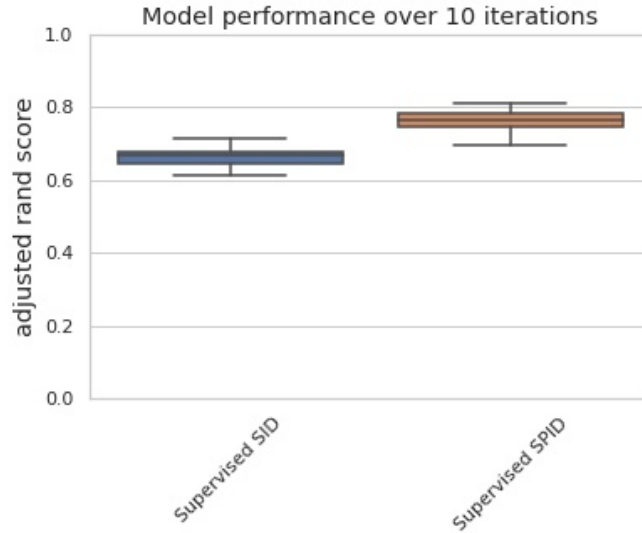


Figure 3.10: Comparing the two supervised models by rand score (performance over 10 iterations)

Our next question is how well the models know the morpho-syntactic makeup of each clause type. This can be inferred from the coefficients that the two models assign to the morpho-syntactic predictors. A bigger coefficient assigned to a predictor means that the model assigns it a higher significance. As declarative clauses were used as the baseline, the coefficients should in particular be interpreted as indicating how much a specific morpho-syntactic cue contributes to the increase/decrease in the log odds of the sentence being classified as interrogative/imperative vs. declarative (i.e. if a sentence has a morpho-syntactic cue, how likely will it have the feature [+int] or [+imp] as opposed to the default clause type [-int, -imp]). Table 3.3 summarizes the morpho-syntactic profile of each clause type. For our purposes, significance was calculated using a two-tail z -test.

Clause Type Feature	Morpho-syntactic cues
Interrogatives (+int)	+auxiliary, +subject-auxiliary inversion, +sentence-initial UFI
Imperatives (+imp)	–subject, –verb suffix

Table 3.3: Morpho-syntactic cues associated with interrogatives and imperatives compared to declaratives, according to the supervised SPID model

As seen in the table, compared to declaratives, interrogatives are more likely to have subject-auxiliary inversions and sentence-initial UFIs; imperatives are more likely to have null subjects and bare verb stems, as expected. Some associations were less expected. For example, [+int] is associated with [+auxiliary], which is not predicted by our theory of clause typing in English. But this in fact reflects another formal property of English [+int], namely *do*-support. Whenever T and V are not syntactically adjacent (either because T to C movement or negation in declaratives and imperatives), *do* is inserted. Since T to C movement is common in interrogatives, but negation is less common in imperatives and declaratives, this elevates the rate of auxiliary in interrogatives relative to the other kinds of clauses.

Of course, these two models do not reflect how infants learn clause typing, as they were given clause type information. Infants must learn clause typing without having access to data of which the clause type label is given. These supervised models are nevertheless useful for appreciating the information contained in the data that children receive as input. They allow us to measure the correlation between the features and clause type categories, which provides an upper limit on how much information could be conveyed by the morpho-syntactic cues.. The results from these two models suggest that the information needed for clause typing is indeed available in the input and, in particular, from the categories with which we annotated our data. In the next section, we will get one step closer to addressing how infants learn clause types by removing the actual clause type labels from our models (switching to what in machine learning is known as “unsupervised learning”). This makes the learning models

more similar to the actual learning task where infants are not given these labels. Using these models, we can ask whether a learner that expects three clause type categories would be able to solve the clustering problem discussed in Chapter 1 on the basis of morpho-syntactic cues alone. We then consider the question of whether adding in speech act information would help. Finally, as speech act information is likely not perceived veridically (e.g. due to developing pragmatic skills at this age), I ask how much pragmatics learners need to learn clause type categories.

3.3 Modeling the learning of clause types

Now that we have seen that the relevant information for clause typing is present in the input, our next step is to see whether a learner that can only perceive the morpho-syntactic cues would have enough information to find the right clause types.

As discussed in Section 3.1.1, I examine the role of pragmatics computationally by building two Bayesian clustering models. The two models simulate a syntactically informed distributional learner and a syntactically and pragmatically informed distributional learner. Both are distributional learning models: as input, they observe the distribution of several morpho-syntactic cues that are identifiable by infants at 18 months old, and they use that information to cluster sentences into three clause type categories. Given that, cross-linguistically, the same three types of clause types (declaratives, interrogatives, and imperatives) are associated with three major speech acts (assertions, questions, and requests/commands), it is reasonable to assume that an expectation for three clause types is innate. Even if learners do not have this innate expectation, determining whether learning can succeed with this expectation is a first step in determining whether it is possible to learn the clause types without it. Our learners thus assume that there are three categories, and only have to discover what these categories are, based on formal features. The syntactically and pragmatically informed distributional learner additionally observes speech act information and assumes that it is re-

lated to clause type. We further manipulated the ratio of noise in the pragmatic information that the pragmatically informed learner is given, to test how much pragmatics the learner needs.

This section is organized as follows. Section 3.3.1 specifies the generative models simulating the two learners. Then I specify how the two models infer clause type categories in Section 3.3.2. Section 3.3.3 reports the simulations with English infant-directed speech demonstrating that pragmatic information is crucial for finding the right clause type clustering, and that this pragmatic information does not have to be perfect, as even with 70% of noise, it still boosts the performance of the syntactically and pragmatically informed distributional learner.

3.3.1 Generative models

The two models simulating the syntactically informed distributional learner and the syntactically and pragmatically informed distributional learner are Bayesian clustering models, illustrated in Figure 3.11.

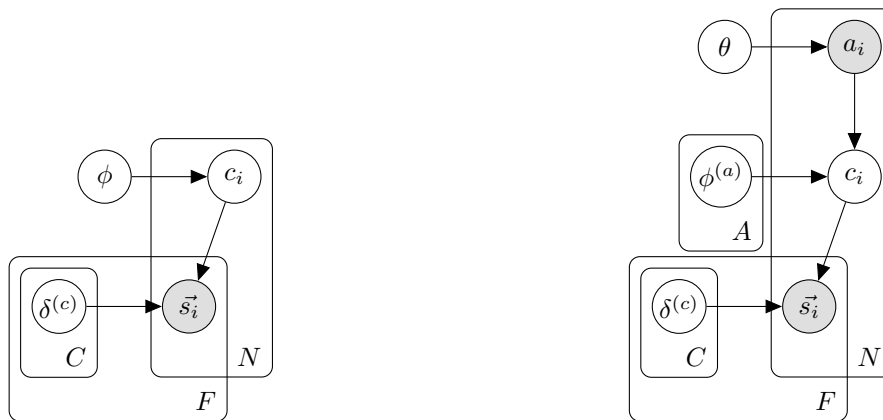


Figure 3.11: The syntactically informed distributional learner (left) and syntactically and pragmatically informed distributional learner (right)

Both models assume that there are three clause type categories, declaratives, interrogatives, and imperatives, that explain the distribution of a set of morpho-syntactic properties in

the input. This is shown in the graphical illustrations in Figure 3.11 as the morpho-syntactic properties \vec{s}_i being conditioned on the clause type variable \vec{c}_i . The SPID model additionally assumes that there are four speech act categories (assertions, questions, requests/commands, or some other speech acts such as exclamatives), and that this pragmatic information helps to identify the clause type category. This is shown in the graphical representation as the variable c_i being conditioned on the variable a_i (for speech act). Tables 3.4 and 3.5 summarize the notation and statistical assumptions for the two models.

c_i	Clause type, declarative, interrogative, imperative $\vec{\phi}^a \sim \text{Dir}(1, 1, 1)$ $c_i \sim \text{Multinomial}(\vec{\phi})$;
\vec{s}_i	Morpho-syntactic properties; See Table 3.2 for the full list of features included $\delta \sim \text{Beta}(1, 1)$ $s_i^{(F)} \sim \text{Bernoulli}(\delta^{(c)})$

Table 3.4: Variables in syntactically informed distributional learner model, their distribution, and explanation

a_i	Speech Act: Assertion, Question, Request/Command, Other $\vec{\theta} \sim \text{Dir}(1, 1, 1, 1)$ $a_i \sim \text{Multinomial}(\vec{\theta})$;
c_i	Clause Type: Declarative, Interrogative, Imperative $\vec{\phi}^{(a)} \sim \text{Dir}(1, 1, 1)$ $c_i \sim \text{Multinomial}(\vec{\phi}^{(a)})$;
\vec{s}_i $s_i^{(F)}$	Set of morpho-syntactic properties; property F in the set; $\delta \sim \text{Beta}(1, 1)$ $s_i^{(F)} \sim \text{Bernoulli}(\delta^{(c)})$

Table 3.5: Variables of the SPID model, their distributions, and explanations

In both models, the clause type variable c_i follows a multinomial distribution with a parameter ϕ . This parameter is assumed to have a uniform Dirichlet prior $\text{Dir}(1, 1, 1)$. This means that it is equally likely *a priori* for a sentence to be a declarative, interrogative or

imperative. This could be interpreted as the learner do not assume any bias at the very beginning. For the model simulating the syntactically informed distributional learner, the parameter ϕ represents the probability that a sentence i is a declarative, interrogative, or imperative. As the SPID model assumes that the speech act category a_i of the utterance expressed by the sentence informs the learner of the clause type category, so the parameter $\phi^{(a)}$ in this model represents the probability that a sentence i , which expresses the speech act a_i , is a declarative, interrogative, or imperative.

In the SPID model, the speech act variable a_i follows a multinomial distribution with parameter θ . This parameter represents the probability that the speech act expressed by the sentence i is an assertion, question, request/command, or some other speech act. This parameter also has a uniform Dirichlet prior $\text{Dir}(1, 1, 1, 1)$, meaning that we assume that it is equally likely *a priori* that the sentence i expresses an assertion, question, request/command, or some other speech act.

In both models, morpho-syntactic properties are represented by a vector of Bernoulli variables \vec{s}_i , and \vec{s}_i is conditioned on c_i . The number of morpho-syntactic properties included is represented by a vector F , and the properties are the same as the ones in the corpus study listed in Table 3.2 in Section 3.2.2. Each $s_i^{(F)}$ in \vec{s}_i takes the value 1 if the sentence contains the property, and 0 otherwise. $s_i^{(F)}$ is conditioned on the parameter $\delta^{(c,F)}$, which represents the probability of observing a given morpho-syntactic property in a particular clause type category. This parameter is assumed to have a uninformative uniform $\text{Beta}(1, 1)$ prior, meaning that *a priori*, it is equally likely for a particular property to be present or absent in a clause type category.

3.3.2 Inferences

The task of both learners is then to infer the category assignment of the clause type variable (i.e. is the sentence a declarative, interrogative, or imperative) from observations of morpho-syntactic properties of the sentence. The SPID learner additionally observes speech act

information and assumes it is related to clause type.

I applied Gibbs sampling (Geman & Geman 1984), a form of a Markov chain Monte Carlo sampler to sample from the posterior distribution of clause type category assignments. The sampling algorithm of the two learners are detailed below. In the following algorithms, \vec{c} represent the set of clause type assignments in the corpus. The variable \vec{a} represents the set of all speech act values in the corpus; \vec{S} represents all values in the corpus for the set of morpho-syntactic properties, and $s^{(F)}$ represents all the values in the corpus for a specific morpho-syntactic feature F . The subscript i represents the current sentence, and $-i$ represents all values of a variable except the current sentence (for example, c_i represents the clause type assignment of the current sentence, and c_{-i} represents the clause type assignments of all sentences in the corpus except the current one). The following counts are calculated:

- (48) n is total number of observations so far;
- n_{c_i} is the number of observations of $c = c_i$ so far;
- n_{a_i} is the number of observations of $a = a_i$ so far;
- $n_{c_i}^{(a_i)}$ is the number of observations of $c = c_i$ while $a = a_i$;
- $n_{s_i}^{(F, c_i)}$ is the number of observations so far where $s^{(F)} = 1$ when $c = c_i$;
- $n_{c_i}^{(F)}$ is the number of observations of feature F so far when $c = c_i$

For both models, we first randomly initialize values of c_i for each sentence with three categories (representing declaratives, interrogatives, and imperatives). After initialization, the SID sweeps to update current c_i with the posterior distribution specified in Equation 3.3, and the SPID updates with Equation 3.4.

$$\begin{aligned}
 p(c_i | c_{-i}, \vec{S}_i) &= \frac{p(\vec{S}_i | \vec{c}) p(c_i | \beta, c_{-i})}{\sum_{c'_i} p(\vec{S}_i | \vec{c}') p(c'_i | c_{-i})} \\
 &= \frac{\prod_F \frac{1+n_{s_i}^{(F, c_i)}}{2+n_{c_i}^{(F)}} \frac{1+n_{c_i}}{4+n}}{\sum_{c'_i} \prod_F \frac{1+n_{s_i}^{(F, c'_i)}}{2+n_{c'_i}^{(F)}} \frac{1+n_{c'_i}}{4+n}} \quad (3.3)
 \end{aligned}$$

$$\begin{aligned}
p(c_i|c_{-i}, \vec{a}, \vec{S}_i, \delta) &= \frac{p(\vec{S}_i|\vec{c}) p(c_i|\beta, c_{-i}, \vec{a})}{\sum_{c'_i} p(\vec{S}_i|\vec{c}') p(c'_i|c_{-i}, \vec{a})} \\
&= \frac{\prod_F \frac{1+n_{s_i^{(F,c_i)}}}{2+n_{c_i^{(F)}}} \frac{1+n_{c_i^{(a_i)}}}{4+n_{a_i}}}{\sum_{c'_i} \prod_F \frac{1+n_{s_i^{(F,c'_i)}}}{2+n_{c'_i^{(F)}}} \frac{1+n_{c'_i^{(a_i)}}}{4+n_{a_i}}} \tag{3.4}
\end{aligned}$$

For each model, I ran the chain for 5000 iterations and analyzed the last sample. Each learner was simulated 10 times.

3.3.3 Simulations with infant-directed speech

The data for this model were taken from the annotated dataset reported in Section 3.2. The labels of speech acts and morpho-syntactic observations were used as input to the models, and the true labels of clause type were used to evaluate the performance of the models. As the learners need to use the surface features of sentences to learn about clause-level properties, instead of one-noun utterances or utterances of only injectives, we eliminated from the dataset sentences that do not contain a verb or an auxiliary. In total, 3923 sentences were fed into the models.

Figure 3.12 shows the adjusted rand scores of the two models over 10 iterations, and in comparison with the adjusted rand scores of the two multinomial logistic regression models ran in Section 3.2.3.3. As explained in the last section, this score ranges between 0 and 1, with higher scores indicating better performance.

Overall, the SPID model substantially outperforms the SID model. The pragmatic information led to an improvement from an average 0.4 adjusted rand score to an average of 0.7 rand score, bringing the model up to the level of supervised performance. This suggests that pragmatic information is crucial for learning the clustering of clause type information.

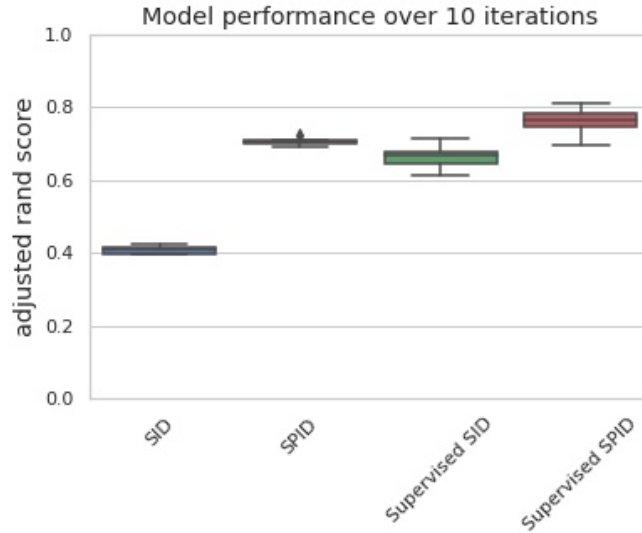


Figure 3.12: Comparing all four learners (distributional, pragmatic distributional, supervised distributional, supervised pragmatic) by adjusted rand score

3.3.3.1 The SID model

For a model to be successful it is not sufficient that it categorizes its input sentences into *any* three categories. The three categories should correspond to the sets of actual declaratives, actual interrogatives and actual imperatives. That is, for example, if in the classification learned by the model, the actual declaratives are split across two categories, the model has not learned to cluster by clause type. Turning to the profile of each cluster identified by the SID model, Figure 3.13 shows the proportion of declaratives, interrogatives, and imperatives in each identified cluster, and Figure 3.14 shows the proportion of sentences clustered together. In other words, Figure 3.13 shows whether each cluster mostly consists of one clause type, and Figure 3.14 shows whether each clause type is put in one cluster.

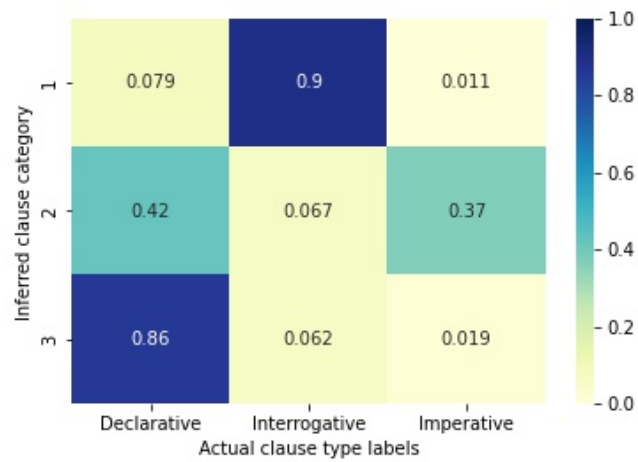


Figure 3.13: The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SID model

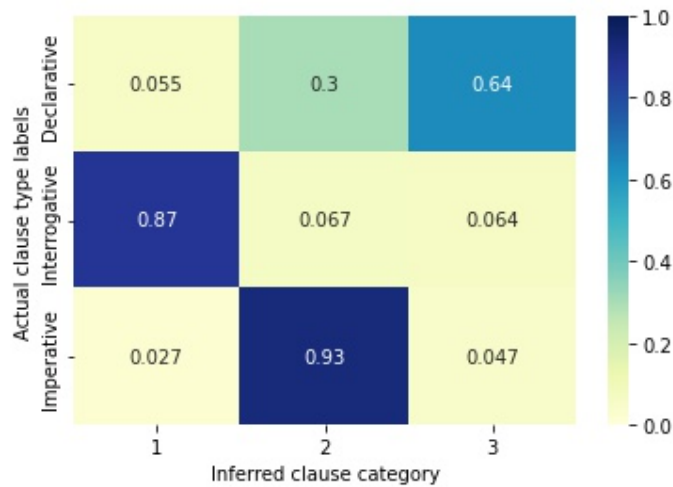


Figure 3.14: The proportion of actual declaratives, interrogatives, and imperatives clustered in one category

Overall, Figure 3.13 shows that the SID model identifies an interrogative cluster and a declarative cluster. Cluster 1 mostly contains interrogative clauses, as 90% of sentences in this cluster are interrogative; Cluster 3 is mostly declarative, and 86% of the data in this cluster are declaratives. Cluster 2 is split between declaratives and imperatives. Figure 3.14 shows that 87% of interrogatives and 93% of imperatives are clustered together in Cluster 1 and 2 respectively. While most of declaratives are classified in Cluster 3, a proportion is classified in Cluster 2.

These results seem to suggest that the SID model found one out of three clause types, but how do these clause types look like? Did the model find the right morpho-syntactic properties to associate with each clause type? Figure 3.15 plots the morpho-syntactic profile of each cluster identified by the SID. Darker colors represent sentences with the morpho-syntactic property, and lighter colors represent sentences without the property.

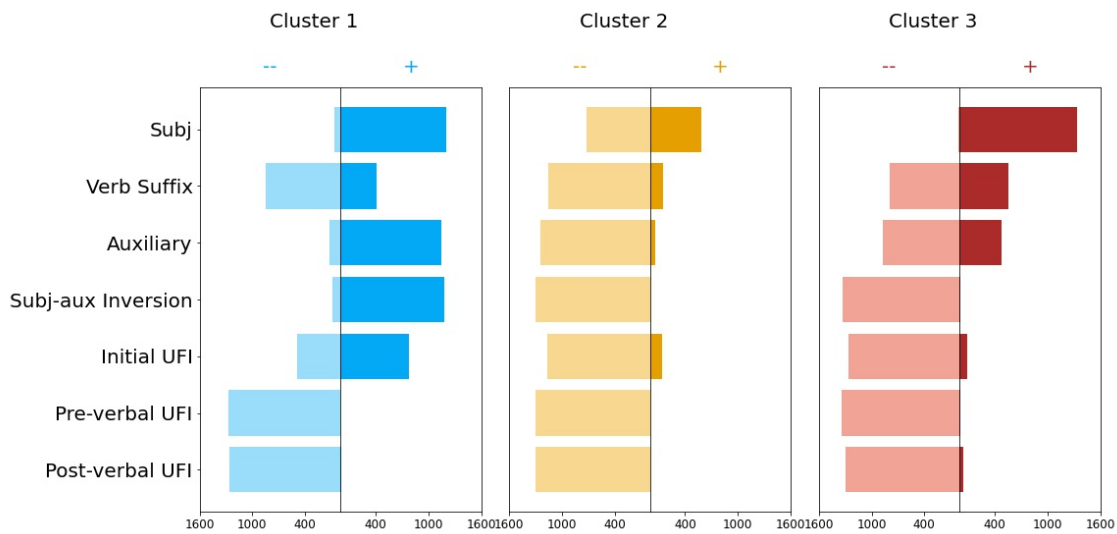


Figure 3.15: Number of sentences with/without certain formal features in each cluster (Cluster 1 ~ Interrogatives, Cluster 2 ~ Imperatives, Cluster 3 ~ Declaratives), darker colors represent the number of sentences with the feature. UFI stands for Unknown Functional Item (e.g. *wh*), see Table 3.2 for details.

We can see that Cluster 1, which is 90% interrogative clauses, is associated with [+int] morpho-syntactic properties such as subject-auxiliary inversion and sentence-initial unknown functional item (e.g. *wh*-phrases).

The other two clusters are not as ideal. Cluster 2 consists of a mix of imperative sentences and simple declarative sentences. While it appears that the cluster mostly consists of sentences with bare verb stem, which is characteristic for imperatives in English, but a quick look at the sentences show that it also includes declaratives with these properties as well:

(49) I love school. Mother of Lily, Session 010423

The sentence is clustered together with imperatives like *Take the bottle!* by the model. Ambiguous sentences like 50 are also put in this cluster:

(50) Wanna read your little Tigger book? Mother of Violet, Session 010407

Moreover, compare to the distribution of morpho-syntactic features of imperatives in English in Figure 3.9, this cluster seems to have more sentences with subjects. It seems that instead of finding a cluster for imperatives and declaratives, the model puts too much weight on the lack of subjects in a sentence, so any sentences without subjects are in Cluster 2.

Overall, the SID model fails to identify the clause type clustering in English, and fails to identify the characteristic morpho-syntactic properties for these clause types.

3.3.3.2 The SPID model

Figure 3.16 shows the proportion of declaratives, interrogatives, and imperatives in each identified cluster, and Figure 3.17 shows the proportion of sentences clustered together.

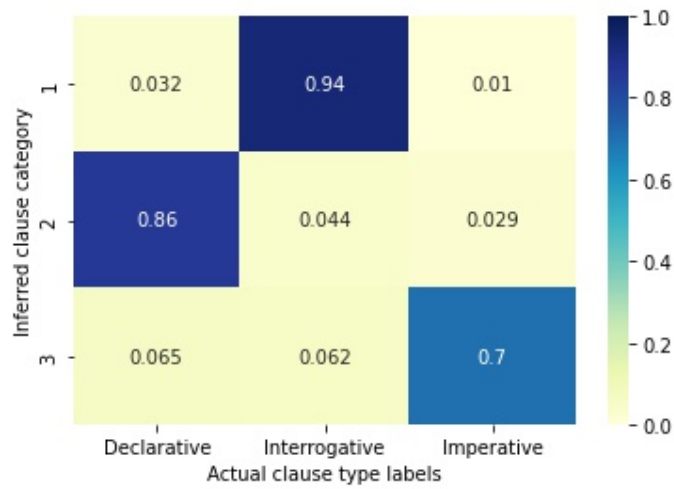


Figure 3.16: The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SPID model

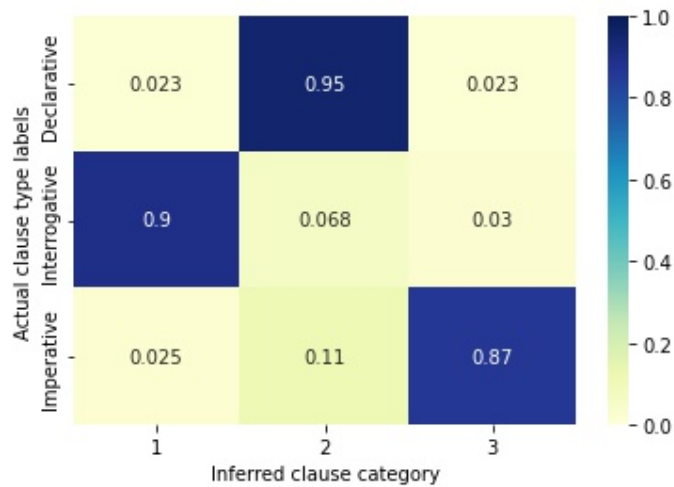


Figure 3.17: The proportion of actual declaratives, interrogatives, and imperatives clustered in one category

We can see that the SPID model clearly identifies a declarative, an interrogative, and an imperative cluster: 94% of Cluster 1 are interrogatives, 86% of Cluster 2 are declaratives, and 70% of Cluster 3 are imperatives. The three clause types in English are also mostly clustered together by the model: 95% of declaratives, 90% of interrogatives, and 87% of imperatives are clustered together in Cluster 2, 1, 3 respectively. In contrast to SID model, this model is able to find the right clause types in English.

Figure 3.18 shows the morpho-syntactic profile of each cluster identified by the SPID.

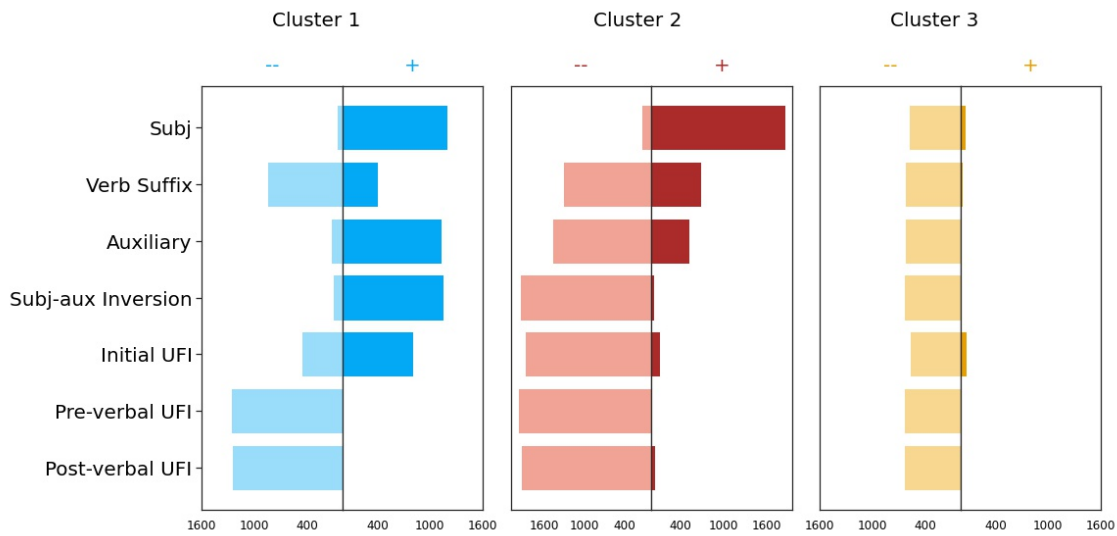


Figure 3.18: Number of sentences with/without a formal property in each cluster (Cluster 1 ~ Interrogatives, Cluster 2 ~ Declaratives, Cluster 3 ~ Imperatives), darker colors represent the number of sentences with the property, lighter colors represent sentences without the property.

The profile of the three clusters resembles the profile of declaratives, interrogatives, and imperatives in our corpus study (Figure 3.9). The cluster for interrogatives has more sentences with subject-auxiliary inversion and clause-initial unknown functional items (e.g. *wh*), and the cluster for imperatives is characterized by the lack of subjects and verb suffixes.

3.3.3.3 Interim Summary

Using data from infant-directed speech in English, our models simulate the syntactically informed distributional learner and the syntactically and pragmatically informed distributional learner. We found that the SPID model performs much better than the SID model; it successfully identifies three clause clusters roughly corresponding to the three major clause types in English. In contrast, the SID model fails to find the right clustering, collapsing declaratives and imperatives. These results suggest that pragmatic information is essential for children to solve the clustering problem for clause types. But how much pragmatics does the learner need? Recall from Chapter 1 that assuming *too much* pragmatics to be available to learners might catch us in a circularity. Considering the fact that 18-month-olds' pragmatic skills might still be developing, it is likely that they are not able to perfectly identify the speech act information of some utterances, and thus the pragmatic information available to the learner might be rather noisy. To probe deeper into this question, I now turn to simulated pragmatic learners with noisy speech act information. Studying such learners will allow us to investigate how much pragmatics is need to solve the clustering problem.

3.3.4 Simulating noisy pragmatic information

As we have seen in the last section, pragmatics information is crucial for clause type clustering. In this series of simulations, I manipulated the amount of noise in the speech act information that the SPID model takes as input. Instead of feeding the model true speech act labels, I replaced a percentage of the data with random speech act labels to simulate the situation where the learner randomly guess the speech act for a proportion of the utterances they hear. Each simulation was run 10 times.

Figure 3.19 shows the simulations with 0-100% noise. The performance of the SID model in our last simulation is indicated by the dotted line. As can be seen, the SPID learner outperforms the SID learner up until there is around 80% of noise in speech act, and even at 80% level there are iterations that outperforms the SID. This suggests that a little pragmatics

goes a long way, as the learners with noisy pragmatic information still outperforms the one without.

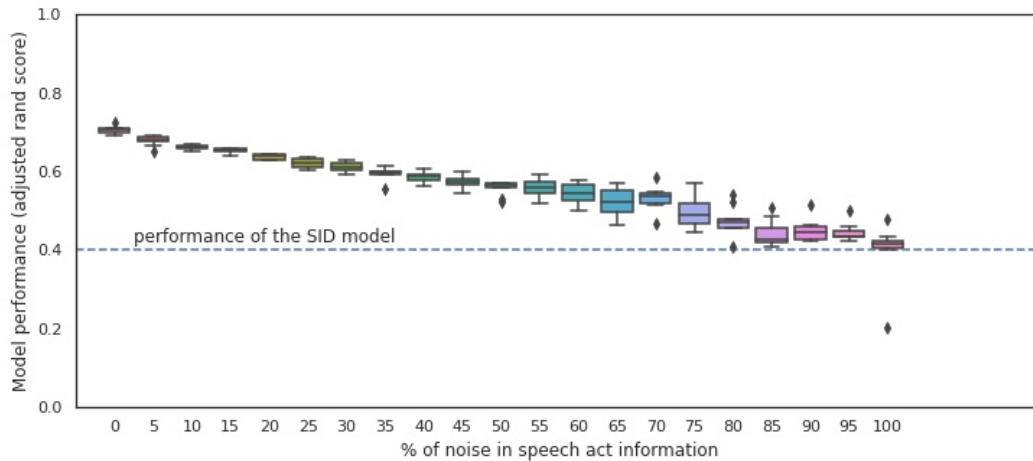


Figure 3.19: Performance of the SPID model with different levels of noise in the speech act information; dotted line indicates the adjusted rand score of the SID learner

Zooming in on the clusters identified by the models, I found that simulations with around 80% noise cannot successfully identify all three clusters correctly while simulations with 70% noise level can, even though the adjusted rand scores of these two levels are roughly the same. At 80%, we can see that the SPID reverts back to the performance of the SID in that it fails to identify a cluster for declaratives (Figure 3.20). Similar to the SID model, the cluster containing most imperatives also contains many declaratives (3.21).

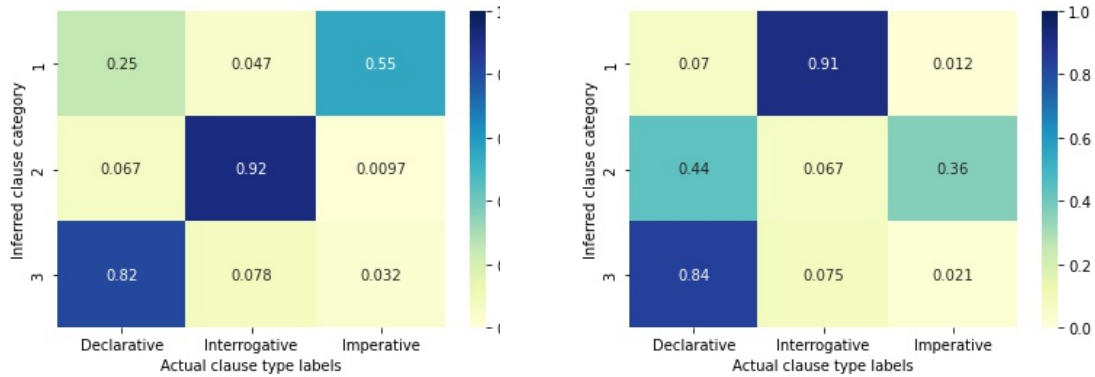


Figure 3.20: Proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SPID model; with 70% (left) and 80% (right) noise in speech act information

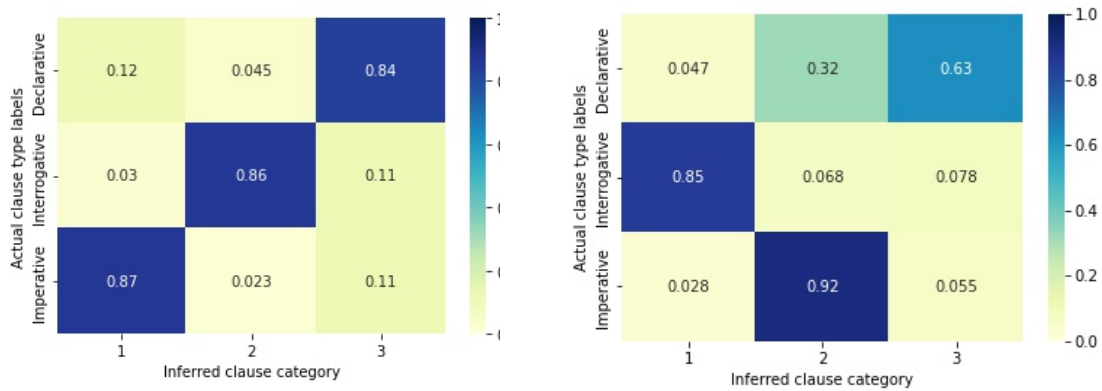


Figure 3.21: Proportion of actual declaratives, interrogatives, and imperatives clustered in one category, with 70% (left) 80% (right) noise in speech act information

The morpho-syntactic profile of different simulations further shows that at 80% noise level, the SPID and SID model behave similarly, as both fail to identify the property [–subject] for imperatives (Figure 3.22). In contrast, with 70% noise level the model can still find the right morpho-syntactic features for interrogatives and imperatives (Figure 3.23).

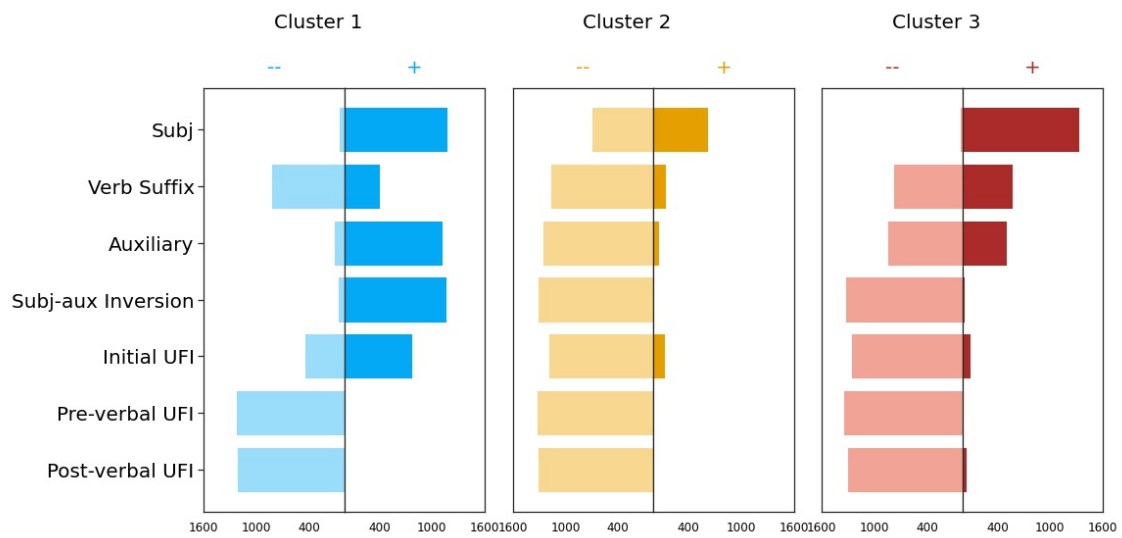


Figure 3.22: Morpho-syntactic profile of each cluster in simulations with 80% noise in speech act information (Cluster 1 ~ Imperatives, Cluster 2 ~ Interrogatives, Cluster 3 ~ Declaratives).

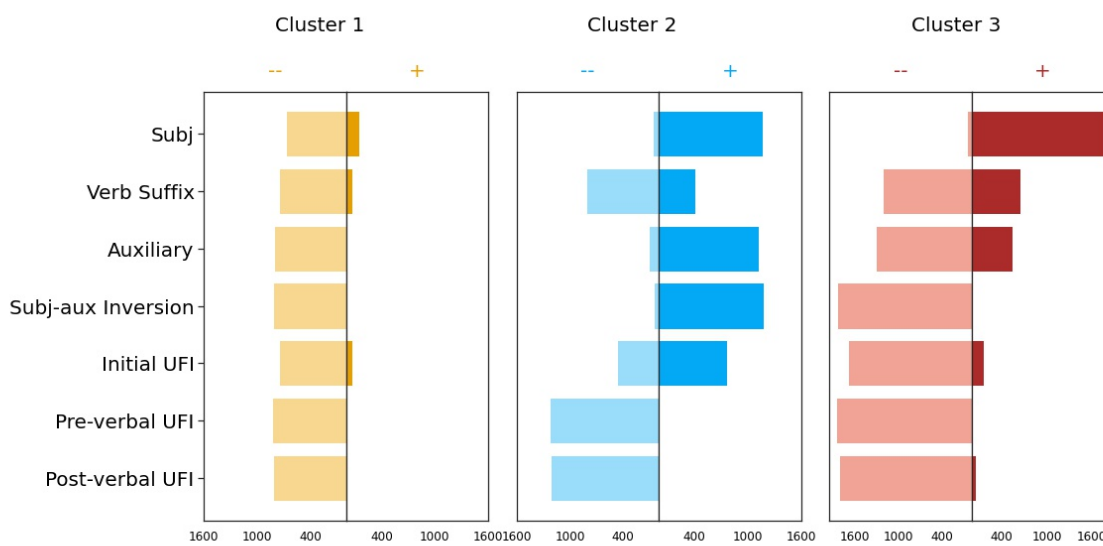


Figure 3.23: Mmorpho-syntactic profile of each cluster in simulations with 70% noise in speech act information (Cluster 1 ~ Interrogatives, Cluster 2 ~ Imperatives, Cluster 3 ~ Declaratives).

Results from our simulations suggest that morphosyntax is not sufficient to solve the clustering problem, and a small amount of pragmatic information is necessary. Even with 70% noise, pragmatic learner can still home in on the correct three clause types.

3.4 Discussion

This chapter investigates how 18-month-olds learn to identify the three matrix clause types in English. Our corpus study provides evidence suggesting that the information that children need to solve the clustering and labeling problems for acquiring clause types is present in their input. Parents use clause types systematically, and the three clause types are predominantly mapped to their canonical functions (declaratives to assertions, interrogatives to questions, imperatives to requests/commands) despite the presence of the theoretically expected exceptions to this mapping. The three clause types also show systematically different

formal profiles in the input: compared to declaratives, interrogatives are more likely to have auxiliaries, subject-auxiliary inversion, and clause-initial functional items like *wh*, which are characteristic of [+int] of C^0 in English; imperatives are more likely to not have subjects or verb suffixes, which is characteristic of [+imp].

I then investigated the extent to which learners need to rely on pragmatic information (i.e., knowing what speech act a given utterance of sentence is conveying), to solve not just labeling, but the clustering itself. I built two Bayesian clustering models simulating a syntactically informed distributional learner (SID) and a syntactically and pragmatically informed distributional learner (SPID). I found that morpho-syntactic information is not sufficient for finding the right clause type clustering and a small amount of pragmatic information is necessary, as the SPID learner outperforms the SID when we ran the two models with the annotated dataset from our corpus study. In addition, a small amount of pragmatics might suffice; even with 80% noise, speech act information still helps improve the performance of the model, and at 70% noise, the model still finds the right clause type categories. So even if the speech act information is not perceived veridically for all utterances, learners can still enjoy its benefits.

3.4.1 Implications

With all of the above results, what conclusion can/can't we draw regarding children's learning of clause types? Here I want to discuss the areas and the models are solving an easier problem than children, where the models are probably given a harder problem.

First, our models assume that learners have *a priori* knowledge regarding the number of clause type categories in their language. While this is a reasonable assumption given the near-universality of declaratives, interrogatives, and imperatives, there are several ways that the actual task that learners have to solve could be more difficult. For one, children might have to figure out the number of clause types in addition to which sentence belongs to which clause type. This additional inference about number could make the problem harder, as

learners could end up with more or less than 3 main clause types, that do not correspond to those of the adult grammar. On the other hand, allowing for an underdetermined number of clause types could help learners pick up more easily on minor clause types, like exclamation marks, which are not shared across languages as universally. However, children could still start with the assumption that there are three clause types, and gradually expand to include minor clause type categories like exclamation marks.

Moreover, I assumed that the learners at this age can perceive a set of morpho-syntactic features that are relevant for clause typing. But this set of features were hand-picked as the most relevant ones for clause typing to give the SID model the best chance to succeed; children may not be so lucky to have this set picked for them and might need to figure out what the relevant features for clause typing are in addition to figuring out which features go with which clause type cluster. This will no doubt make the learning problem harder. One justification for making this assumption though is that (besides giving the SID the best shot) children might have some knowledge that certain features are clausal features, and some are not. For example, they might know that verb objects might be relevant for learning properties about verbs, but not necessarily relevant for learning the properties of the whole clause. This knowledge might help them pick the relevant morpho-syntactic features for figuring out clause type clusters.

Another area where we could improve the model is how speech act information is accessed by the learner. In our simulations with different noise levels, we replaced a proportion of true speech act labels with random speech act labels. In other words, the models assumed that every utterance you hear is associated with a speech act category; this models a learner, who, in the face of uncertainty would always make a guess. But it is likely that in cases of uncertainty, instead of make a guess, and the learner simply tosses out the utterance, so they might not learn from this datapoint. In fact, we adults do this too: sometimes when we can't decide whether our interlocutor is asking a question or making a statement, instead of making a guess right away, we might decide to wait for more data. Therefore, my

next step is to adjust the model to incorporate an internal signal of confidence about speech act perception. This signal could then be used as a filter to help learners only learn from utterances that they are confident about.

Uncertainty may not be the only source of noise: learners could very well make systematically wrong assumptions about speech acts, particularly in cases of indirect speech acts where the force of the primary act differs from sentential force. In the future, I will investigate whether such a pragmatic model could still find the right clause type categories.

One area that the models are given a disadvantage is the absence of prosodic features. As mentioned in Chapter 2, a final rising contour could be particularly helpful to identify polar interrogatives. We will come back to how parents use prosody, and if taking into account prosodic features would improve SID's performance in Chapter 5.

Now that we have established the importance of pragmatics, and that a small amount of pragmatics might suffice, the question then is, how children may get this pragmatic information. How can children infer speech act independently of clause types? As we have discussed in Chapter 1, with the innate knowledge of a set of speech act categories and what speech acts do, and a theory of communication, children might have certain expectations about how each speech acts are performed in the input. If these expectations are met, then they can use these cues to infer which utterance is performed with what speech act. But just as our discussion with clause types, behaviors related to speech acts might be absent or misleading. We therefore need to verify empirically whether some of these expectations about speech acts can be met. Specifically, do certain prosodic contours and non-linguistic behavior of parents correlate with the use of questions? Before turning to these questions in Chapter 6, I will first explore the role of morpho-syntactic information in solving the clustering problem in another language, Mandarin.

Chapter 4: Learning to identify clause types in Mandarin

In this chapter, we will explore how Mandarin-acquiring children learn clause typing. As we have seen in the last chapter, pragmatic information is crucial for finding the right clause type clustering in English. We found that a learner must have access to some pragmatic information in order to find the right clause types but this learner can succeed with very limited access to pragmatic information. Can we show the same for learners of a different language, like Mandarin?

Mandarin offers an interesting testing case not only because it has a different set of surface features for [+int] and [imp]. Due to its impoverished morphology, the problem with absent and misleading surface features might be even more pronounced. Is it possible for a learner to infer the surface forms associated with these features from the input? Is pragmatics also crucial for Mandarin-acquiring children? And if so, how much pragmatics is required? In this chapter, I address these questions computationally following the same methodology as in Chapter 3. I find that, for Mandarin too, surface morpho-syntactic features alone are not sufficient for learners to cluster sentences into the three clause types, and pragmatics is needed, even more so than for English.

This chapter is organized as follows: we will review the properties of Mandarin [+int] and [+imp], and what we know of children's knowledge regarding these properties. I then report results from a corpus study for Mandarin-speaking parents' input to children in 4.2. The data from the corpus study was used to simulate the two learners, syntactically informed distributional learner and syntactically and pragmatically informed distributional learner.

4.1 Background

4.1.1 Clause types in Mandarin

In Mandarin, the [+int] value in C^0 does not trigger movements, but instead shows up in surface form as *wh*-phrases, sentence final *ma*, and the A-not-A form. Let's go through one by one.

The presence of a *wh*-phrase could signal [+int]. Mandarin *wh*-interrogatives are famously known to have *wh*-phrases stay in situ, and do not need to be fronted to clause-initial position (Huang 1982, Cheng 1991 among many others). Compare the declarative sentence in (51) with the *wh*-interrogative in (52): the *wh*-phrase *shenme* occurs in the same position in the interrogative (52) as the noun phrase *zaocan* in (51).

- (51) Xiaoxiao chi-le zaocan.
Xiaoxiao eat-ASP breakfast
“Xiaoxiao ate breakfast.”

- (52) Xiaoxiao chi-le shenme.
Xiaoxiao eat-ASP what
“What did Xiaoxiao eat?”

Polar interrogatives in Mandarin also have SVO word order, but are distinguished from declaratives by surface features like the question-forming particle *ma* or the A-not-A construction, as in (53) and (54).

- (53) Xiaoxiao chi-le zaocan ma
Xiaoxiao eat-ASP breakfast SFP
“Did Xiaoxiao eat breakfast?”

- (54) Xiaoxiao chi-meishi zaocan?
Xiaoxiao eat-NEG-eat breakfast
“Did Xiaoxiao eat breakfast?”

Sentence final particles (SFPs) like *ma* are particles at the right edge of a clause (Chao 1968, Zhu 1982, Huang 1982, Cheng 1991, Li 2006 among others). Many SFPs such as *ba*,

ya can occur in any clause type, but *ma* can only occur in polar interrogatives that do not have the A-not-A form seen in (54).

To summarize the discussion so far, the surface form of [+int] in Mandarin is sentence final particle *ma*, A-not-A constructions, and *wh* with prominence.

In theoretical discussions, it is extremely difficult to pin down what are imperatives in Mandarin. One diagnostic is the negative modal *bie* (and its etymologically related modal *beng*), which cannot occur with other values of C^0 (Chao 1968, Li & Thompson 1981, Chen-Main 2005a, C.-H. Han 1998 among many others):

- (55) a. **Bie** pao!
 Don't run
 "Don't run!"
- b. *Zhangsan **bie** pao.
 Zhangsan don't run
 (intended) Zhangsan doesn't run.

Using *bie* as a diagnostic for imperatives, we can see that Mandarin [imp] can be embedded (Li & Thompson 1981, Chen-Main 2005b), and certain verbs like *zhuzhang* selects [imp]:

- (56) Wo zhuzhang Lisi **bie** chu-guo.
 I advocate Lisi don't exit-country
 "I have the opinion that you don't leave the country." Li & Thompson (1981:p.458)

While some have argued that certain SFPs like *ba* are imperative particles (Zhu 1982, Chao 1968, Li & Thompson 1981, Huang, Li, & Li 2009), they can actually appear in declaratives, imperatives, and interrogatives, and do not necessarily convey a request/command force (Y. Han 1995, Li 2006, Ettinger & Malamud 2014, Yu'an Yang 2021):

- (57) Xiayu le **ba**?
 rain ASP SFP
 "It's raining?"

Similarly, when these particles append to a declarative clause, they do not change the clause type to [+int]:

- (58) Xiang zuo a?
 want sit SFP
 “Want to sit on it?”

Declaratives as question

Mother of Tong, Session 01;08;22

The particle *a* elicits a discourse effect along the lines of “we both know *p* (or the answers to the question *q*), but let’s put it on the table.” I do not consider this particle to be changing the clause type to be interrogative, just like the final rise in English does not change the clause type of declaratives. This is notwithstanding the fact that the discourse effect of an utterance with *a* is similar to that of a question, namely the speaker wants the addressee to respond. This too is akin to the English final rise. I therefore do not assume that sentence final particles like *ba* and *a* that can occur across clause types are related to [imp] or [+int].

But this left us with only one other possible surface feature for Mandarin might be the lack of subjects. But as we have briefly mentioned, this feature might create many problems for the learner. First, as it is possible to have second person pronoun as a subject with [imp]. In (60), the sentence has the imperative marker *bie*, but also has subject *ni*:

- (59) Ni **bie** pao!
 You don’t run
 “Don’t run!”

Moreover, as Mandarin is a pro-drop language, other types of sentences could appear without subjects.

Another surface feature for English [+imp] is the absence of verbal suffixes, but Mandarin verbs do not have suffixes. Mandarin does have aspect morphemes *le*, *zhe*, *guo*, but they can appear with [+imp]:

- (60) Bie yao-zhe zuichun!
 Don’t bite-ASP lips
 “Don’t bite your lips!”

Mother of Tong, Session 02;00;19

Thus, [+int] in Mandarin may show up in the surface form of the sentence as the presence of *wh* with prominence, sentence final *ma*, or A-not-A structure; [imp] shows as sentences

without subjects or with second person subjects, and as special negative modal *bie* or *beng* in negative imperatives.

But prosody could be important for [+int]. The *wh*-phrases in Mandarin can be interrogative or indefinite (61), but interrogative *wh* is normally associated with prosodic prominence (Hu 2002, Dong 2009, Yang Yang 2018 a.o.).

- (61) Xiaoxiao mei chi shenme
 Xiaoxiao NEG eat what
 a. What didn't Xiaoxiao eat?
 b. Xiaoxiao didn't eat anything.

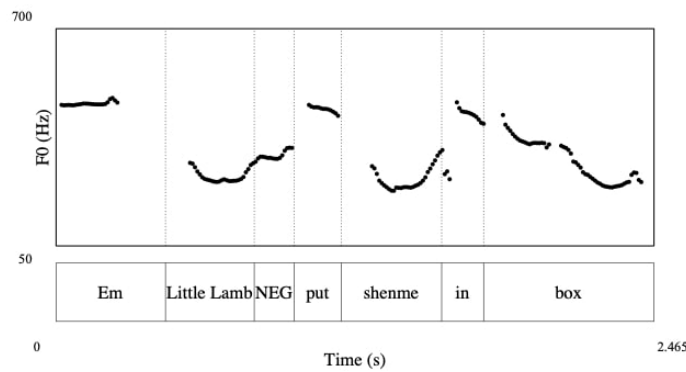


Figure 4.1: Pitch contour associated with a *wh*-interrogative sentence

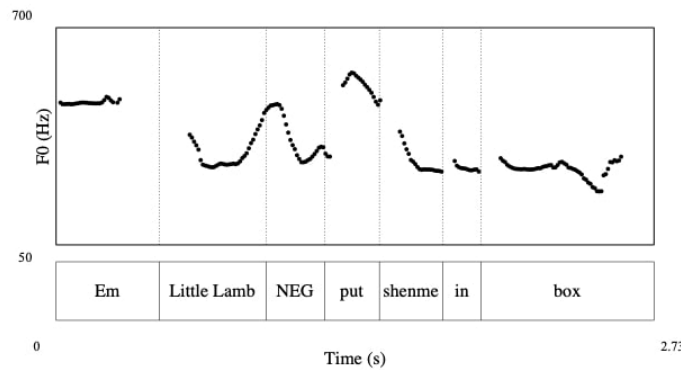


Figure 4.2: Pitch contour associated with a *wh*-indefinite sentence

Figure 4.1 shows the pitch contour of the interrogative interpretation of *wh* and Figure 4.2 shows the indefinite interpretation. The former prosody is associated with [+int] and the latter with [-int], and the crucial difference is that the interrogative *shenme* is produced with prosodic prominence (longer duration, extended lexical tone).

Prominence on the *wh*-phrases is not the only feature. Experimental studies controlled for lexical tone find that compared to declaratives, the F0 of the entire utterance is raised for interrogatives (Ho 1977, Shen 1991, Yuan, Shih, & Kochanski 2002, Ni & Kawai 2004, O. J. Lee 2005, Liu & Xu 2005, Yuan 2006, 2011, Fang 2009, Jiang & Chen 2011 among many others). In addition, the nuclear contour of *ma*-interrogatives and rising declaratives tend to have a final rise contour (Shen 1991, Zeng, Martin, & Boulakia 2004), and A-not-A and *wh*-interrogatives tend to end on falling intonation (Shen 1991, Ni & Kawai 2004, cf. Liu & Xu 2005, Yang, Gryllia, & Cheng 2020). Perception experiments show that speakers tend to use the overall F0 of the whole sentence to distinguish interrogatives from declaratives (Jiang & Chen 2011, Gryllia et al. 2020), and it is more difficult to use clause-final prosodic cues, especially if the final syllable has a rising tone (Yuan 2011). However, most of these studies involve sentences where the lexical tones are strictly controlled. It is therefore unclear how speakers make use of intonation in real speech, where the lexical tones will interfere with sentential prosody. This question is beyond the scope of this dissertation, so here we focus on morpho-syntactic features.

In sum, Mandarin learners face the same learning problem as English learners, but even more dire: the clause type feature is abstract and is related to a variety of surface forms, none of which is obligatorily present and many of which can occur in sentences with a different clause-type feature on C^0 . Since the surface features indicative of clause type differ between English and Mandarin, however, the problem manifests itself in its own way in each language. For example, even though [imp] is related to null subjects, but since Mandarin is a pro-drop language, it is far more likely that the learner observes a sentence without subject and yet the feature is not [imp]. Interrogatives in Mandarin tend to be associated with raised overall

F0 and in some cases raised boundary tone. However, these results about prosodic features are mostly from experimental studies where the lexical tone is controlled. It is difficult to discern the intonation pattern with real speech data. In this chapter, I therefore focus on the morpho-syntactic features.

4.1.2 Mandarin-acquiring children’s knowledge of clause types

In this section we will look at what we know about Mandarin-acquiring children’s knowledge regarding the features reviewed in the last section. For many of these features, we do not have a lot evidence from the comprehension side, and we have to rely on data from children’s production alone to make inferences about their knowledge. As production might not be an accurate reflection of children’s grammar, we will be conservative in our assessment.

±SUBJECTS As Mandarin does not have subject-verb agreement, and does not have expletive subjects, it is hard to assess children’s knowledge regarding subjecthood. But observations of children’s production data before 18 months old suggest that they understand the basic word order of Mandarin as SVO (Tardif 1993, 1996). They also produce subject control sentences (e.g. *Zhangsan tried PRO to help me*) correctly before turning 2 (Yang & Yang 2015). Although this falls out of our interested age range, studies have demonstrated that 2.6-year-olds can correctly distinguish subjects from topics (Chien & Lust 1985).

±VERBS By 18 months old, Mandarin-acquiring infants produce many verbs, and use verbs productively in many contexts (Tardif 1993, 1996, Xiao, Cai, & Lee 2006, Zhang, Shi, & Li 2014, among many others); while this does not necessarily mean that they have a verb category, it is consistent with the hypothesis that they do. 13-month-olds also demonstrate the ability to categorize novel words as verbs using frequent frames related to verbs, such as auxiliaries (e.g. *bie* “don’t”, Zhang, Shi, & Li 2015) and focus particles (e.g. *ye* “also”, Zhang, Shi, & Li 2014; Ying, Yang, & Shi 2021 for 19-month-olds).

±AUXILIARY Zhang, Shi, & Li 2015 find that 12-month-olds use functional words to categorize content words, specifically that they could use the negative imperative modal *bie*

to identify the follow-up item is a verb, suggesting that children might be sensitive to the presence of this negative modal auxiliary.

$\pm wh$ -PHRASES Children start producing interrogative *wh* as early as 14 months old (T. H.-T. Lee 1989, Fan 2012, Lin, Weerman, Zeijlstra, et al. 2014); both comprehension experiments and corpus data show that they are also found to answer *where*, *who*, *what* questions appropriately at 18 months old (Fan 2012, Moradlou et al. 2020). While we have evidence that 3-year-olds can use prosodic prominence to infer whether a *wh*-sentence is [+int] (Yang et al. 2022), we do not have evidence about younger children. In the corpus study described in the next section, I follow the practice from the last chapter and classify *wh*-phrases with focus particles and connectives (e.g. *yaobu* “if”) as ‘Unknown Functional Items’ (UFI).

A-NOT-A STRUCTURE As noted earlier, the A-not-A structure and the presence of sentence-final *ma* both distinguish interrogatives from declaratives in Mandarin. But while there is evidence suggesting that children start producing negation around 1.5 years old (T. H.-T. Lee 1982, Fan 2007, Jingyao 2019, R. L. Huang et al. 2022 among others), we do not have evidence regarding whether children perceive A-not-A sentences differently from simple negation sentences. Similarly, we do not have evidence regarding whether children treat *ma* differently from sentences with other sentence final particles like *ya*. I therefore simulated a conservative learner who does not have access to A-not-A and *ma* features, and compare it with a knowledgeable learner who does have access to these two features.

4.2 Corpus study

4.2.1 Corpus and methods

For this corpus study, we used data from the Tong subcopora (Deng & Yip 2018) from CHILDES (MacWhinney 2000), which contains audio and video recordings of weekly hour-long free play sessions between Tong and his caregivers from 1;7-3;4 in Shenzhen, China

where Mandarin is the language of the community. Although this corpus only contains data from one child, it is the source of Mandarin data that is most comparable to the Providence Corpus in terms of the child’s age range and availability of audio/video data. If more corpora from Mandarin-speaking children become available in the future, the methodology developed here can be applied to them. Another problem with the corpus is that it does not have data from when before the child was 18 months old, which is older than the age that we assume children figure out the clause type categories; 18 months is also older than the children in the English corpus study. However, while the pragmatics of parent-child interaction might change with children’s age, the morpho-syntactic properties of parents’ sentences should stay constant. Therefore, we assumed that the input sentences share similar properties as parents’ sentences before Tong turns 18 months old. Once corpus data from younger children becomes available, we would apply the same methodology to these data.

We sampled 500 conversational turns from each session from when the child was 01;07;18 to 2;2;16. Each session was coded by two annotators independently for the Clause Type and Speech Act information (mean cohen’s κ : 0.8). For the morpho-syntactic features, initial annotation was generated by a script using the morphological tagging provided by the corpus, and then manually corrected. In total, 4501 utterances were annotated.

4.2.1.1 Clause Type

As in the English corpus study reported in the last chapter, each sentence was annotated with their clause type category, declaratives, interrogatives, and imperatives (62a-62c). Sentences with only one noun phrase or injectives were annotated as “fragments”:¹

- (62) a. Wazi shi le.
 Sock wet SFP

¹In contrast to English, sentences without verbs might not be fragments in Mandarin, as the copula *shi* “be” is optional:

1. Zhe wode.
 This mine.
 ‘This is mine.’

	“Your socks got wet.”	Declarative
b.	Kandao le ma? See ASP SFP “Do you see it?”	Interrogative
c.	Gei wo hongse de na-ge. Give me red POSS that-CL “Give me the red one.”	Imperative
d.	Ai-ya! INTJ “Wow!”	Fragments

The three major speech acts were also annotated, same as in the English corpus study.

4.2.1.2 Morpho-syntactic features for clause typing

As reviewed in the last section, Mandarin-acquiring children at 18 months old can perceive many morpho-syntactic properties related to clause typing. In particular, they might be able to identify the subject, verb, auxiliary of the sentence, and distinguish functional from lexical items. We additionally adopted the conservative assumption that children at this age might not be able to identify all the *wh*-items at this stage, but still assume that they might be able to classify them as functional elements, as they may know the distinction between functional and content elements. We therefore put *wh*-items, quantifiers, connectives (e.g. *haishi*, the interrogative “or”), and focus particles in one category “unknown functional item (UFI),” and annotated their position in a sentence: sentence initial, sentence-medial but before the verb, after the verb, or sentence final. In addition to such surface features that were also annotated in the English corpus study, we also annotated for whether the sentence contains an A-not-A structure, and whether there is a sentence-final *ma* particle. But since we do not have evidence for whether children around 18 months can perceive these two cues, as discussed in the last section, when simulating children’s learning process, we will compare conservative models without these two cues, and knowledgeable models with these two cues.

Each sentence was annotated with whether or not a surface feature is present. (4.2.1.2)-(68) demonstrate the surface features we annotated and their examples.

- (63) **kan** zhe-ge.
 Look this-CL
 ‘Look at this one!’ +Verb
- (64) **Wo** zhidao.
 I know
 ‘I know.’ +Subject
- (65) a. Xiaopengyou bu-**neng** peng.
 Children NEG-can touch
 ‘Children can’t touch (this).’ +Auxiliary
- (66) + Unknown functional items:
- a. **Yaobu** na zhe-ge kapiān lái jiāo ba
 What-if take this-CL card to teach SFP
 ‘What if (you) teach with this card.’ Sentence-initial UFI
- b. Zheli **hai** you labi.
 Here also have crayon
 ‘There’s also crayons.’ Pre-verbal UFI
- c. Limian you **shenme** ya?
 Inside have what SFP
 ‘What’s inside?’ Post-verbal UFI
- d. Guolai **ba**
 Come SFP
 ‘Come here!’ Sentence final particle
- (67) + A-not-A:
- a. Jintian **leng-bu-leng** a?
 Today cold-NEG-cold SFP
 ‘Is it cold today?’ *Adj-not-Adj*
- b. Ni **hui-bu-hui** xiēzi?
 You can-NEG-can write
 ‘Can you write?’ *Aux-not-Aux*

c. Ni **you-mei-you** xiaoqiche?
 You have-NEG-have car
 “Do you have cars?”

V-not-V

d. **Chuan-hao** yifu **meiyou**?
 Put-well cloth NEG
 “Did you put on your coat?”

A-not

(68) Xiayu le **ma**?
 Rain ASP MA
 “Is it raining?”

+ma

4.2.2 Results

4.2.2.1 Overview

In total, 3077 utterances were annotated. Figure 4.3 shows the distribution of clause types in the dataset. As in the English dataset, declarative clauses are the most frequent clause type, followed by interrogatives and imperatives.

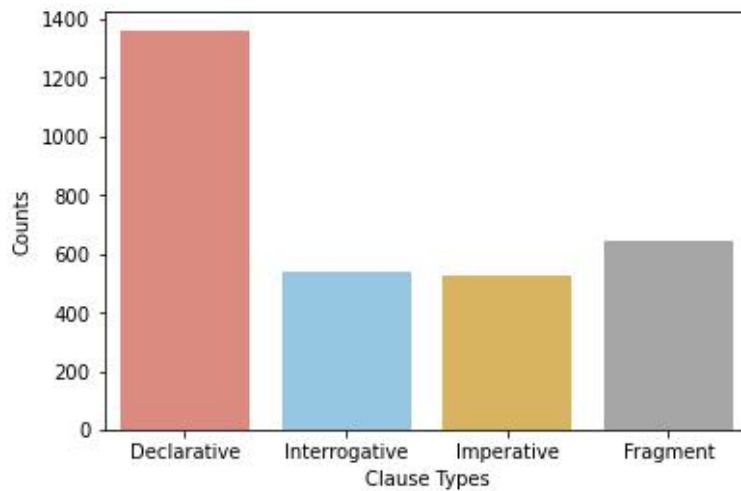


Figure 4.3: Distribution of clause types in the corpus

Within interrogatives (Figure 4.4), *wh*-interrogatives (69a) are more frequent than other types of interrogatives, a pattern similar to the one observed in English; followed by A-not-A (69b) and *ma*-interrogatives (69c). Disjunctive interrogatives with *haishi* “or” (69d) are relatively rare.

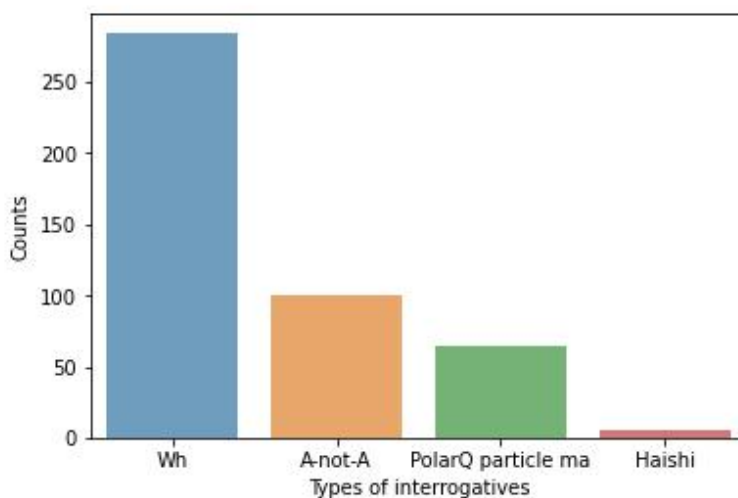


Figure 4.4: Subcategories of interrogatives

- (69) a. Zhe **shenme** a?
 this what SFP
 “What is this?” *wh*-interrogative
 Mother of Tong, Session 01;10;17
- b. Na **xi-bu-xihuan** he naifen a?
 Then like-NEG-like drink baby formula SFP
 “Do you like baby formula then?” A-not-A interrogative
 Mother of Tong, Session 02;00;19
- c. Ni Zhidao **ma**?
 You know MA
 “Do you know?” Ma-interrogative
 Mother of Tong, Session 01;10;17

- d. Qiezi haochi **haishi** baicai haochi?
 Eggplant tasty or cabbage tasty
 “Is eggplant tastier or cabbage tastier?”

Haishi interrogative

Father of Tong, Session 01;08;22

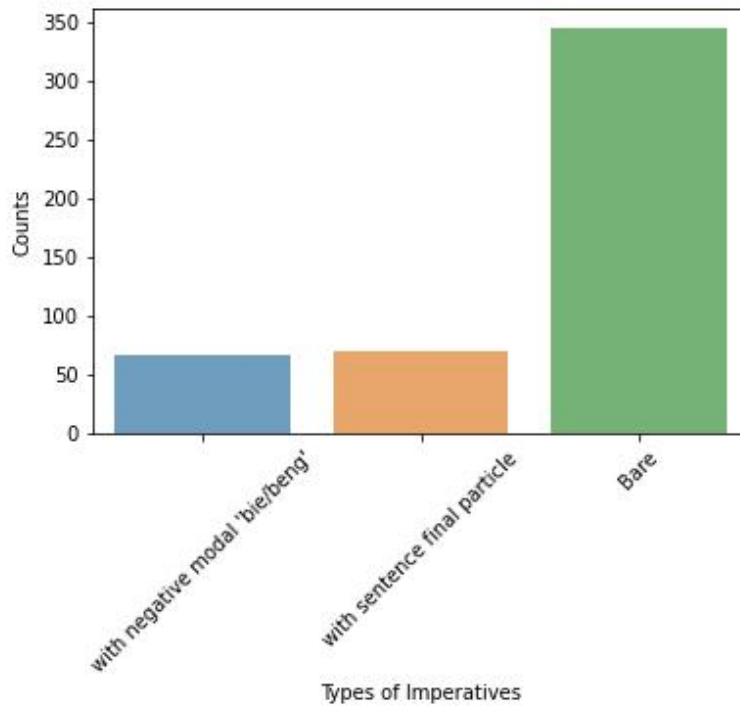


Figure 4.5: Subcategories of imperatives

Imperative sentences (Figure 4.5) mostly come without any marker like (70a); *bie*-imperatives (70b) and SFP-imperatives (70c) are equally frequent.

- (70) a. Kan zhege shi shenme dongxi
 look this is what thing
 “See what this is!”

Imperative (used as a question)

Mother of Tong, Session 02;00;19

- b. **Bie** fan le!
 Don't turn ASP

“Stop messing around!”

Bie-imperative

Mother of Tong, Session 02;00;09

- c. Reng zheli **ba**
throw here SFP

“Throw it here!”

SFP imperative

Mother of Tong, Session 02;01;17

As for speech acts, Figure 4.6 shows the distribution of speech acts. Same as in English, assertions are the most frequent speech acts, followed by questions and requests.

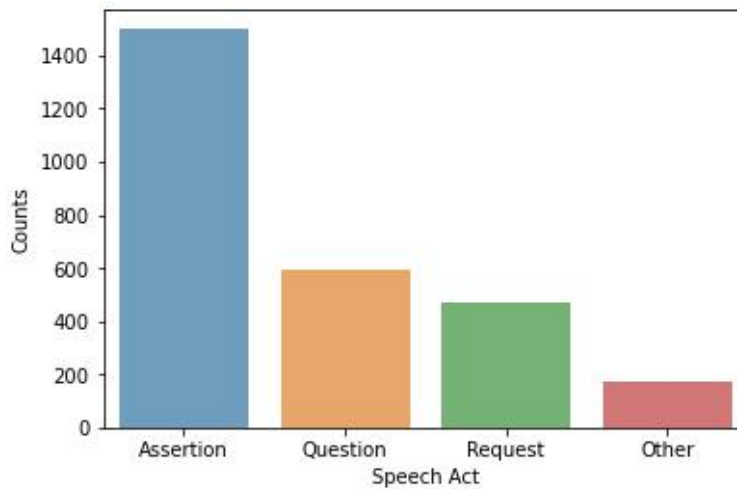


Figure 4.6: Distribution of speech acts in the corpus

The mapping between speech acts and clause types also shows a similar pattern as in English, as illustrated by Figure 4.7 and 4.8. Declaratives are mostly used as assertions, interrogatives as questions, and imperatives as requests.

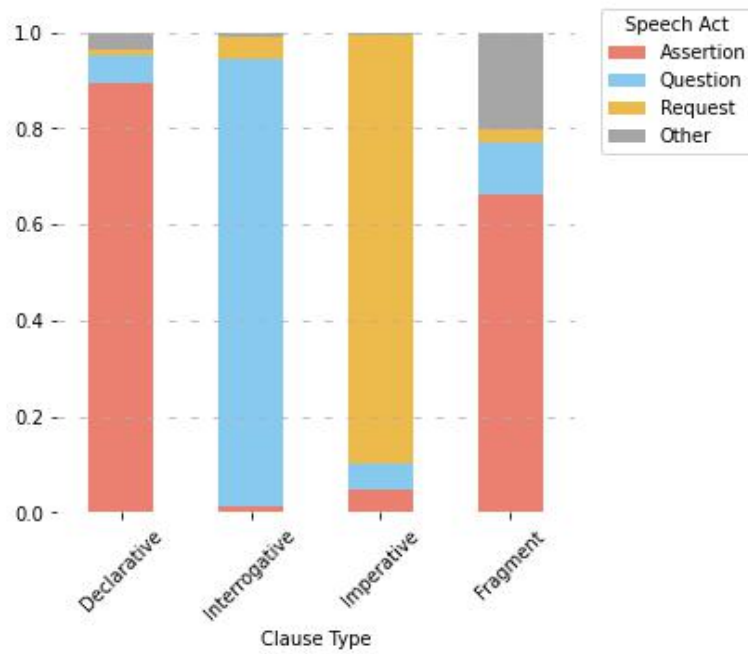


Figure 4.7: The speech acts performed by each clause type in parents' speech

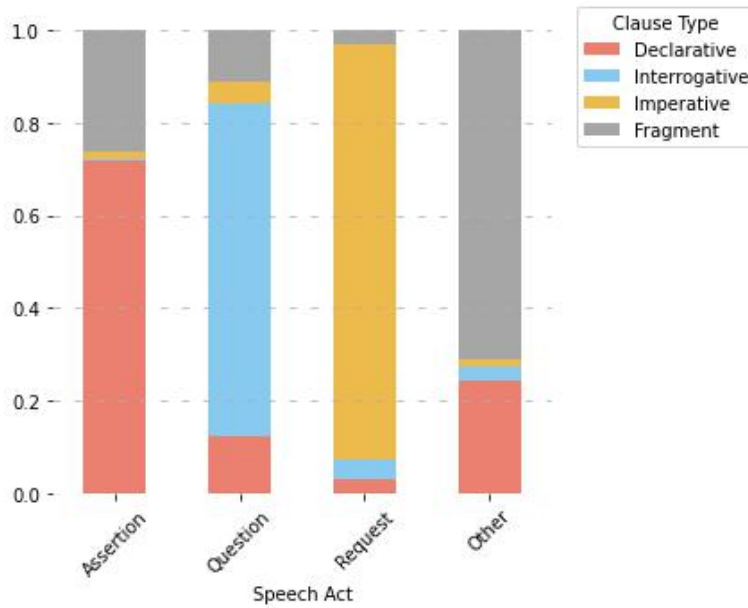


Figure 4.8: The clause type used to express each speech act in parents' speech

The proportion of mismatches is small and they form a systematic subcategory. All interrogatives that are used as assertions are rhetorical questions (71). The proportion of interrogatives as requests is smaller than what we observed in the English study (72). Even in the small portion of cases we found here, it could be debated whether the non-literal speech act of this sentence is request, as the mother is still asking about the child’s ability whereas indirect requests in English like *can you pass the salt* are much more saliently to be interpreted as requests instead of inquiries about the addressee’s ability.

(71) Ni xia gaoxing ge shenme ya
 you blindly happy CL what SFP
 “What are you happy about.” Interrogative as question

Mother of Tong, Session 02;00;19

(72) Tongtong hui bei ma?
 Tongtong can recite MA
 “Can you recite (this poem), Tongtong?” Interrogative as Requests

Mother of Tong, Session 02;00;19

When imperatives are not used as requests, it is almost exclusively with *kan* “look” and *gaozu* “tell”:

(73) Gaosu mama zhe shi shenme ya?
 Tell mom this is what SFP
 “Tell mom what this is” Imperative as question

Mother of Tong, Session 02;00;19

Declaratives as questions often occur with sentence final particles like *a*:

(74) Xiang zuo a?
 want sit SFP
 “Want to sit on it?” Declaratives as question

Mother of Tong, Session 01;08;22

As we have discussed, the final particle *a* does not change the clause category of the sentence, but still elicits a discourse effect similar to a question’s effect, namely that the speaker wants the addressee to settle an issue.

4.2.2.2 Morpho-syntactic cues

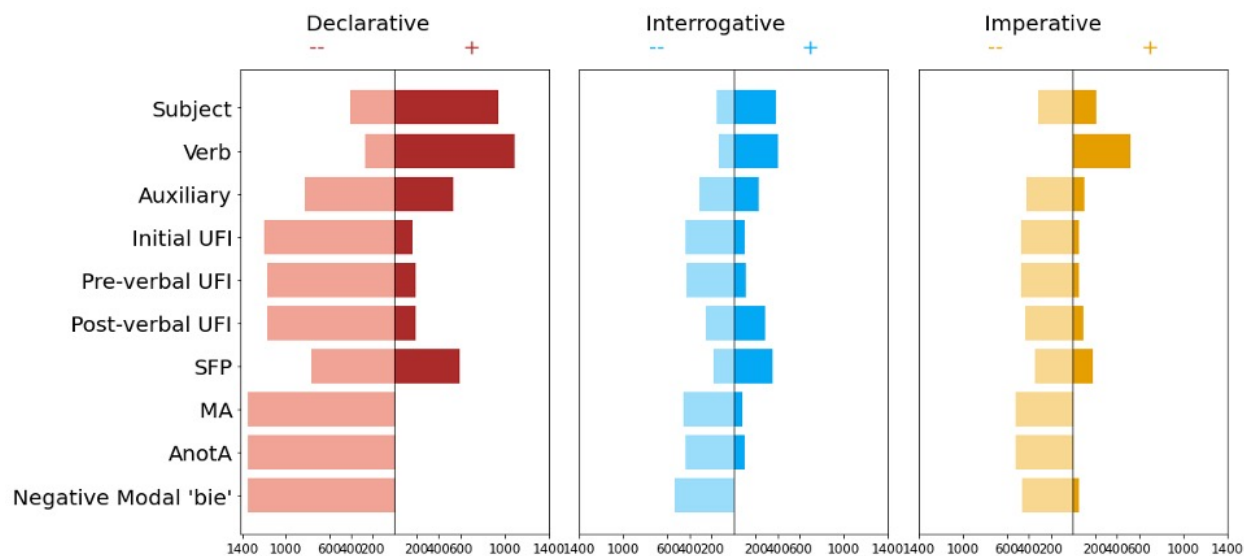


Figure 4.9: Number of sentences with/without various formal cues in each clause type; darker colors represent number of sentences with the cue, lighter colors, number of sentences without the cue

4.3 Modeling the learning of clause type in Mandarin

In this section, I applied the same Bayesian cluster models as reported in the last chapter to the Mandarin data. As mentioned earlier, we do not have evidence for whether Mandarin-speaking children identify A-not-A constructions or sentence final *ma*. Therefore, I ran two separate simulations for each model (four simulations in total): a conservative learner that do not have access to whether A-not-A or *ma* are present, and a knowledgeable learner that do have access. (75) lists the morpho-syntactic cues that each learner have access to.

(75) For the conservative learner:

\pm subject, \pm verb, \pm auxiliary, \pm sentence-initial UFI, \pm pre-verbal UFI, \pm post-verbal UFI, \pm sentence final particle

(76) For the knowledgeable learner: \pm subject, \pm verb, \pm auxiliary, \pm sentence-initial UFI, \pm pre-verbal UFI, \pm post-verbal UFI, \pm sentence final particle, \pm A-not-A, \pm *ma*

We used the same methodology and learning models as in the English study in Chapter 3. The results are as follows.

4.3.1 Performance of the SID model

4.3.1.1 The conservative simulation

The conservative simulation of the SID model does not use interrogative-specific features $\pm A$ -not-A construction and $\pm ma$. From the results, it does not seem like the model identified any meaningful grammatical class in Mandarin.

Figure 4.10 shows the proportion of declaratives, interrogatives, and imperatives in each identified cluster, and Figure 4.11 shows the proportion of sentences clustered together.

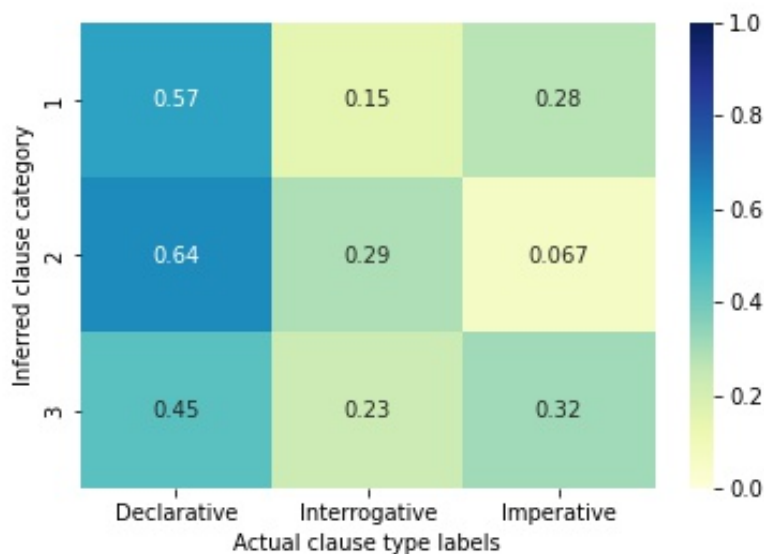


Figure 4.10: The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SID model

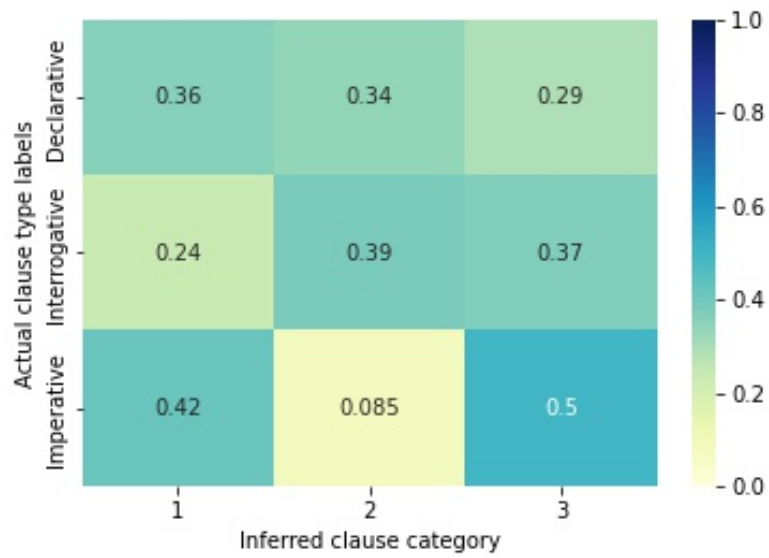


Figure 4.11: The proportion of actual declaratives, interrogatives, and imperatives clustered in one category

As we can see, the conservative SID fail to identify any clause types. Each cluster contains all three clause types, and all three clause types are distributed across the three clusters.



Figure 4.12: The number of sentences with/without certain formal features in each cluster, darker colors represent the number of sentences with the feature.

Figure 4.12 shows the distribution of morpho-syntactic features across all three clusters. The three clusters do not seem to form meaningful grammatical classes.

4.3.1.2 The knowledgeable simulation

The knowledgeable simulation of the SID model uses interrogative-specific features, $\pm A$ -not- A construction and $\pm ma$. But it also failed to identify any meaningful grammatical class in Mandarin.

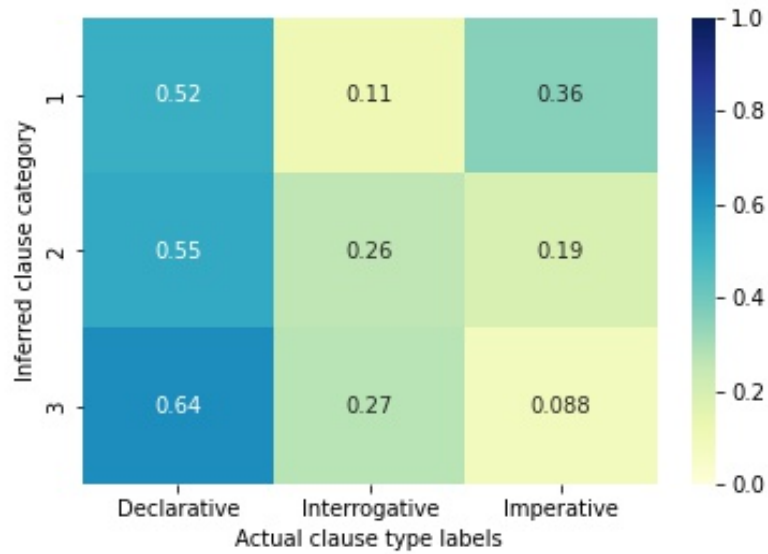


Figure 4.13: The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SID model



Figure 4.14: The proportion of actual declaratives, interrogatives, and imperatives clustered in one category

Overall, Figure 4.13 shows that all three clusters identified by the SID model are pre-

dominantly declarative sentences, which simply reflects the fact that declaratives are the most frequent clause type in the dataset. But compared to the conservative simulation, it puts more interrogatives in Cluster 2, as shown in Figure 4.14. But a closer look at the distribution of morpho-syntactic cues within each cluster (Figure 4.15 suggests that none of the clusters are meaningful grammatical classes in Mandarin.

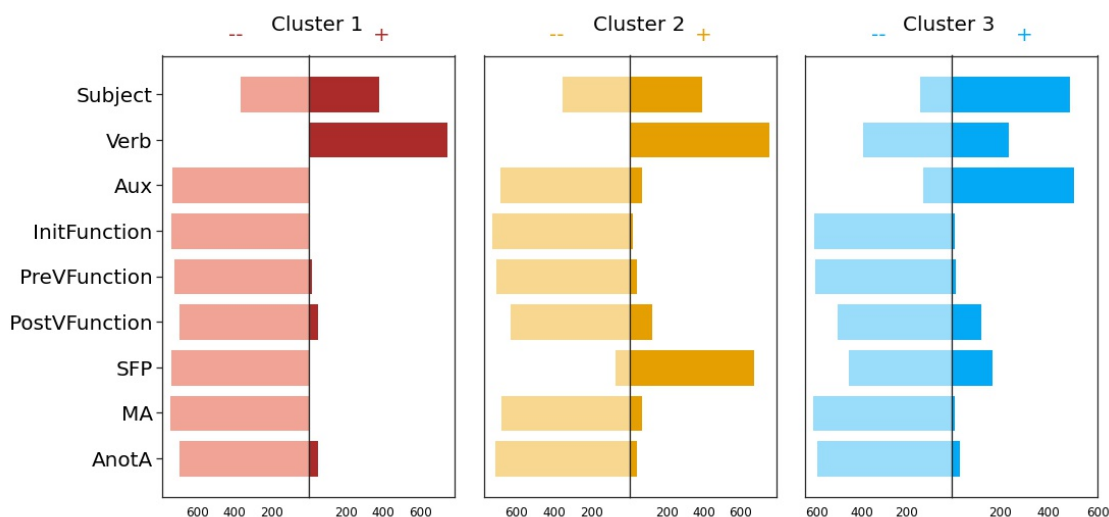


Figure 4.15: The number of sentences with/without certain formal features in each cluster, darker colors represent the number of sentences with the feature.

These results suggest that regardless of whether the model has access to interrogative-specific features or not, the SID failed to identify any of the clause types. It failed to find any meaning grammatical classes.

4.3.2 Performance of the SPID model

4.3.2.1 The conservative simulation

By having access to pragmatic information, the conservative SPID model could identify two clause types. Figure 4.16 shows the proportion of declaratives, interrogatives, and impera-

tives in each identified cluster, and Figure 4.17 shows the proportion of sentences clustered together.

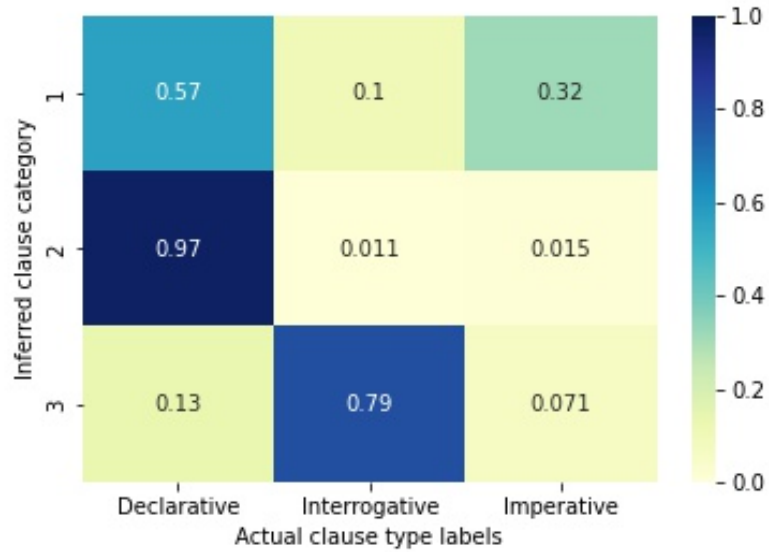


Figure 4.16: The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SPID model

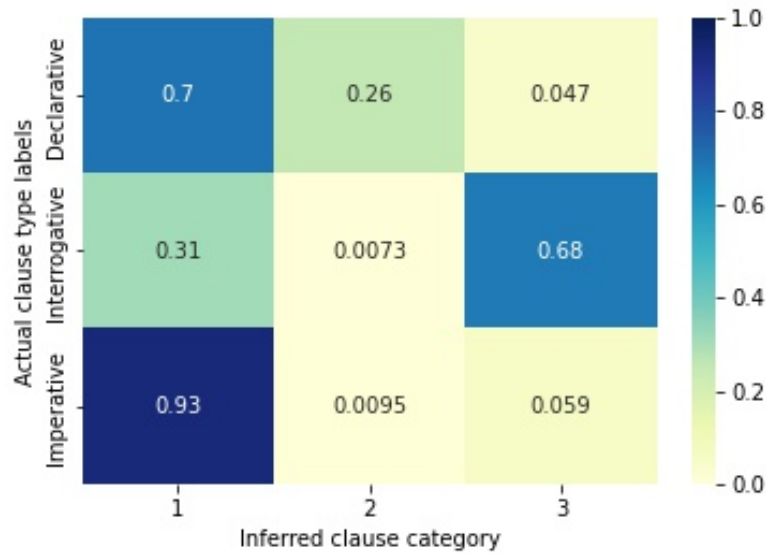


Figure 4.17: The proportion of actual declaratives, interrogatives, and imperatives clustered in one category

We can see that the SPID model clearly identifies a declarative and an interrogative cluster: 97% of Cluster 1 are declaratives, 79% of Cluster 3 are interrogatives. Declaratives and interrogatives in Mandarin are also mostly clustered together by the model: 70% of declaratives, 90% of interrogatives, and 93% of imperatives are clustered together in Cluster 1, 3, 1 respectively. In contrast to SID, this learner is able to find at least one clause type in Mandarin.

Figure 4.18 shows the morpho-syntactic profile of each cluster identified by the SPID.

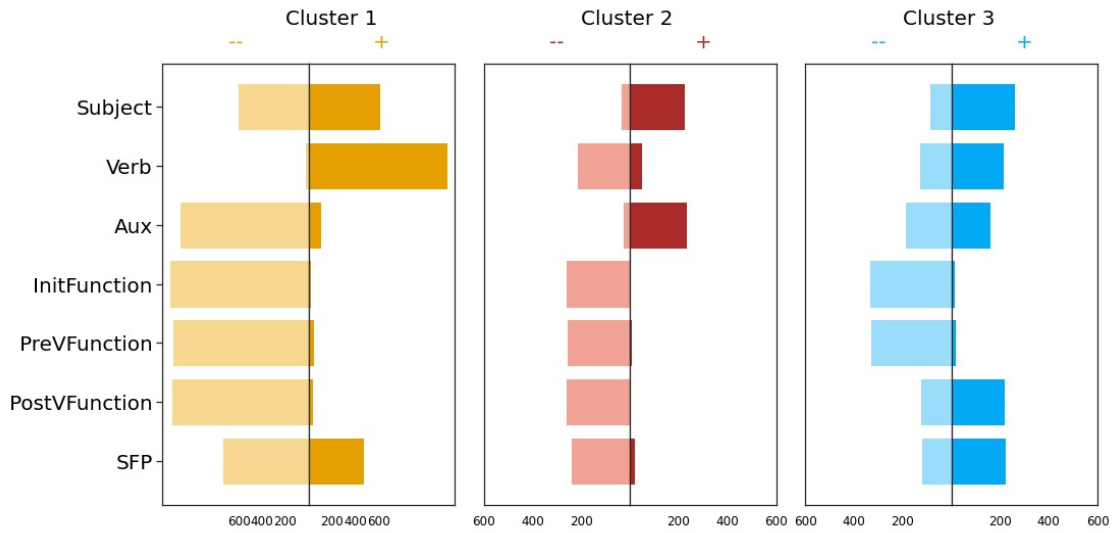


Figure 4.18: The number of sentences with/without a formal property in each cluster (Cluster 1 ~ imperatives, Cluster 2 ~ declaratives, Cluster 3 ~ interrogatives), darker colors represent the number of sentences with the property, lighter colors represent sentences without the property.

4.3.2.2 The knowledgeable simulation

By having access to pragmatic information, the knowledgeable SPID model could identify two clause types, one for interrogatives and one for declaratives, but imperatives and declaratives are still collapsed into one cluster. Figure 4.19 shows the proportion of declaratives, interrogatives, and imperatives in each identified cluster, and Figure 4.20 shows the proportion of sentences clustered together.

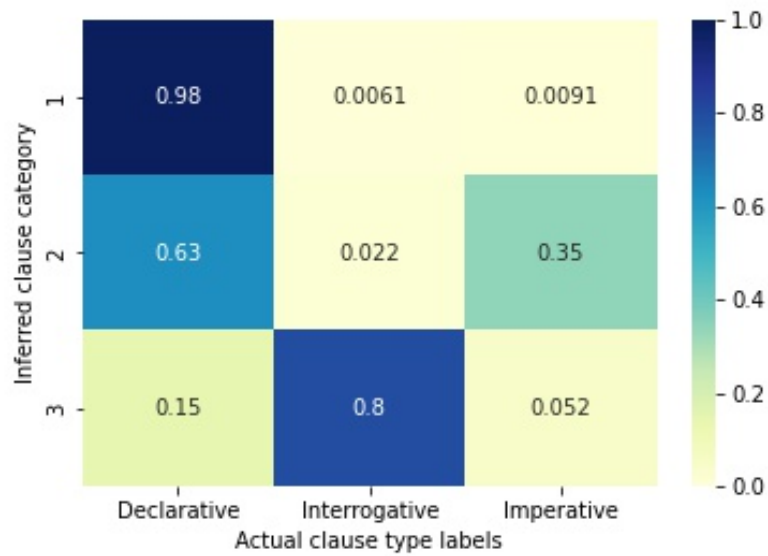


Figure 4.19: The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SPID model

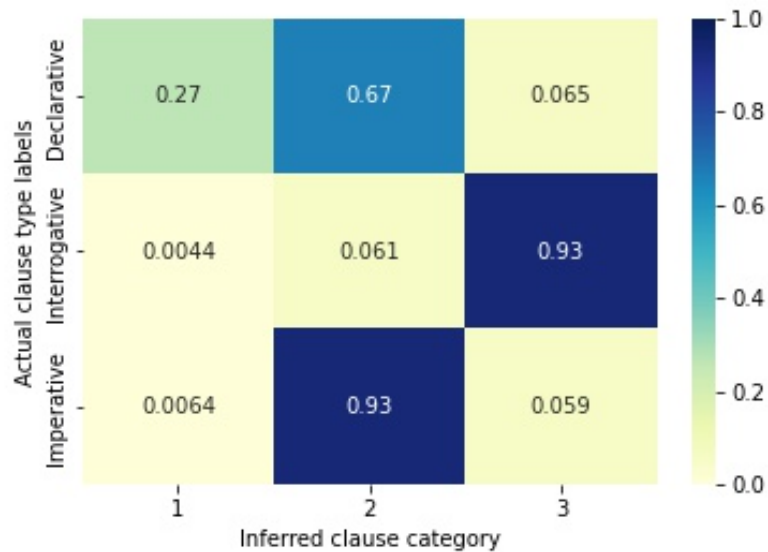


Figure 4.20: The proportion of actual declaratives, interrogatives, and imperatives clustered in one category

The SPID model identifies a declarative and an interrogative cluster: 98% of Cluster 1

are declaratives, 80% of Cluster 3 are interrogatives. The three clause types are also mostly clustered together by the model: 67% of declaratives, 93% of interrogatives, and 93% of imperatives are clustered together in Cluster 2, 3, 2 respectively. In contrast to SID, this learner is able to find at least one clause type in Mandarin.

Figure 4.21 shows the morpho-syntactic profile of each cluster identified by the SPID. The model correctly associates +A-not-A and +*ma* as the surface features for [+int], but it fails to identify the right property for [imp], as Cluster 2 contains almost all of the imperatives, but also a large portion of declaratives. It seems that Cluster 1 is a cluster specifically for ‘copula’ sentences like (77).

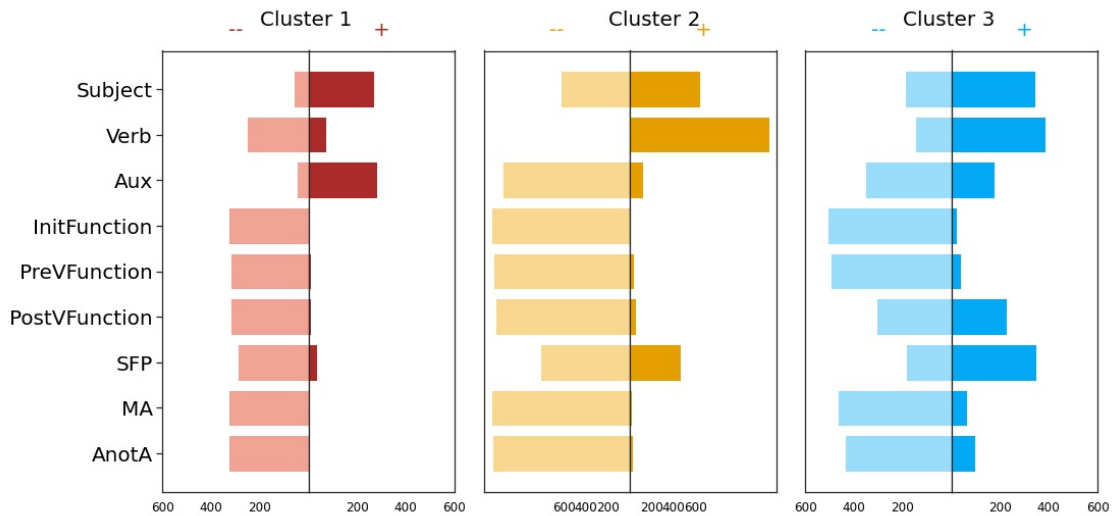


Figure 4.21: The number of sentences with/without a formal property in each cluster (Cluster 1 ~ declaratives, Cluster 2 ~ imperatives+declaratives, Cluster 3 ~ interrogatives), darker colors represent the number of sentences with the property, lighter colors represent sentences without the property.

- (77) Zhe xia, (zhe shang).
 This down (this up)
 “This is down. (This is up)”

4.3.2.3 Compare all four simulations

Figure 4.22 compares the performance of all four simulations. We can see that the two SID simulations are both bad at identifying clause types in Mandarin, but with pragmatics and some knowledge about A-not-A, the SPID outperforms the other three models.

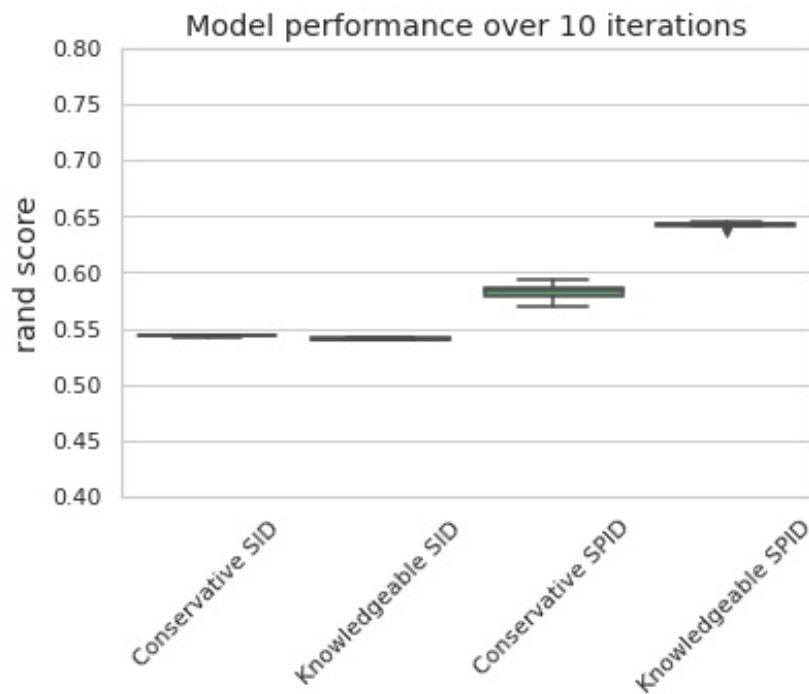


Figure 4.22: Performance of conservative and knowledgeable SID and SPID

Our results suggest that pragmatics is extremely crucial for Mandarin-acquiring children to identify clause types. Without pragmatics, learners might not be able to find any

meaningful categories. Next, we are going to look at how much pragmatics the learners need.

4.3.3 Simulations with noisy pragmatic information

Figure 4.23 and 4.24 show the simulations with 0-100% noise in the pragmatic information, as a conservative learner and a knowledgeable learner respectively. The performances of the SID models are used as baselines and marked by the dotted line. As can be seen, the conservative SPID model tolerates 20% of noise in the pragmatics, and the knowledgeable model tolerates about 40% of noise. Both are lower than the threshold for their English counterparts. This suggests that Mandarin learners rely heavily on the pragmatic information to identify clause type categories, as compared to their English counterparts.

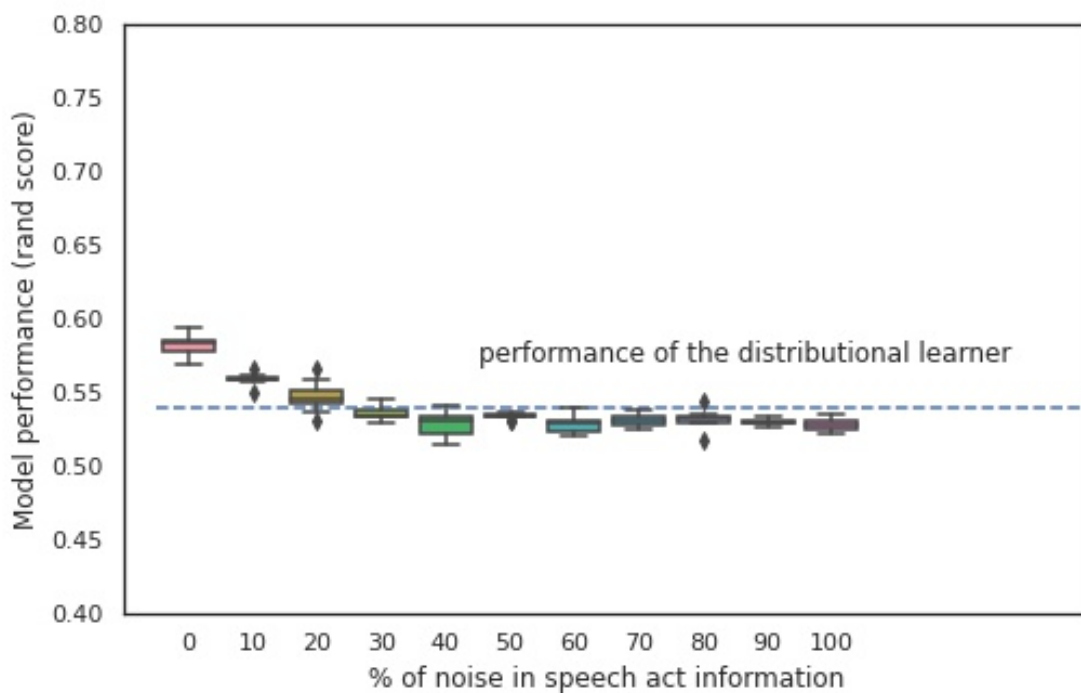


Figure 4.23: Performance of the conservative SPID model with different levels of noise in the speech act information; dotted marks the rand score of the SID learner

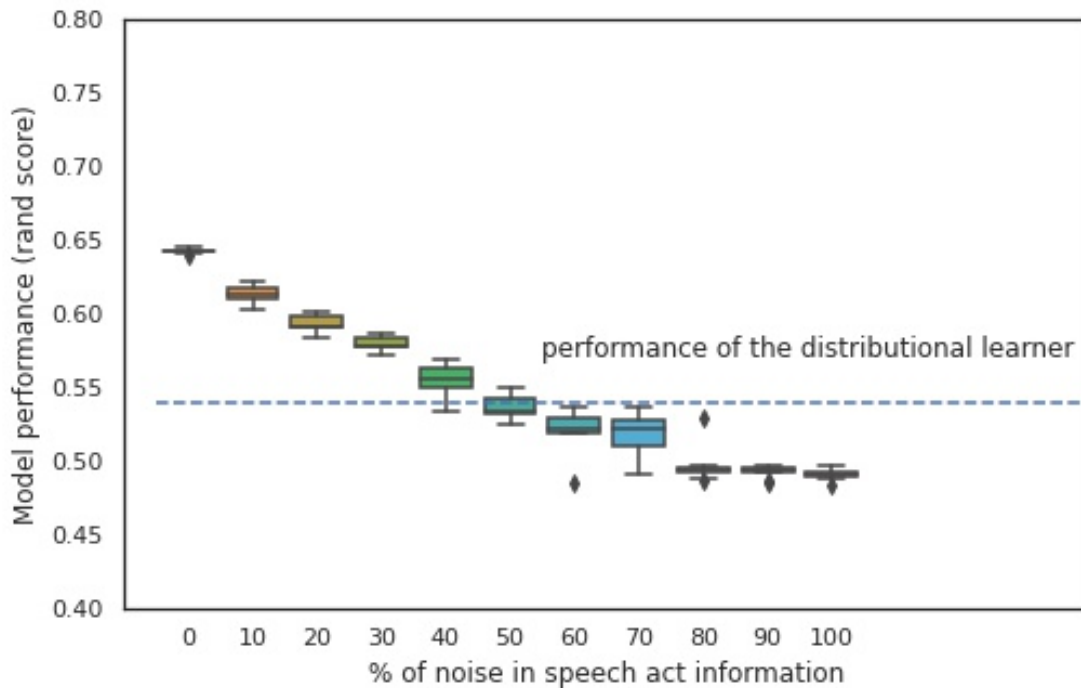


Figure 4.24: Performance of the knowledgeable SPID model with different levels of noise in the speech act information; dotted marks the rand score of the SID learner

4.4 Discussion

In this chapter, we investigated how Mandarin-acquiring 18-month-olds can learn to identify the three clause types. As [+int] and [imp] in English and Mandarin are related to different surface features, comparing these two languages provides us with an opportunity to see the role of pragmatics in different languages.

Just like English parents, Mandarin parents also use clause types systematically, and the three clause types are predominantly mapped to their canonical functions (declaratives to assertions, interrogatives to questions, imperatives to requests/commands). The three clause types show distinctive formal profiles in the input available to children: compared to declaratives, interrogatives are more likely to have post-verbal UFIs, and more likely to

have A-not-A and sentence final *ma*; imperatives are more likely to not have subjects and negative modal *bie*.

We then investigated the extent to which Mandarin learners need to rely on pragmatic information to identify clause types. I showed that, SID performs worse in Mandarin than in English, as it simply cannot identify any of the right categories, even when we loosen our assumptions on learners' morpho-syntactic knowledge. But adding pragmatic information improves the performance significantly, as the SPID could identify interrogative and declarative clauses, but still have problem identifying imperatives. Whereas high level of noise in pragmatic information still helps virtual learners figure out clause types in English, Mandarin learners require more accurate pragmatic information to succeed; when the noise in speech act exceeds 40 % (20% for learners with less morpho-syntactic knowledge), performance drops to the same level as syntax-only models. Thus, pragmatics is essential for both English and Mandarin learners. But pragmatics is even more crucial for Mandarin learners.

While adding pragmatic information improves the performance of the model, it still fails to identify all three clusters. Why do both models fail when simulating with Mandarin data? The problem seems to be that declaratives and imperatives simply do not look that different in the input. While most scholars agree that imperatives form a separate clause type in Mandarin (C.-H. Han 1998, Chen-Main 2005b among many others), there are few cases of sentences with the special marker *bie* that is unique to imperatives, which is not significant enough to differentiate the two clause types.

One possibility is that we need to include more morpho-syntactic features. In this study, I only included clausal features that are relevant for clause typing as reported in the literature. But it is likely that even though some features do not in principle differentiate [+imp] and [-imp], it is likely that there is a bias in the distribution in the input.² For example, it is likely that the learner not only need to track whether there is a sentence final particle in the sentence, but they might also need to track *which* particle is used. The particle *ba*,

²See N. Huang et al. (2022) for a discussion on a similar problem with embedded clauses.

for example, theoretically cannot distinguish [+imp] from [−imp], but it is likely that the distribution of this particle in the input is informative for the learner. The same for aspect markers: even though in principle aspect markers can be found in both imperatives and declaratives, it is possible that they are more frequent with declaratives. Since infants seem to be sensitive to the presence of aspect markers from early on (e.g. Yang, Shi, & Xu 2018), observing the distribution of this feature might help the learner find the right clustering. I plan to annotate more features to see if it would improve the models' performances.

But it is also possible that this is exactly what the learners are doing: they mix declaratives with imperatives at the beginning. As they hear more sentences with *bie*, they can then have a dedicated category for imperatives. As we don't have clear evidence from the literature about when Mandarin learning children actually figure out clause types/speech acts, it is possible that the knowledge of clause types (and imperatives in particular) might come later for Mandarin children than their English peers.

Another way the performance could be improved is to add in the prosodic information. Due to the fact that Mandarin is a tonal language, to extract meaningful intonation information from real speech (as opposed to experimental data where lexical tones are well-controlled) is a huge challenge. Additionally, it is unclear whether 18-month-olds are sensitive to these intonational features. With the complication of lexical tones, children might first need to learn to discern tonal and intonational features. It is likely that they might need some knowledge of clause type and speech act to find the correct clustering of prosodic features. As English does not have the complication of lexical tone, it provides us a simpler case to study the role of prosody in speech act and clause type learning. In the next chapter, I report how parents use prosody and whether it's useful for learning clause typing in English, in the hope that we can extend our investigation to Mandarin in the future.

Chapter 5: Prosody, clause types, and speech acts in English

One important source of information that we have not yet discussed is prosody. As mentioned in Section 2.1.3, the role of prosody in clause typing and speech act identification is quite complicated. This chapter focuses on English prosody and its relation with clause type and speech act. English offers an interesting case, as its interrogative clauses could be marked by morpho-syntactic features (subject-auxiliary inversion), but are also tend to be associated with a rising intonation (polar interrogative only). How should we understand the role of prosody in clause typing and speech act identification? How could the learners utilize this information in learning clause type and speech act categories? In Section 5.1, I discuss the different ways people have linked English prosodic features with clause typing, and the implications for the learners. In Section 5.1.2, I review the literature on children’s knowledge of prosody. Even though we do not have direct evidence for English-acquiring 18-month-olds, it is likely that they are at least sensitive to the distinction between final rise and fall. If they are indeed sensitive to this distinction, would this distinction correlate with clause types/speech acts in the input? I report results from a corpus study in Section 5.2, and use the annotated dataset to simulate whether children could learn clause types with both morpho-syntactic and prosodic information (Section 5.3). Results from both studies suggest that there might not be a clear correlation between final rise and clause types/speech acts, and that this prosodic feature is not informative for finding the right clause type clusters in English. I discuss the reasons for seeing these patterns in the input, and some alternative ways that prosodic information could be utilized in learning clause type clusters

in Section 5.4.

5.1 Background

5.1.1 Prosody, clause types, and speech acts in English

In English, whether prosody is a surface feature for clause typing or simply an indicator for questioning is debated. As we have seen repeatedly, English interrogatives have their designated morpho-syntactic features, i.e. subject-auxiliary inversion. But prosodic information could also inform us about the clause type category. English declaratives tend to be associated with falling intonation, polar interrogatives tend to bear a final rising intonation, while *wh*-interrogatives also generally bear a falling contour. This final rise and fall are generally believed to be the nuclear contour (Ladd 1981, 2008, Hedberg, Sosa, & Fadden 2004, Hedberg, Sosa, & Görgülü 2014 among many others). The nuclear contour is the contour from the nuclear pitch accent to the end of the utterance. The nuclear pitch accent is the final pitch accent (a pitch accent is a marker of greater relative prominence on a syllable). In other words, the relevant contour is the nuclear contour and whether it rises or falls.

As we mentioned in Chapter 2.1.3, cross-linguistically languages could use morpho-syntactic features alone, prosodic alone, or a combination of both for clause typing. We also tentatively put English in the “combination” column. But is the final rise a property of English interrogatives? Or is it simply a property of questionhood, and that some interrogatives are associated with this feature because they are used to perform questioning acts? To address this issue, we should look at the cases where speech acts and clause types misalign. Specifically, when imperatives and declaratives and imperatives are used to ask questions, and when interrogatives are used to perform assertions and requests.

In Chapter 3.1.2, I argued that rising declaratives are declaratives that can be used to ask questions. But declarative questions do not need to be associated with final rise. Embedded questions like (78) are declaratives used to perform questions, but the most natural prosodic

contour associated with it should be the same as declarative assertions.

(78) I wonder if you could tell me what time it is.

Similarly, imperatives used to perform questions like (79) could also have falling intonation:

(79) Tell me what time it is.

When used to perform assertive acts, both polar and *wh*-interrogatives tend to be associated with a final fall. The following example comes from a production study by Dehé & Braun (2020):

(80) Context: One of your friends suggest that you should all go to the museum. But it is known to all of your friends that none of your friends likes to go to the museum, and therefore nobody responded to his suggestion. You say to your friend:

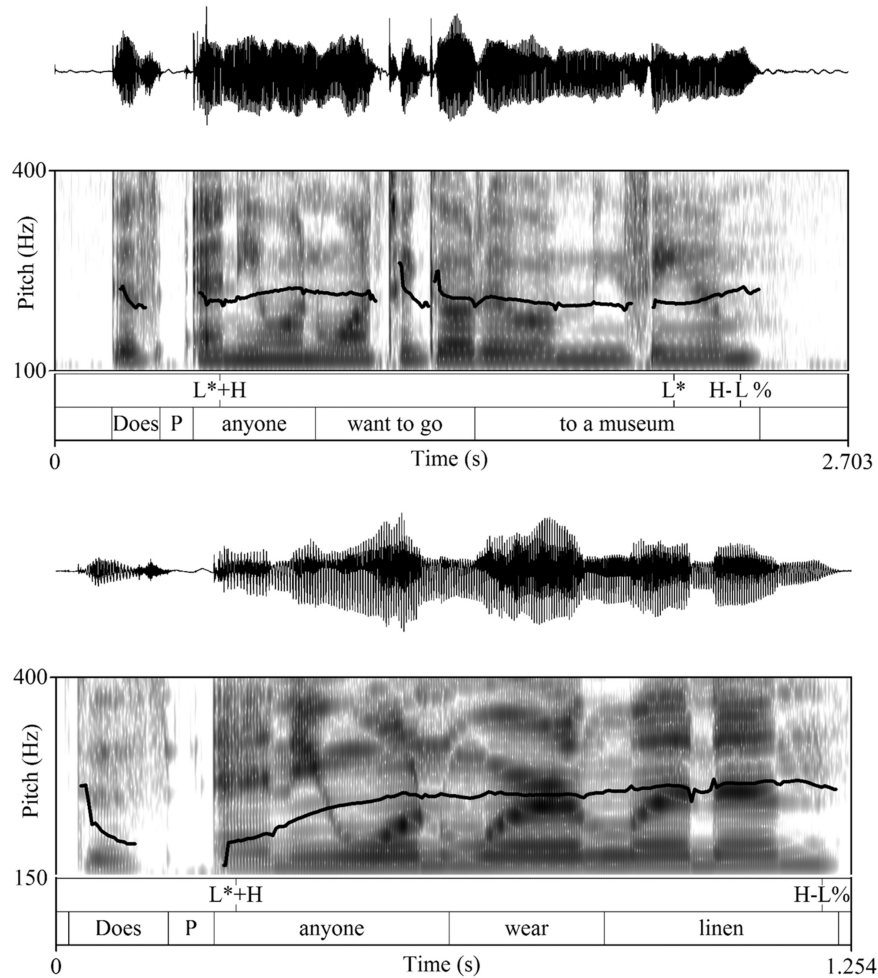


Figure 5.1: Two types of frequent rhetorical polar interrogatives produced by a female (up) and male (down) participant with contexts similar to (80) (Figure 11 from Dehé & Braun 2020)

But Dehé & Braun (2020) also note that a substantial amount of sentences are produced with final rise H-H% (examples are not given in the paper).

When interrogatives are used to perform requests like *Can you pass the salt?* the most natural prosodic contour seems to be the regular polar interrogative contour.

To summarize, declaratives and imperatives used to ask questions have final rises (rising declaratives), but also have falling intonation (embedded questions). Interrogatives used to perform an assertion or a request are associated with a rising intonation. These patterns seem to suggest that English final rise (a rising nuclear contour) could be triggered by a question

act, but could also be triggered by the right matrix clause type (polar interrogatives).

5.1.2 Children’s knowledge of prosodic features

As we have seen in Chapter 2, rising intonation tends to associate with interrogatives (particularly polar interrogatives) and questions cross-linguistically. Results from previous experimental and corpus studies suggest that children might be sensitive to the distinction between a final rise (specifically polar interrogative rise) and a final fall.

From as early as 6 months old, infants are able to identify utterance boundaries, and are sensitive to the edge prosody (Johnson, Seidl, & Tyler 2014). They also can use distinctions in prosodic contours (e.g. final rise vs. fall) to distinguish clause types (polar interrogative vs. declaratives). For example, Frota, Butler, & Vigário (2014) show that as young as 6 months old, European Portuguese-acquiring infants are sensitive to the prosodic distinction between polar interrogatives and declaratives, where prosody is the only cue that distinguishes these two clause types in the language. Geffen & Mintz (2011) find that when given a combination of word order and prosodic cues, English-acquiring 7-month-olds can distinguish polar interrogatives from declarative assertions. Soderstrom, Nelson, & Jusczyk (2005) test English-speaking infants between 4.5 months and 2 years old (average 14 months old), and find that they are sensitive to the distinction between declaratives with a falling a rising contour. Before 18 months old, infants acquiring non-lexical tone languages such as English are shown to be sensitive to some tonal distinctions in lexical tone languages such as Mandarin. They are particularly sensitive to the distinction between the rising tone that is similar to English polar-interrogative rise, and the falling tone similar to English declarative fall (Shi et al. 2017, Hay, Cannistraci, & Zhao 2019), although it is unclear whether we can draw any conclusions about their sensitivity to English polar interrogative rise.

While infants might be sensitive to the distinction between rise and fall, it is unclear, at least for English-acquiring infants, whether they can use prosodic cues alone to detect questions. When given only prosodic cues without any morpho-syntactic information, Keitel

et al. (2013) and Casillas & Frank (2017) both find that children younger than 2 years old cannot use intonation alone to infer transition after a turn (i.e. casting more anticipatory looks to the addressee), but they can infer such transitions with morpho-syntactic cues associated with interrogatives. However, the low-pass filtered utterances might be difficult for infants to process, so their failure at casting anticipatory looks after rising intonation might be an artifact of task effect.

On the production side, studies have found that children are able to use different prosodic contours for different functions from when they start speaking. Menyuk & Bernholtz (1969) analyze the prosodic contours of one child's utterances, and annotates their perceived speech acts between 18 and 20 months, and find that even though the utterances are mostly one word or two words, there's a correlation between the intended speech act and prosody. For example, an intended request with a one-word utterance "door!" is typically associated with a sharp rise and then fall; the same utterance as a question tends to end with a rising intonation. Since the speech act labels in the study are annotations by adults inferred from one-word utterances, it is unclear whether these are actually the speech acts that the child perform, and the conclusion seems to be that adults systematically associate certain prosodic contours to assertions, questions, and requests/commands, even with one-word utterance. Nonetheless, it seems that at the child is using different prosodic contours at this age, and it is possible that these contours are systematically associated with different speech acts.

Furthermore, for their own production, children seem to associate the rising contour with response elicitation. Flax et al. (1991) conduct a longitudinal study observing three children interacting with their mothers before they can speak (when they have a vocabulary of 10 words, and again when their vocabulary consists of 50 words), and code whether the child's utterance (or vocalization, at the pre-verbal stage) is produced with a final rise. They find when children request a response from their addressee, they tend to produce the utterance with a final rise.

Previous studies have also shown that there are prosodic cues that distinguish clause

types and speech acts, in particular to distinguish polar interrogatives and declaratives, in child-directed speech. Geffen & Mintz (2017) show that polar interrogatives and declaratives differ in the pitch of the last two syllables, with polar interrogatives generally have a rising contour; they also find that there is no distinction between *wh*-interrogatives and declaratives in the last two syllables. Chiang, Geffen, & Mintz (2018) examine sentence-initial prosodic cues, and find that polar interrogatives and declaratives both have a higher starting pitch than declaratives, and that echo *wh*-questions tend to have a higher pitch than other types of questions.

In sum, prosodic cues are in principle helpful for distinguishing different speech acts, and there is evidence suggesting that children can perceive these cues. In particular, there might be cues at sentence-final and sentence-initial position that are both useful and available to children. We do not have sufficient data to know whether 18-month-olds are sensitive to the prosodic distinctions between declaratives and polar interrogatives, but evidence from other languages suggests that infants around this age are at least sensitive to the final fall vs. rise distinction. If children are indeed sensitive to this distinction, would they find anything in the input? In the next section, I report results from a corpus study examining the prosodic features of parents' speech acts and clause types.

5.2 Corpus study

5.2.1 Methods

This study also used data from the Providence Corpus (Demuth, Culbertson, & Alter 2006) from CHILDES system (MacWhinney 2000). The audio and video of the sessions sampled in Chapter 3 were extracted for annotation.

For the audio data, I adapted a Kaldi forced alignment system (Povey et al. 2011) to obtain a time-aligned dataset containing the beginning and ending timestamp of each utterance in a session. I also manually aligned 20% of this dataset to compare for accuracy. The

mean difference between the manually aligned dataset and the forced-aligned dataset is 0.1s at the beginning of the utterance (0.001s to 20s), and 0.08s at the end of the utterance. We then extracted the pitch information of each utterance using a Python library for the Praat software (Boersma 2001), Parselmouth (Jadoul, Thompson, & de Boer 2018).

The pitch information from 3366 utterances were extracted by using Praat (Boersma 2001). I further applied a low-pass filter at 500 Hz that removes most of the phonetic information used to distinguish between phonemes.

As discussed in the previous section, what we are most interested in is the nuclear contours and whether they rise or fall. The nuclear contour is defined as the contour from the nuclear pitch accent to the end of the utterance; the nuclear pitch accent is the final pitch accent, and a pitch accent is just a marker of greater relative prominence on a syllable (Büring 2016). To my knowledge, there is currently no perfect way to automate recognition of (nuclear) pitch accents (see Büring 2016 for a discussion). I therefore applied the following algorithm to identify the nuclear contours of these 3366 utterances, and hand-annotated 100 for comparison. I first applied a peak identification algorithm (`scipy.signal.find_peaks`; from the `scipy` python package). This function takes a 1-D array and finds all local maxima by simple comparison of neighboring values. The peaks in F0 would be our approximation for the pitch accent (most prominent syllable). But prominence is a perceptual category: a syllable that is prominent is in some sense intuitively "stronger" than surrounding syllables, according to human intuition. This could either be the highest or lowest F0. I therefore inverted the pitch track and applied the find peak algorithm again; the results would be a set of valleys in the pitch track. The last peak or valley was then considered the last pitch accent. If this last "pitch accent" is a peak like (5.2), then the utterance was labelled as having a final rising contour.¹ If the last "pitch accent" is a valley, then the "nuclear contour" was considered a rise.

¹The script for coding pitch information can be found at <https://osf.io/u9378/>.

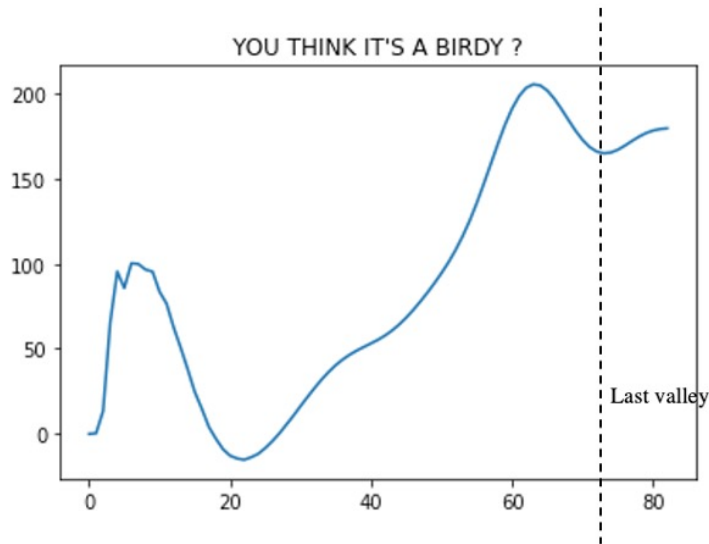


Figure 5.2: “You think it’s a birdy?”

I then manually annotated 100 utterances to check for accuracy; 52 of these were correctly labeled by the algorithm. The low agreement rate could be due to a number of factors. Besides the fact that the recordings themselves are noisy, one of the problems with automating the process is that the tail end of all pitch tracks can be distorted in some way. The reason is that pitch periods are irregular due to dissipating energy, which may cause spurious events in the waveform to be misinterpreted as a pitch period, which Praat then incorrectly interprets as pitch. I plan to hand annotate more utterances, and improve the algorithm.

For prosodic patterns, we should see that final rising contour is more frequently associated with questions, especially polar interrogatives, than with assertions.

5.2.2 Results

Figure 5.3 and 5.4 show the proportion of final rise, as identified by the algorithm. As we can see, there is no difference in the proportion of rises among speech acts.

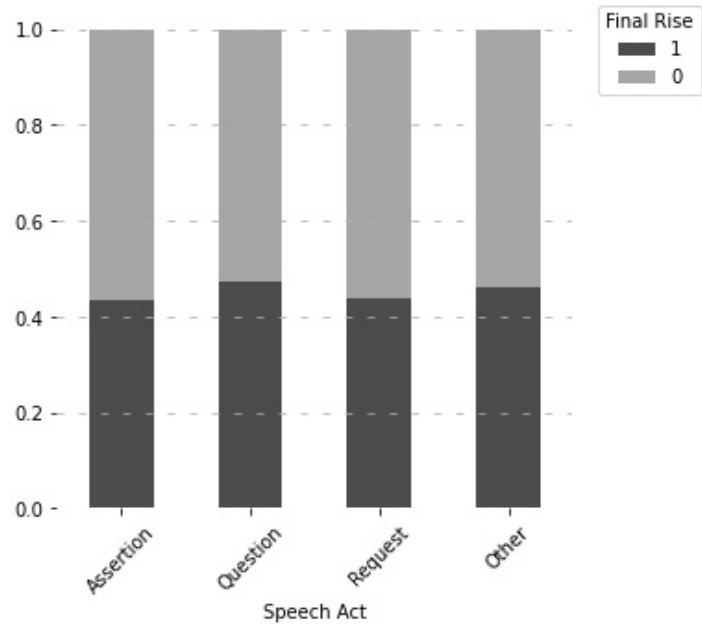


Figure 5.3: Proportion of utterances with final rise across different speech acts

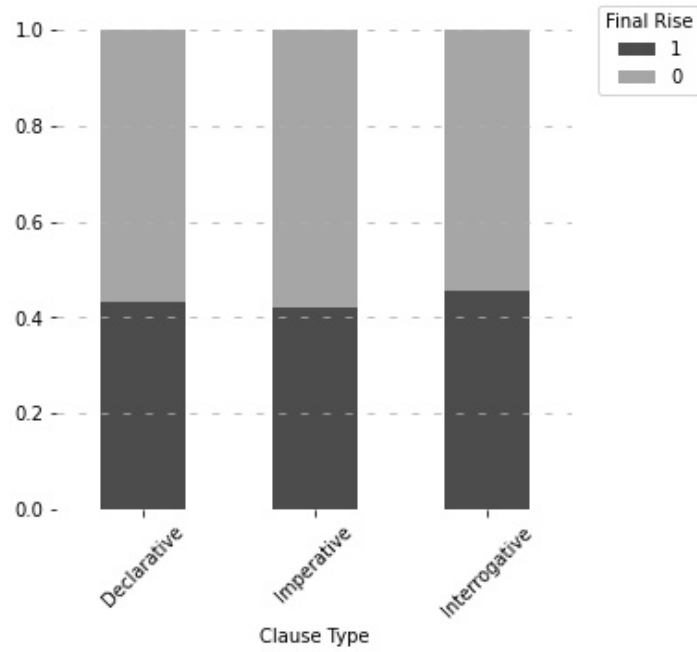


Figure 5.4: Proportion of utterances with final rise across different clause types

But if we look into sub-categories of interrogatives, we can see that the proportion of final rise is much higher with polar interrogatives than with declaratives.

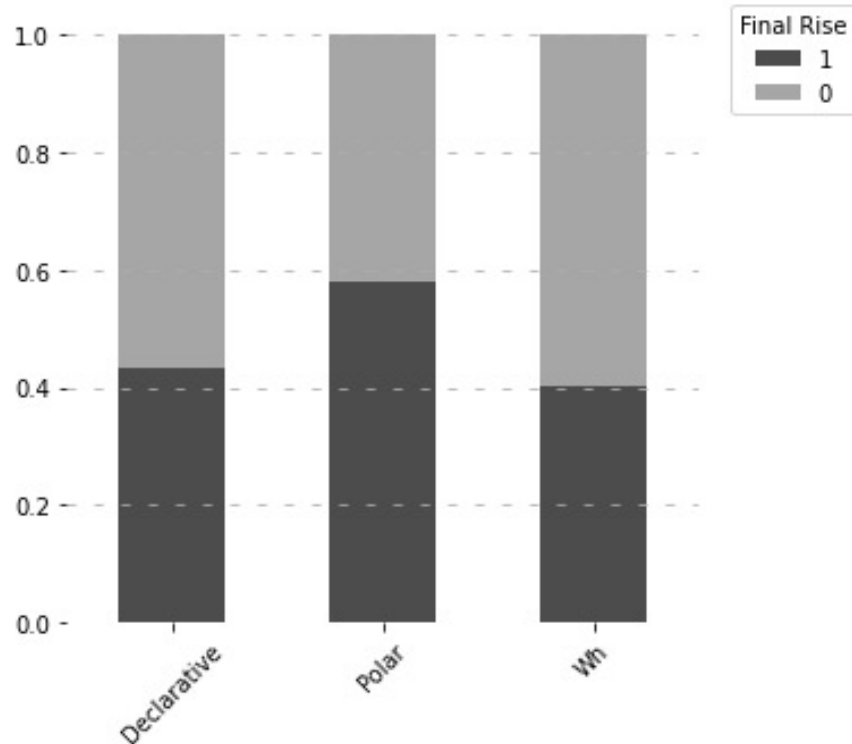


Figure 5.5: Proportion of declaratives, *wh* and polar interrogatives with final rise

Despite being a noisy approximation of the actual data, I found that some signal for polar interrogatives, as predicted. Specifically, polar interrogatives are more often associated with final rises than declaratives and *wh*-interrogatives.

These results also seem to suggest that the presence of final rise might not be informative of the speech act of the sentence, unless morpho-syntactic features like subject-auxiliary inversion is also present. However, this could be a result of the algorithm I applied here, and do not reflect the pattern of the data. I plan to hand annotate more cases in the future to have more reliable data.

5.3 Learning clause type categories with prosody

Our next question is, would prosodic information help the two learners (syntactically informed distributional learner and syntactically and pragmatically informed distributional

learner) find the right clause type clusterings. If prosody helps, then the performance of SID will improve with the addition of prosodic information, and the model should be able to find the right clusterings. To this end, I modified the models given in Chapter 3 to include prosodic information. The modified models are given below:

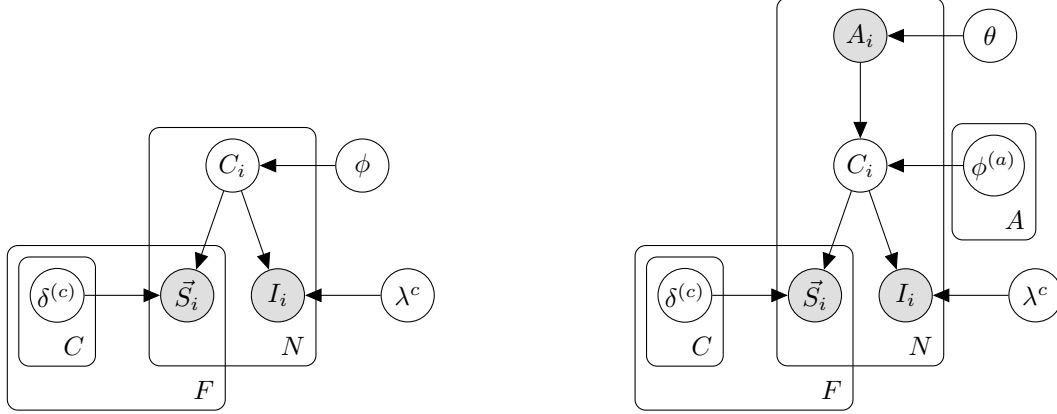


Figure 5.6: The syntactically informed distributional learner model with prosody (left) and syntactically and pragmatically informed distributional learner model with prosody (right)

Both model now learns from prosodic data, represented in Figure 5.6 as I_i . This variable is a Bernoulli variable conditioned on c_i . It takes the value 1 if the sentence is associated with a final rise intonation, and 0 otherwise. It is also conditioned on the parameter $\lambda^{(c)}$, which represents the probability of observing a final rise in a particular clause type category.

The same sampling method as reported in Chapter 3.3 was used. The posterior of c_i is updated with the following equations:

$$\begin{aligned}
 p(c_i | c_{-i}, \vec{a}, \beta, \vec{S}_i, \delta, \gamma, l_i, \lambda, \eta) &= \frac{p(\vec{S}_i | \vec{c}, \delta, \gamma) p(l_i | \vec{c}, \lambda, \eta) p(c_i | \beta, c_{-i})}{\sum_{c'_i} p(\vec{S}_i | \vec{c}', \delta, \gamma) p(l_i | \vec{c}', \lambda, \eta) p(c'_i | \beta, c_{-i})} \\
 &= \frac{\prod_F \frac{\gamma_0 + n_{s_i^{F, c_i}}}{2r_0 + n_{c_i^F}} \frac{\eta_0 + n_{l_i^{c_i}}}{2\eta_0 + n_{c_i}} \frac{\beta_0 + n_{c_i}}{3\beta_0 + n}}{\sum_{c'_i} \prod_F \frac{\gamma_0 + n_{s_i^{F, c'_i}}}{2r_0 + n_{c'_i^F}} \frac{\eta_0 + n_{l_i^{c'_i}}}{2\eta_0 + n_{c'_i}} \frac{\beta_0 + n_{c'_i}}{3\beta_0 + n}} \quad (5.1)
 \end{aligned}$$

$$\begin{aligned}
p(c_i|c_{-i}, \vec{a}, \beta, \vec{S}_i, \delta, \gamma, l_i, \lambda, \eta) &= \frac{p(\vec{S}_i|\vec{c}, \delta, \gamma) p(l_i|\vec{c}, \lambda, \eta) p(c_i|\beta, c_{-i}, \vec{a})}{\sum_{c'_i} p(\vec{S}_i|\vec{c}', \delta, \gamma) p(l_i|\vec{c}', \lambda, \eta) p(c'_i|\beta, c_{-i}, \vec{a})} \\
&= \frac{\prod_F \frac{\gamma_0 + n_{s_i^{F, c_i}}}{2r_0 + n_{c_i^F}} \frac{\eta_0 + n_{l_i^{c_i}}}{2\eta_0 + n_{c_i}} \frac{\beta_0 + n_{c_i^{a_i}}}{4\beta_0 + n_{a_i}}}{\sum_{c'_i} \prod_F \frac{\gamma_0 + n_{s_i^{F, c'_i}}}{2r_0 + n_{c_i^F}} \frac{\eta_0 + n_{l_i^{c'_i}}}{2\eta_0 + n_{c_i'}} \frac{\beta_0 + n_{c_i^{a_i}}}{4\beta_0 + n_{a_i}}} \quad (5.2)
\end{aligned}$$

The data for this model were taken from the annotated dataset reported in the last section. In total, 3366 sentences were fed into the models.

Overall, these two models do not improve on the models without prosody, as shown by the similar adjusted rand score in Figure 5.7.

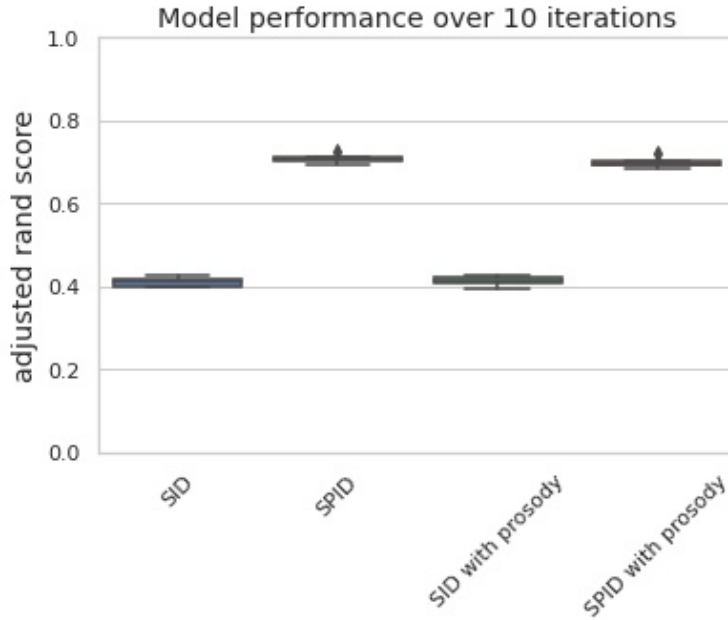


Figure 5.7: Performance of all four models as measured by adjusted rand score

The clusters identified by the two models with prosody also do not improve on previous models. As shown by Figure 5.8 and Figure 5.9, SID with prosody still cannot identify the

right clusters, and SPID could find the right clusters.

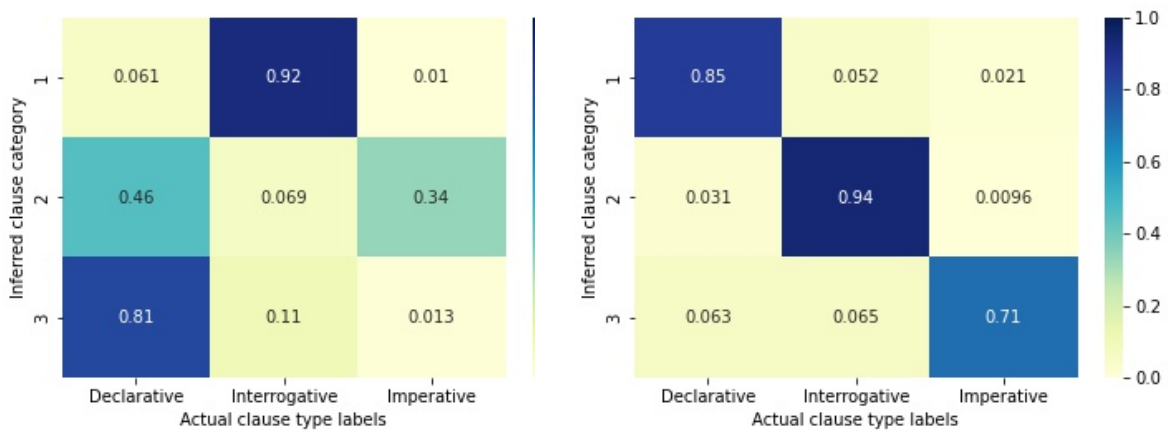


Figure 5.8: The proportion of declaratives, interrogatives, and imperatives in each of the three clusters identified by the SID with prosody (left) and the SPID model with prosody (right).

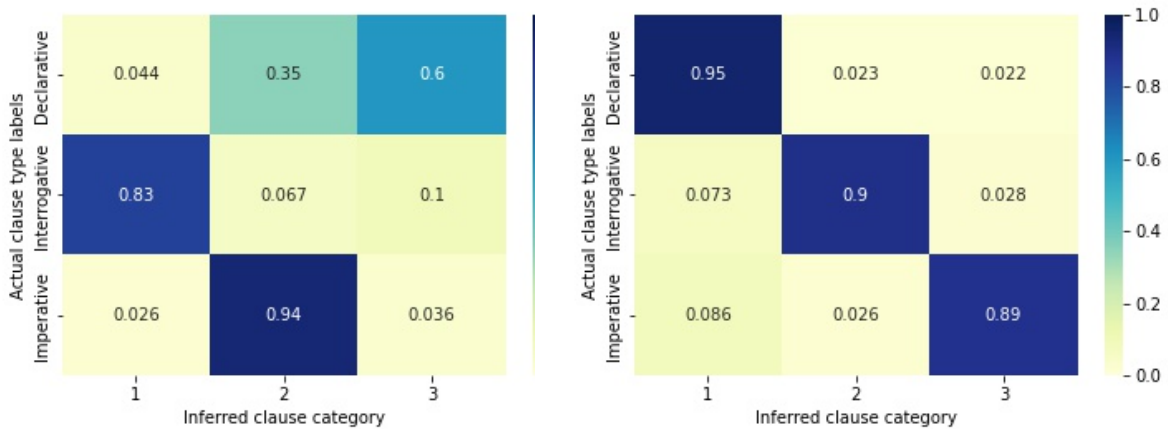


Figure 5.9: The proportion of actual declaratives, interrogatives, and imperatives clustered in one category as identified by the SID+prosody model (right) and SPID+prosody model (left)

Figure 5.10 and Figure 5.11 show the surface features identified by the two models with prosody. Again, the SID with prosody cannot identify [– subject] for imperatives, and SPID can.

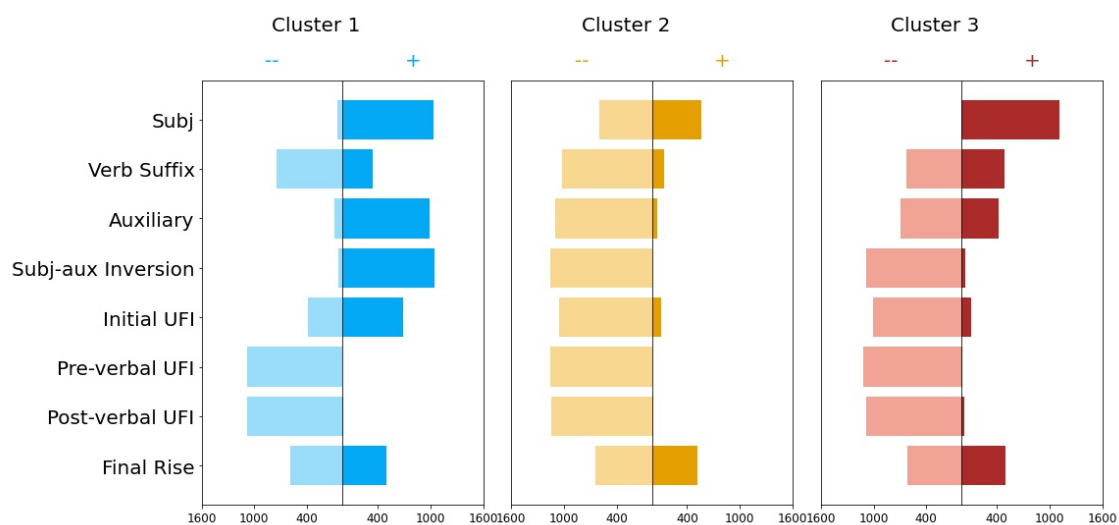


Figure 5.10: Morpho-syntactic profile of each cluster identified by the SID+prosody model (Cluster 1 ~ Interrogatives , Cluster 2 ~ Imperatives, Cluster 3 ~ Declaratives).

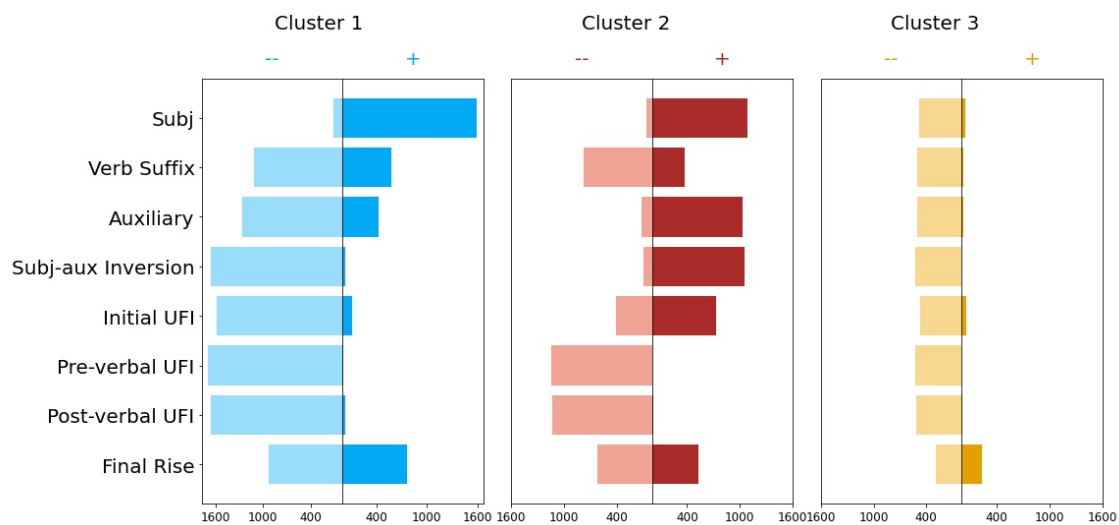


Figure 5.11: Morpho-syntactic profile of each cluster identified by the SPID+prosody model (Cluster 1 ~ Declaratives , Cluster 2 ~ Interrogatives, Cluster 3 ~ Imperatives).

Results from our simulations suggest that prosodic features do not improve the performance of SID; pragmatics is still necessary for finding the right clusters.

5.4 Discussion

Cross-linguistically, pitch rises tend to signal questions and pitch falls signal assertions; and some argue that this universality reflects the innate knowledge that high pitch connects to the speech act of questioning (Ohala 1984, Gussenhoven & Chen 2000, Gussenhoven 2002 among others). If children are armed with the knowledge that questions tend to be associated with rising contours, they might expect rising contours to be somewhat correlated with the act of asking a question. But just as not all interrogatives have subject-auxiliary inversion, not all questions have final rises. With preliminary data, I found that parents do not use final rises more often with questions, but polar interrogatives have more final rises than other types of speech acts and clause types, including *wh*-interrogatives and declaratives. Adding prosody to the SID and SPID models does not improve their performances.

The correlation between final rise is not with the questioning speech act, but with a specific type of interrogative. Since our model assumes that the presence of final rise predicts the use of declaratives and interrogatives as a whole, this additional information did not improve the models' performances. However, it is possible that the contribution of prosody is more complicated. For example, the learner might need to combine prosody with some morpho-syntactic features for this information to be useful. Learners might first come to associate polar interrogatives with questions. They might then be able to bootstrap other interrogatives to questions by noticing similarities in morphosyntax (e.g., whether there is subject-auxiliary inversion).

It is also possible that instead of three categories, the models perform better identifying four categories, with polar and *wh*-interrogatives separated. These two types of interrogatives share the same sentential force, but not the same set of morpho-syntactic or prosodic features.

Polar interrogatives do not have *wh*-phrases and are associated with final rises, but *wh*-interrogatives have *wh*-phrases and are associated with final falls. It is then likely that they should not be treated as a unified category by the learner. I leave these explorations to future research.

Chapter 6: Learning speech acts

As we have discussed in the last three chapters, pragmatics is crucial for children to learn clause type categories, and even noisy pragmatics can benefit the learner. But then, there still remains the question of how children learn speech act categories? Given that the way we generally identify a speech act is via its clause type, there is a potentially vicious circle here: you need to know the clause type to identify the speech act, but you need to know the speech act to identify the clause type.

But as we have discussed in Chapter 1, it is likely that children learn to identify speech act and clause type in tandem and mutually informative ways: children learn to identify clause types by tracking morpho-syntactic and prosodic regularities in conjunction with their growing knowledge of speech act and its associated behavioral cues; similarly, they learn to identify speech acts by tracking social pragmatic cues and prosodic cues in conjunction with their growing understanding of the formal features of various clause types.

In Chapter 3 and 4, we have seen that even if children have limited access to the speech act information, they might still benefit from it when learning to cluster clause types. If children do not need *perfect* speech act information to figure out clause types, then we do not need to assume that by the time children are learning to figure out clause types, they have already figured out speech acts.

But what about the learning of speech act categories? Even if children must rely on clause type information to figure out speech acts, they could have access to additional information that is unrelated to clause typing, but informative for recognizing speech act type. For example, in conversations, we use questions to elicit responses and information, which leads to behaviors like pausing after questions, or looking directly at our interlocutor, to nudge them

to answer our questions. Thus, armed with an innate category for questions, and a theory of what questions do in conversations, the child can expect certain kinds of nonlinguistic behavior to be correlated with the act of asking a question. If these behaviors are also correlated with the act of asking a question in the input, then they can tell whether their parent is asking them a question.

But just as surface formal features might be absent or misleading in the input, the social pragmatic features might also be absent or misleading. Therefore, I ask in this chapter whether it is possible to see some social pragmatic and prosodic features correlated with speech acts. I focus on the questioning act here, as this speech act is relatively well understood, and plan to extend to requests/commands in the future.

In conversations, questions are devices for us to engage with an addressee. When we ask questions, we are looking for information, or sometimes simply for a response. This makes questions perfect turn-transition points (Duncan 1972 among many others), signaling that the current speaker is done with the turn and the next speaker needs to get ready. This makes the use of questions to be correlated with communicative signals like direct eye gaze: speakers gaze at their interlocutor longer after questions (Argyle 1972). When we are looking at our interlocutor for an answer, we might have certain facial expressions. For example, Domaneschi, Passarelli, & Chiorri (2017) find people tend to be associated with the facial expressions pictured in Figure 6.1 with questions. If children are equipped with the knowledge about facial expressions, they might be able to exploit the correlation between questioning and facial expressions to learn about questions.



Figure 6.1: Upper face action units (AU) identified by participants as indicating the act of questioning, as reported by Domaneschi, Passarelli, & Chiorri (2017)

Another similar conversational signal might be the time in-between conversational turns. A property related to a turn-transition point is that interlocutors use these points as cues for when to take over the turn. In a natural face-to-face conversation, the gap between a question and its answer might be smaller than other pairs of utterances in a conversation (Tice & Henetz 2011 a.o.). But the questioner might also pause and wait if nobody steps in immediately (Sacks, Schegloff, & Jefferson 1978). Therefore, tracking the gap after utterances might tell us whether the utterance is a question.

So if children understand the mechanism of human communication, which as we have in Chapter 2 that they do, and have the prior knowledge that certain speech acts expects responses, they might expect to see a correlation between questions and the length of eye gaze, the facial expressions of their interlocutor, and the length of pauses. They then can exploit this correlation to learn questions.

We express our desire to know with questions (e.g. Searle 1975a, Krifka 2011, Carruthers 2018). As a result, when we ask questions, we might be ignorant of certain information. If children can keep track of who knows what, and as we saw in Chapter 2 that they can, they can exploit this correlation between lack of information with questions to infer questionhood. However, this correlation might not show up in the input, as parents tend to use questions as a tool for teaching. A parent asking “What’s that?” is likely not to solicit some information they don’t know, but to see if the child knows it too. Previous corpus study suggests

that about 38% of parents' questions when the child is before 2 years old are pedagogical questions (Yu, Bonawitz, & Shafto 2019), and for middle class parents, this number could reach 60%. The child might still be able to exploit the cue in some way, but it might not be as straightforward.

Due to the limitation of our corpus data, in this chapter we will examine the following features: duration of pauses between utterances and eye gaze. This chapter is organized as follows: Section 6.1 reviews previous evidence for children's perception of these cues. I then report results from a corpus study with Providence Corpus in Section 6.2, where I looked at the length of pauses between parents' utterances (Section 6.2.3.1) and parents' attentional behavior (6.2.3.2). Section 6.3 concludes the chapter.

6.1 Background

6.1.1 Pauses in conversation

As mentioned earlier, questions are turn-transition points (Duncan 1972 among others), and this means that speakers anticipate a change of turn after questions. In adult-to-adult conversations, this property translates to shorter speech gaps after questions in one-on-one in-person conversations (Stivers 2010, Enfield, Stivers, & Levinson 2010, Hilbrink et al. 2013 among many others). Speakers might also wait after questions for responses (Sacks, Schegloff, & Jefferson 1978), but for adults, longer inter-turn silence leads to feelings of unease and tend to be avoided (Roberts, Francis, & Morgan 2006 among others). But when interacting with pre-linguistic infants, this silence might be helpful rather than awkward: if children can track the length of pause after an utterance, they might be able to infer that longer pauses are associated with questions and request. Are children sensitive to the turn-taking properties of conversations?

As early as 3 months old, infants show sensitivity to the structure of conversation. The overlap between infant vocalization and parents' speech decreases overtime, and the average

gap between parent utterance and infant vocalization decreases, suggesting that they respect the turn-taking rules (Hilbrink, Gattis, & Levinson 2013). When they start producing one-word utterances, slow down at trying to change turn, but quickly pick up the speed again as their linguistic ability develops (Hilbrink et al. 2013).

Studies also show that same as adults, children interpret pauses as meaningful. Craig & Gallagher (1983) investigate whether or not 22-36-month-olds abandon their request (e.g. wanting a cookie) after the parent's initial responses (classified as neutral, e.g. *you want what?*, negative, *No.* or pause). They found that the older children (3-year-olds) abandon requests after adult pauses for more than 1s, suggesting that children treat long pauses as a meaningful response. This is suggestive, but based on this research alone it remains unclear whether children associate inter-turn silence with the interlocutor trying to elicit a response.

On the other side the problem, the length of pauses results from the dynamics of conversation; if parents are interacting with a pre-linguistic infant, would they follow the same rules? In other words, if we assume that children can keep track of the length of pauses, are there any patterns found in the silences? Previous studies show that parents seem to treat even pre-linguistic infants as competent conversationalists, and any vocalizations (e.g. crying, babbling) or even body movements, are treated as a conversational turn (Beebe et al. 1988, Jaffe et al. 2001). The length of pauses also correlates with the vocabulary of children (Marklund et al. 2015). Different from adult-to-adult interactions, parents of 14-month-olds tend to follow questions with another question (Reimchen & Soderstrom 2017).

While these studies show that parents respect turn-taking rules when interacting with pre-verbal infants, little is known about whether parents' speech gaps are informative of the speech act performed. To address this problem, Section 6.2.3.1 examines the correlation between the speech act of an utterance and the length of pauses after an utterance, to see if the gap between utterances is informative of an utterance's speech acts.

6.1.2 Attentional behavior

Another consequence of the response-expectation property of questions is that by the end of a question, the speaker tends to appoint the next speaker. A common device for turn allocation is eye gaze (Argyle 1972, Kendon 1967, Duncan, Brunner, & Fiske 1979, Rossano, Brown, & Levinson 2009, Csibra 2010). In dyadic interactions, we tend to look at our interlocutor when we need their response. Thus, it is possible that by tracking where the parent attends to, specifically whether the parent is directly looking at the learner, the learner might be able to infer whether a question is being asked.

We have seen in Chapter 2.2.1 that infants are sensitive to the direction of eye gaze since birth. 3-day-old newborns prefer to look at the face that appears to make eye contact with them, suggesting that they are sensitive to the position of the pupils/irises within the eye (Farroni et al. 2004).

While these studies show that infants can perceive parents' direction of eye gaze, and that adults use eye gaze for turn allocation, so far little is known about whether parents' gaze pattern correlates with turn allocation, and furthermore, with the speech act of an utterance. If questions signal a transition of turn, parents' gaze should fall on the addressee—the child—more after questions. I address this question in Section 6.2.3.1 by examining parents gaze pattern during parent-child interactions, to see if this cue is informative of an utterance's speech acts.

6.1.3 Interim Summary

When interacting with other people, we use many cues, besides the clause type information, to infer what kind of speech act is being performed. In particular, since questions usually signal the transition of a conversational turn, the questioner usually shifts gaze to the next designated speaker. Adults can also infer from a question being performed that the questioner wants someone to pick up the turn, and thus the lengths of speech gap after questions are normally longer than lengths of pause after other speech acts. However, if we look at dyadic

interactions between a parent and a child, especially a child as young as 18 months old, these cues might not show up, or show up in a different way. We should ask, for example, whether parents in child-directed interaction also use direct eye gaze to indicate the next speaker, i.e. the child? Also, since 18-month-olds are not mature conversationalists, the speech gap after utterances might be longer across the board, so are there any differences in pauses that correlate with the speech act being performed? Another potential cue is prosody. In English, yes/no questions are usually associated with a special final rise contour ($L^* H-H\%$), which can be produced with either polar interrogatives and declaratives syntax. Would final rises be a cue that distinguishes between the three speech acts?

To address these questions, we built a multi-modal dataset with dyadic parent-child interactions for infants before 18 months old, to see if there is any non-clause type cues that may help infants distinguish speech acts.

6.2 Corpus study

This section details the corpus study we conducted to investigate whether speech gap, direct gaze, and prosody in dyadic parent-child interactions correlate with parents' use of different speech acts.

6.2.1 Methods

This study also used data from the Providence Corpus (Demuth, Culbertson, & Alter 2006) from CHILDES system (MacWhinney 2000). The audio and video of the sessions sampled in Chapter 3 and Chapter 5 were extracted for annotation.

For annotating gaze, we used the video data available in the Providence corpus. Trained annotators viewed the muted videos using ELAN (Lausberg & Sloetjes 2009). Each video was annotated first with whether the parent and the child are visible on screen and the parent's focus of attention can be identified. If the parent is only half on screen, but we can

still identify the focus of attention of the parent, this proportion was counted as on-screen. If the parent is on screen but the focus cannot be identified due to bad lighting, this proportion of the video was counted as off-screen.

Then, for the parts of the video that both the parent and the child are both on screen, the focus of parents' attention was annotated. Specifically, whether the parent is paying attention to the child, to an object, or unidentifiable. The parent was annotated as paying attention to the child when they attended to the child through (1) direct eye gaze (when the eyes could be seen), (2) head or body orientation toward the child (when the eyes could not be seen), (3) physical interaction with the child. I then used a script to calculate the proportion of looks to the child during the utterance, and at 1s, 2s, 3s before and after the utterance.

6.2.2 The correlations we expect

If speech gaps and gaze are indeed useful cues to obtain speech act information, we would expect the following predictions to bear out.

For speech gaps, we should see that parents pause longer after questions and requests than after assertions, to wait for responses. And for eye gaze, we should see that parents direct attention to the child more often after questions and requests than after assertions.

6.2.3 Results

6.2.3.1 Pauses

For this cue, we measured the length of silence between parents' consecutive utterances like (81). To do so, we first followed the methodology of Bloom, Rocissano, & Hood (1976) and annotated for each utterance whether it is a contingent utterance of the previous one, i.e. whether it shares the same topic and adds new information relative to the topic of the prior utterance. We will refer to such pairs of utterances, with one contingent on the previous

utterance like in (81), as consecutive turn sequences. Within the sequences, I measured the length of pauses between utterances.

(81) **Consecutive turn sequence**

a. You can't take your rake on the swing.

PAUSE: 0.014S

You wanna take your big bird rake on the swing?

ASSERTION -QUESTION

b. You don't wanna swing?

PAUSE: PAUSE: 0.23S

You don't have to.

QUESTION-ASSERTION

c. Here you use this one.

PAUSE: 0.033S

This one works better.

ASSERTION-ASSERTION

In total, 4066 utterances were found to have a contingent follow-up that allows us to measure the inter-turn silence. Figure 6.2 shows the length of pause after each speech act category.

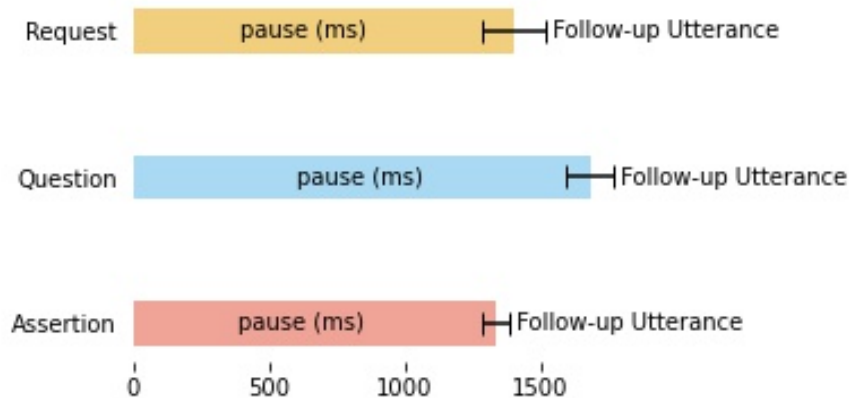


Figure 6.2: Duration of pause (ms) after each type of speech act

These results show that parents are equally likely to ask another question as they are to answer their own questions, but parents tend to pause longer after questions (mean = 1728ms) than assertions (mean = 1321ms; $t(2925) = -2.23, p = 0.02$), suggesting that parents pause after asking a question but proceed with the conversation after an assertion.

6.2.3.2 Parents' attentional behavior

In total, 857 utterances were found to have attentional data suitable for annotation. Figure 6.3 shows the proportion of looks to the child before, during, and after an utterance.

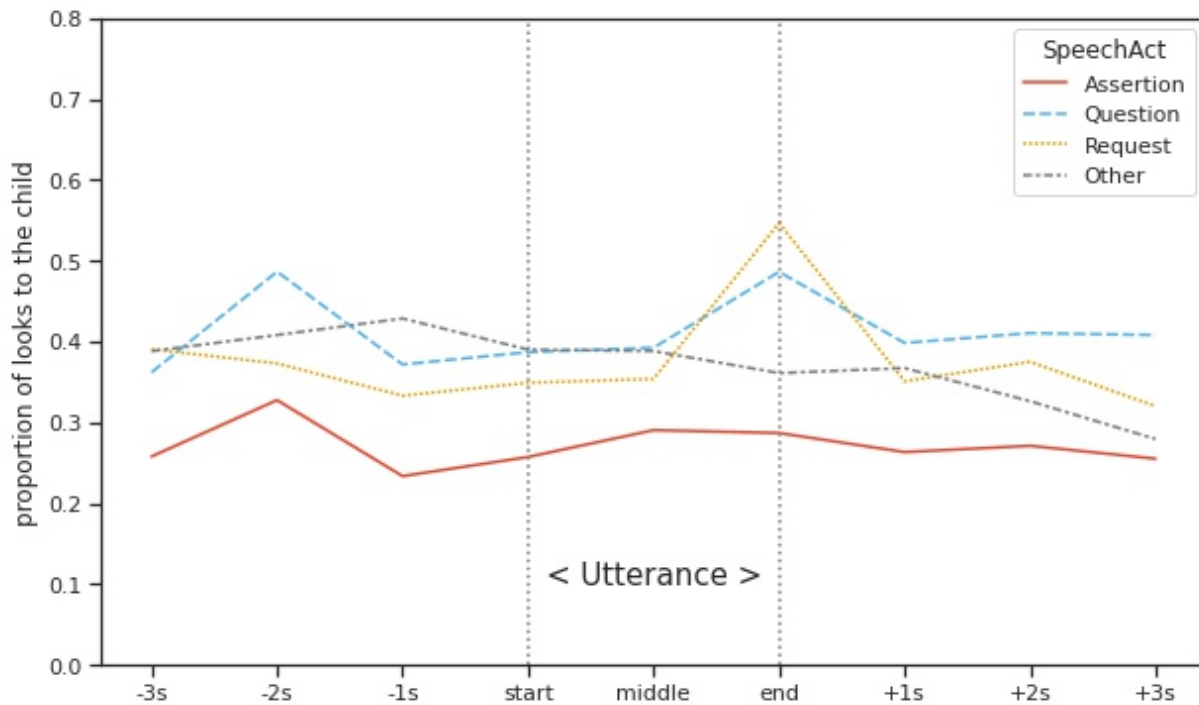


Figure 6.3: Proportion of parents' looks to the child before, during, and after an utterance

From Fig. 6.3, we can see that when the utterance is a question, the proportion of looks to the child is higher (average 0.41) than when it is an assertion (average 0.27; $t(16) = 2.53, p < 0.05$), but roughly equivalent to when the utterance is a request (0.37; $t(16) = 1.27, p = 0.21$). These results suggest that parents look at the child longer after questions and requests.

6.3 Conclusion

This chapter investigated the question of how children might be able to pick out speech acts independently of clause types or prosodic information. As questions are often used to elicit responses and information, we can expect different kinds of behavior to be correlated with these acts. Armed with an innate category for questions, and a theory of what questions do in conversations, the child can expect certain kinds of non-clause type cues to be correlated with the act of asking a question.

Some candidates for cues that could potentially differentiate questions from other speech acts are pauses and direct eye gaze. The canonical function of questions is to solicit responses or seek information, so when we use questions, it is likely that we pause after questions, or look directly at our interlocutor, as a way to signal to them that they need to take over the conversational turn. As we have reviewed, children have a theory of communication early on, and understand the structure of conversations from as young as three months old (Casillas, Bobb, & Clark 2016). With this theory of communication, and the prior knowledge that certain speech act is used for response-elicitation, they may expect questions to be correlated to some degree with longer pauses and direct eye gaze. I found that these features are correlated with the act of questioning in the input: parents tend to pause longer after questions, and attend the child more when asking questions. To the extent that they are weakly correlated with the questioning act, it is in principle plausible that (a) a child could use these features, in addition to their growing knowledge of clause types to infer the speech act category of an utterance; and (b) this little bit of information about speech act could then be used to provide the 20% of pragmatic information that the child needs in order to get the clause type clusters identified accurately.

Of course pauses and eye gaze are not the only cues for speech acts. In the future, I will explore other cues like facial expressions and gestures to see how they correlate with the use of different speech acts.

Chapter 7: Conclusion and discussion

7.1 Summary of findings

In language after language, we find three clause types (declaratives, interrogatives, and imperatives) that are dedicated to three speech acts (assertions, questions, commands). By 18 months old, children seem to be able to differentiate these clause types and associate them with their canonical speech act. To gain this ability, they need to identify the right categories of clauses (i.e. solve the clustering problem) and figure out what speech act they are canonically used for (i.e. solve the labeling problem). To solve the labeling problem, some speech act information must be available to the learner, but since mismatches are possible – sentences used to perform speech acts that are not canonically associated with their clause types – it is not immediately clear how useful it is for learners to be able to identify speech acts, if it is useful at all. To solve the clustering problem, children surely need to pay attention to the surface morpho-syntactic features of each sentence in their input. But in the input that learners actually receive, many surface features might be absent or misleading. Again, it seems plausible that the availability of some information related to speech acts might be helpful to the learner, but this information too is potentially misleading, again due to cases where sentences are used to perform speech acts that are not canonically associated with their clause type.

This dissertation investigates how children figure out clause type categories. In particular, are the surface formal features of the sentences in the input sufficient for children to figure out the clustering of clause types? If not, is speech act information helpful for solving the clustering problem? And how might learners access such speech act information?

I addressed these questions computationally by simulating two learners, a syntactically informed distributional learner (SID), and a syntactically and pragmatically informed distributional learner (SPID). Both learners use the surface morpho-syntactic features of the input sentences to attempt to cluster sentences into three categories, i.e. to learn the clause type categories. But the SPID additionally has access to some information about which speech act is performed by a sentence. I found that in English, the SID model could identify interrogative clauses but could not identify the other two clauses, but in Mandarin, this model could not find any of the right categories. The SPID model in both languages outperforms the SID. In English, the SPID model can find all three clause types; in Mandarin, the model has problem with imperative clauses. These results suggest that pragmatics is helpful, indeed crucial, to solve the clustering problem.

Another potential source of information for clause typing is prosody. Indeed, final rises tend to be associated with polar interrogatives cross-linguistically. However, I find that, overall, parents do not use final rises more often with questions. When I incorporate prosodic information into the two models, I find that this information does not improve SID's performance. Speech act information is still necessary for learners to find the right clause type clustering. When we look at subcategories of interrogatives, polar interrogatives do have more final rises than declaratives and *wh*-interrogatives. A bootstrapping strategy relying on prosody to figure out clause typing would have to be incremental: learners would first come to associate a subset of interrogatives to questions, and then would have to rely on additional correlations in morpho-syntactic and/or pragmatic features to figure out the other interrogatives.

But if the speech act information is useful for clause type learning, how do children figure out speech act information? Given that the way we generally identify a speech act is via its clause type, there is a potentially vicious circle here – you need to identify a sentence's clause type to infer the speech act that is being performed, but you need to be able to infer speech act information to learn to identify clause types. How do learners avoid this circularity?

One way to break this circularity is to *not* think of the learning of speech acts and clause types as two processes that need to happen sequentially, but as a joint learning process. Children learn to identify speech act and clause type in tandem and mutually informative ways.

To get one step closer to understanding and evaluating this joint learning hypothesis, I first addressed the question of how much speech act information children need to identify clause types. With the SPID model, I simulated the learning of clause type with various degrees of noise in the speech act information. I showed that even if children can only perceive speech act information a small proportion of the time, they can still benefit from this information, as a noisy pragmatic percept is superior to no pragmatics at all.

I then explored what kind of non-clause type cues for speech act information are present in the input. Even if children must rely on clause type information to figure out the speech acts, they could have access to additional information that is unrelated to clause typing, but is informative for recognizing speech act type. When speakers perform speech acts, because of the conventional functions of these speech acts on the discourse, the performance might be associated with certain socio-pragmatic features. For example, because of questions' response-elicitation function, we would expect pauses after questions. With prior knowledge about the functions of communication, and expectations about what questions do, children might be able to use these socio-pragmatic features to figure out this speech act. If children have such expectations, could they find any useful socio-pragmatic cues in the input?

I explored two cues that could potentially differentiate questions from other speech acts: pauses and direct eye gaze. I found that parents tend to pause longer after questions, and attend to the child more when asking questions. Therefore it is in principle plausible that there are some socio-pragmatic features that children can use, in addition to their growing knowledge of clause types to infer the speech act category of an utterance. This little bit of information about speech act could then be used to provide the pragmatic boost that the child needs in order to get the clause type clusters identified accurately.

7.2 The pragmatic syntactic bootstrapping hypothesis

As we have discussed, sentential force is an abstract semantic category that is connected to a formal category, clause type, on the one hand, and to speech act on the other. How do they figure out the various sentential forces of their language, i.e., the forms that each force can take (clause types), and the speech acts that they are canonically associated with?

In Chapter 1, we have seen that when we put the learning of clause type and learning of speech act together, we seem to run into a chicken-and-egg problem. On the formal side, clause type is an abstract formal feature of a sentence related to a variety of surface forms, none of which is obligatorily present and many of which can occur in sentences with a different clause-type feature. While in principle the surface formal features could be sufficient, this dissertation showed that they in fact fall short. To compensate for this insufficiency, learners can use the speech act information – which is systematically related to clause types – to learn to identify clause types and which surface features are relevant for clause typing.

On the function side, speech act is also an abstract category. As we have seen in Chapter 6, speech act categories are related to a variety of human behaviors (e.g. pauses and eye gaze), none of which are obligatorily present (you don't *need* to pause after a question) and many of which can occur when performing other speech act categories. While I showed that the features are correlated with the use of questions, it is likely that the learners use the clause type information — which is systematically related to speech acts — to inform how they infer speech act types.

How can learners break this vicious cycle to figure out the various sentential forces? The hypos proposes that the learner can break the cycle by learning clause type and speech act jointly:

(82) **The pragmatic syntactic bootstrapping hypothesis:**

Children jointly learn clause types (and their associated sentential forces) and speech acts from observations of the socio-pragmatic, prosodic, and morpho-syntactic fea-

tures of input sentences. On the one hand, children learn to identify speech acts by exploiting the prosody of the utterance and the socio-pragmatic features of the utterance (such as the social function of the utterance and the social attentional behavior of the speaker); on the other hand, they learn to identify clause types and their associated sentential force by exploiting the morpho-syntactic features of the sentence. Crucially, however, learning to identify speech acts and learning to identify clause types and sentential forces are mutually informative: children could use speech act information to learn the makeup of clause types and their associated sentential forces, and use clause type information to learn the socio-pragmatics of speech acts.

By learning speech act and clause type in tandem and mutually informative ways, we no longer need to assume that learners has to learn one before moving on to the other. Learners can exploit socio-pragmatic features to indirectly identify clause types via identifying the speech act, and exploit the morpho-syntactic features to indirectly identify speech acts via identifying clause types.

In future work, I plan to test the feasibility of this hypothesis computationally by simulating the learning process with a Bayesian clustering model. The proposed model will (i) track the prosodic and socio-pragmatic features of each utterance and infer the speech act that would generate these features; (ii) using knowledge of prosodic and morpho-syntactic features of each sentence, the model will be able to identify the clause type and sentential force that would generate these features; and (iii) speech act on the one hand, and clause type and sentential force on the other, will be inferred jointly.

Building on this model, we could also probe more into the role of prosody. As discussed in Chapter 2, the role of prosody for clause typing might be complicated, which affects how learners might use prosody. There are two possible starting points: learners could either expect prosody (and in particular final rise) to be informative of clause types, especially given the fact that in some languages like Italian, polar interrogatives only differ from declaratives by prosody, or they might not. If not, they might still expect prosody to be indirectly related

to clause types by being correlated to speech acts. We can test these options by modifying how prosody is represented in the syntax-pragmatic bootstrapping model mentioned above: in one version, prosodic information depends on both clause type and speech act information, and we can compare this with when the prosodic information only depends on speech act. We can then simulate the learning of languages like Vata, where prosody is not informative of clause typing, and languages like Italian, where prosody is critical for clause typing.

In this thesis, I've assumed that learners focus on matrix clauses to figure out clause types, and learn to associate them with particular sentential forces. But the same clause types can also occur in embedded contexts, but crucially not with their own sentential force. How do learners eventually relate the two? Hacquard & Lidz (2018) introduce pragmatic syntactic bootstrapping hypothesis as a proposal for how children acquire the meaning of attitude verbs. Under their view, learners exploit parallels between syntax and pragmatic function to figure out attitude meanings: namely, the kinds of clause types these verbs embed (in particular whether these embedded clauses share features with matrix declaratives, e.g., *think*), and the kinds of indirect speech acts that these verbs tend to lend themselves to (e.g., indirect assertions for *think*). Thus, learners could first learn to identify clause types by focusing on matrix clauses and associating them with particular speech acts, and later, build on this knowledge to then acquire the meaning of verbs that embed these clauses.

7.3 Conclusion

This dissertation investigates when and how children figure out the main clause types of their language (declaratives, interrogatives, imperatives) and their sentential forces, and the speech acts they are canonically associated with (assertions, questions, requests). Infants as young as 18 months old seem to be able to differentiate these clause types and associate them with their canonical speech acts. To gain this ability, children need to identify the right categories of clauses (the “clustering problem”) and figure out what functions they

are canonically used for (the “labeling problem”). By comparing a learner with speech act information and one without speech act information, I found that morpho-syntactic and prosodic features are not enough for English learners to find right clause type clustering, and that speech act information is crucial. Similar patterns were found when simulating with Mandarin data. However, the learning of clause types does not require *perfect* speech act information, as noisy speech act information is better than no speech act information. Thus, it is in principle, possible that children learn clause type information while still trying to figure out the speech act information. Additionally, I found that there are socio-pragmatic cues associated with the use of the speech act of questioning. Based on these results, I propose that children learn to figure out clause types and their associated sentential force, and speech acts jointly with observations of morpho-syntactic, prosodic, and socio-pragmatic features: learners use prosodic and morpho-syntactic features, as well as their growing knowledge of speech act information to infer clause types and their related sentential force; meanwhile, they use prosodic and socio-pragmatic features, and their growing knowledge of clause typing to learn about speech acts.

Bibliography

- Aikhenvald, Alexandra Y. 2016. In Jan Nuyts & Johan Van Der Auwera (eds.), *The oxford handbook of modality and mood*, chap. Sentence types, 141–165. Oxford University Press.
- Akmajian, Adrian. 1984. Sentence types and the form-function fit. *Natural Language & Linguistic Theory*. 1–23.
- Ambridge, Ben, Caroline F Rowland, Anna L Theakston, & Michael Tomasello. 2006. Comparing different accounts of inversion errors in children’s non-subject wh-questions: ‘what experimental data can tell us?’ *Journal of Child Language* 33(3). 519–557.
- Argyle, M. 1972. Eye-contact and the direction of gaze. *The psychology of interpersonal behavior*. 301.
- Atlas, Jay David & Stephen Levinson. 1981. *It*-clefts, informativeness, and logical form. In P. Cole (ed.), *Radical pragmatics*, 1–62. New York: Academic Press.
- Austin, John Langshaw. 1975. *How to do things with words*. Oxford university press.
- Babineau, Mireille, Rushen Shi, & Anne Christophe. 2020. 14-month-olds exploit verbs’ syntactic contexts to build expectations about novel words. *Infancy* 25(5). 719–733.
- Bach, Kent & Robert M Harnish. 1979. *Linguistic communication and speech acts*. Cambridge: M.I.T. Press.
- Baldwin, Dare A, Jodie A Baird, Megan M Saylor, & M Angela Clark. 2001. Infants parse dynamic action. *Child development* 72(3). 708–717.
- Bartels, Christine. 1999. *The intonation of english statements and questions: a compositional interpretation*. Routledge.
- Bates, Elizabeth. 1976. *Language and context: the acquisition of pragmatics*. Academic Press.
- Bateson, MC. 1975. *Before speech: The beginning of interpersonal communication*. 63–77.

- Beebe, Beatrice, Diane Alson, Joseph Jaffe, Stanley Feldstein, & Cynthia Crown. 1988. Vocal congruence in mother-infant play. *Journal of Psycholinguistic Research* 17(3). 245–259.
- Behne, Tanya, Malinda Carpenter, Josep Call, & Michael Tomasello. 2005. Unwilling versus unable: infants' understanding of intentional action. *Developmental psychology* 41(2). 328.
- Behne, Tanya, Malinda Carpenter, & Michael Tomasello. 2005. One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental science* 8(6). 492–499.
- Behne, Tanya, Ulf Liszkowski, Malinda Carpenter, & Michael Tomasello. 2012. Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology* 30(3). 359–375.
- Berninger, Ginger & Catherine Garvey. 1981. Questions and the allocation, construction, and timing of turns in child discourse. *Journal of Psycholinguistic Research* 10(4). 375–402. <https://doi.org/10.1007/BF01067165>.
- Bloom, Lois, Lorraine Rocissano, & Lois Hood. 1976. Adult-child discourse: developmental interaction between information processing and linguistic knowledge. *Cognitive Psychology* 8(4). 521–552.
- Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9). 341–345.
- Bögels, Sara & Francisco Torreira. 2015. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics* 52. 46–57.
- Bolinger, Dwight. 1978. Intonation across languages. In J Greenberg (ed.), *Universals of human language: phonology*, vol. 2, 471–524. Palo Alto, CA: Stanford University Press.
- Bresnan, Joan & Jane Grimshaw. 1978. The syntax of free relatives in english. *Linguistic inquiry* 9(3). 331–391.
- Büring, Daniel. 2016. *Intonation and meaning*. Oxford University Press.
- Butler, Joseph, Sonia Frota, & Marina Vigário. 2012. Infants' perception of intonation: is it a statement or a question. In *Fifth european conference on tone and intonation, oxford, reino unido*. http://labfon.letras.ulisboa.pt/texts/TIE5_Infant_Disc_BF_V.pdf.
- Caponigro, Ivano. 2003. *Free not to ask: on the semantics of free relatives and wh-words cross-linguistically*. Citeseer dissertation.

- Carpenter, Malinda, Nameera Akhtar, & Michael Tomasello. 1998. Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior and development* 21(2). 315–330.
- Carruthers, Peter. 2018. Basic questions. *Mind & Language* 33(2). 130–147.
- Casillas, Marisa, Susan C Bobb, & Eve V Clark. 2016. Turn-taking, timing, and planning in early language acquisition. *Journal of child language* 43(6). 1310–1337.
- Casillas, Marisa & Michael Frank. 2017. The development of children’s ability to track and predict turn structure in conversation. *Journal of Memory and Language* 92. 234–253. DOI: <https://doi.org/10.1016/j.jml.2016.06.013>. <http://www.sciencedirect.com/science/article/pii/S0749596X16300596>.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chen-Main, Joan. 2005a. Characteristics of mandarin imperatives.
- Chen-Main, Joan. 2005b. Characteristics of Mandarin imperatives. *Georgetown University working papers in theoretical linguistics* 4.
- Cheng, Lisa Lai Shen. 1991. *On the typology of wh-questions*. MIT dissertation.
- Chiang, Cindy, Susan Geffen, & Toben H Mintz. 2018. Distinguishing questions and statements using sentence-initial prosodic cues. In.
- Chien, Yu-Chin & Barbara Lust. 1985. The concepts of topic and subject in first language acquisition of mandarin chinese. *Child Development*. 1359–1375.
- Chierchia, Gennaro & Sally McConnell-Ginet. 1990. *Meaning and grammar: an introduction to semantics (1st ed.)* Cambridge, Massachusetts: The MIT Press.
- Chierchia, Gennaro & Sally McConnell-Ginet. 2000. *Meaning and grammar: an introduction to semantics (2nd ed.)* Cambridge, Massachusetts: The MIT Press.
- Chomsky, Noam. 1957. Syntactic structures. In *Syntactic structures*. De Gruyter Mouton.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT press.
- Chomsky, Noam & Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8(3). 425–504. <http://www.jstor.org/stable/4177996> (16 May, 2022).
- Clark, Eve V. & Kate L. Lindsey. 2015. Turn-taking: a case study of early gesture and word use in answering WHERE and WHICH questions. *Frontiers in Psychology* 6.

890. DOI: [10.3389/fpsyg.2015.00890](https://doi.org/10.3389/fpsyg.2015.00890). <https://www.frontiersin.org/article/10.3389/fpsyg.2015.00890>.

- Craig, Holly K. & Tanya M. Gallagher. 1983. Adult-child discourse: the conversational relevance of pauses. *Journal of Pragmatics* 7(4). 347–360. DOI: [https://doi.org/10.1016/0378-2166\(83\)90023-1](https://doi.org/10.1016/0378-2166(83)90023-1). <http://www.sciencedirect.com/science/article/pii/0378216683900231>.
- Csibra, Gergely. 2010. Recognizing communicative intentions in infancy. *Mind & Language* 25(2). 141–168.
- Dehé, Nicole & Bettina Braun. 2020. The prosody of rhetorical questions in english. *English Language and Linguistics* 24(4). 607–635. DOI: [10.1017/S1360674319000157](https://doi.org/10.1017/S1360674319000157).
- Demuth, Katherine, Jennifer Culbertson, & Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech* 49(2). 137–173.
- Deng, Xiangjun & Virginia Yip. 2018. A multimedia corpus of child mandarin: the tong corpus. *Journal of Chinese Linguistics* 46(1). 69–92.
- Domaneschi, Filippo, Marcello Passarelli, & Carlo Chiorri. 2017. Facial expressions and speech acts: experimental evidences on the role of the upper face as an illocutionary force indicating device in language comprehension. *Cognitive processing* 18(3). 285–306.
- Dong, Hongyuan. 2009. *Issues in the Semantics of Mandarin Questions*. Cornell University dissertation.
- Duncan, Starkey. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology* 23(2). 283.
- Duncan, Starkey, Lawrence J Brunner, & Donald W Fiske. 1979. Strategy signals in face-to-face interaction. *Journal of Personality and Social Psychology* 37(2). 301.
- Echols, Catharine H & C Nathan Marti. 2004. Weaving a lexicon. In D. G. Hall & S. R. Waxman (eds.), chap. The identification of words and their meanings: From perceptual biases to language-specific, 41–78. Cambridge, MA: MIT Press.
- Enfield, Nick J, Tanya Stivers, & Stephen C Levinson. 2010. Question-response sequences in conversation across ten languages: an introduction.
- Erreich, Anne. 1984. Learning how to ask: patterns of inversion in yes–no and wh-questions. *Journal of Child Language* 11(3). 579–592.

- Ervin-Tripp, Susan. 1978. Some features of early child-adult dialogues. *Language in Society* 7(3). 357–373. <http://www.jstor.org/stable/4167022>.
- Ettinger, Allyson & Sophia A. Malamud. 2014. Mandarin utterance-final particle *ba* in the conversational scoreboard. In Eva Csipak & Hedde Zeijlstra (eds.), *Proceedings of sinn und bedeutung 19*.
- Fahn, Rueih-Lirng Sharon. 2003. Chinese-speaking children's production of wh-questions. *Concentric: Studies in Linguistics* 29(2). 82–117.
- Fan, Li. 2007. Ertong ddui putonghua zhong foudingci de zaoqi huode [The acquisition of negation by Mandarin-speaking children]. *Xiandai waiyu* 30. 11.
- Fan, Li. 2012. Ertong zaoqi yuyan zhong yiwenci de yiwen he feiyiwen yongfa [The interrogative and non-interrogative use of Mandarin wh-words in children at early age]. *TSCOL Studies* 45. 87–97.
- Fang, Liu. 2009. *Intonation systems of Mandarin and English: A functional approach*. The University of Chicago dissertation.
- Farkas, Donka F & Floris Roelofsen. 2017. Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics* 34(2). 237–289.
- Farroni, Teresa, Stefano Massaccesi, Donatella Pividori, & Mark H Johnson. 2004. Gaze following in newborns. *Infancy* 5(1). 39–60.
- Feiman, Roman, Susan Carey, & Fiery Cushman. 2015. Infants' representations of others' goals: representing approach over avoidance. *Cognition* 136. 204–214.
- Feldman, Naomi H, Thomas L Griffiths, Sharon Goldwater, & James L Morgan. 2013. A role for the developing lexicon in phonetic category acquisition. *Psychological review* 120(4). 751.
- Flax, Judy, Margaret Lahey, Katherine Harris, & Arthur Boothroyd. 1991. Relations between prosodic variables and communicative functions. *Journal of child language* 18(1). 3–19.
- Frege, Gottlob. 1918. Logische untersuchungen. erster teil: der gedanke. *Beiträge zur Philosophie der deutschen Idealismus* 2. 58–77.
- Frota, Sonia. 2002. Nuclear falls and rises in european portuguese: a phonological analysis of declarative and question intonation. 14(1). 113–146. DOI: [doi:10.1515/prbs.2002.001](https://doi.org/10.1515/prbs.2002.001). <https://doi.org/10.1515/prbs.2002.001>.

- Frota, Sonia, Joseph Butler, & Marina Vigário. 2014. Infants' perception of intonation: is it a statement or a question? *Infancy* 19(2). 194–213. DOI: <https://doi.org/10.1111/infa.12037>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/infa.12037>.
- Gagliardi, Annie, Naomi H Feldman, & Jeffrey Lidz. 2017. Modeling statistical insensitivity: sources of suboptimal behavior. *Cognitive science* 41(1). 188–217.
- Gagliardi, Annie, Tara M Mease, & Jeffrey Lidz. 2016. Discontinuous development in the acquisition of filler-gap dependencies: evidence from 15-and 20-month-olds. *Language Acquisition* 23(3). 234–260.
- Gazdar, Gerald. 1981. Speech act assignment.
- Geffen, Susan & Toben H Mintz. 2011. Seven-month-olds discrimination of statements and questions. In *Proceedings of the 36th boston university conference on language development*.
- Geffen, Susan & Toben H Mintz. 2015. Can you believe it? 12-month-olds use word order to distinguish between declaratives and polar interrogatives. *Language Learning and Development* 11(3). 270–284.
- Geffen, Susan & Toben H Mintz. 2017. Prosodic differences between declaratives and interrogatives in infant-directed speech. *Journal of Child Language* 44(4). 968–994. DOI: [10.1017/S0305000916000349](https://doi.org/10.1017/S0305000916000349).
- Geman, Stuart & Donald Geman. 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* (6). 721–741.
- Genzel, Susanne & Frank Kügler. 2020. Production and perception of question prosody in akan. *Journal of the International Phonetic Association* 50(1). 61–92.
- Ginzburg, Jonathan. 1995. Resolving questions, ii. *Linguistics and Philosophy* 18(6). 567–609.
- Gleitman, Lila. 1990. The structural sources of verb meanings. *Language Acquisition* 1(1). 3–55.
- Goodhue, Daniel. 2021. A unified account of inquisitive and assertive rising declaratives. *Proceedings of the Linguistic Society of America* 6(1). 951–965.
- Goodhue, Daniel, Valentine Hacquard, & Jeffrey Lidz. 2022. Knowledge of clause type-speech act mappings in preschoolers and infants. University of Maryland.

- Goodhue, Daniel, Lyana Harrison, YT Clémentine Su, & Michael Wagner. 2016. Toward a bestiary of english intonational contours. *The Proceedings of the North East Linguistics Society (NELS)* 46. 311–320.
- Gordon, Matthew. 1999. The intonational structure of chickasaw. In *Proceedings of the 14th international congress of phonetic sciences*, vol. 3, 1993–1996.
- Gredebäck, Gustaf, Carolin Theuring, Petra Hauf, & Ben Kenward. 2008. The microstructure of infants' gaze as they view adult shifts in overt attention. *Infancy* 13(5). 533–543.
- Grice, Paul H. 1957. Meaning. *The philosophical review* 66(3). 377–388.
- Gryllia, Stella, Jenny S Doetjes, Yang Yang, & Lisa L Cheng. 2020. Prosody, clause typing, and wh-in-situ: evidence from Mandarin. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1).
- Gunlogson, Christine. 2003. *True to form: rising and falling declaratives as questions in English*. New York: Routledge.
- Gunlogson, Christine. 2008. A question of commitment. *Belgian Journal of Linguistics* 22(1). 101–136.
- Gussenhoven, Carlos. 2000. The boundary tones are coming: on the nonperipheral realization of boundary tones. *Papers in laboratory phonology V: Acquisition and the lexicon*. 132–151.
- Gussenhoven, Carlos. 2002. Why questions intonations rise and why they sometimes fall. In *Questions' conference*. university of liverpool press.
- Gussenhoven, Carlos & Aojun Chen. 2000. Universal and language-specific effects in the perception of question intonation. In *6th international conference on spoken language processing (icslp 2000)*, 91–94.
- Hacquard, Valentine & Jeffrey Lidz. 2018. Children's attitude problems: bootstrapping verb meaning from syntax and pragmatics. *Mind and Language*. 1–24.
- Hamblin, C. L. 1971. Mathematical models of dialogue1. *Theoria* 37(2). 130–155. DOI: [10.1111/j.1755-2567.1971.tb00065.x](https://doi.org/10.1111/j.1755-2567.1971.tb00065.x).
- Han, Chung-Hye. 1998. *The structure and interpretation of imperatives: mood and force in universal grammar*. University of Pennsylvania dissertation.
- Han, Yang. 1995. A pragmatic analysis of the *ba* particle in Mandarin Chinese. *Journal of Chinese Linguistics* 23(2). 99–128.

- Hay, Jessica F., Ryan A. Cannistraci, & Qian Zhao. 2019. Mapping non-native pitch contours to meaning: perceptual and experiential factors. *Journal of Memory and Language* 105. 131–140. DOI: <https://doi.org/10.1016/j.jml.2018.12.004>. <http://www.sciencedirect.com/science/article/pii/S0749596X18301037>.
- He, Angela Xiaoxue & Jeffrey Lidz. 2017. Verb learning in 14-and 18-month-old english-learning infants. *Language Learning and Development* 13(3). 335–356.
- Hedberg, Nancy, Juan M Sosa, & Lorna Fadden. 2004. Meanings and configurations of questions in english. In *Speech prosody 2004, international conference*.
- Hedberg, Nancy, Juan M. Sosa, & Emrah Görgülü. 2014. The meaning of intonation in yes-no questions in American English: A corpus study. *Corpus Linguistics and Linguistic Theory* 13(2). 321–368. DOI: <https://doi.org/10.1515/cllt-2014-0020>. <https://www.degruyter.com/view/journals/cllt/13/2/article-p321.xml>.
- Hilbrink, Elma E, Merideth Gattis, & Stephen C Levinson. 2013. Turn-taking and its timing in infancy: a longitudinal study at 3-, 4-and 5-months. In *Child language seminar*.
- Hilbrink, Elma E, Merideth Gattis, & Stephen C Levinson. 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in psychology* 6. 1492. <https://doi.org/10.3389/fpsyg.2015.01492>.
- Hilbrink, Elma E, Merideth Gattis, Elena Sakkalou, Kate Ellis-Davies, & Stephen C Levinson. 2013. Development of turn-taking during infancy: does the infant contribute? In *The 5th joint action meeting*.
- Ho, Aichen T. 1977. Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative. *Phonetica* 34(6). 446–457.
- Holzman, Mathilda. 1972. The use of interrogative forms in the verbal interaction of three mothers and their children. *Journal of Psycholinguistic Research* 1(4). 311–336.
- Hood, Bruce M, J Douglas Willen, & Jon Driver. 1998. Adult’s eyes trigger shifts of visual attention in human infants. *Psychological Science* 9(2). 131–134.
- Hu, Fang. 2002. A prosodic analysis of wh-words in Standard Chinese. *Speech Prosody* 2002. 403–406.
- Huang, James C-T. 1982. *Logical relations in Chinese and the theory of grammar*. MIT dissertation.
- Huang, James C-T, Audrey Y-H Li, & Yafei Li. 2009. *The syntax of Chinese*. Cambridge: Cambridge University Press.

- Huang, Nick, Aaron Steven White, Chia-Hsuan Liao, Valentine Hacquard, & Jeffrey Lidz. 2022. Syntactic bootstrapping attitude verbs despite impoverished morphosyntax. *Language Acquisition* 29(1). 27–53.
- Huang, Ruoyu Lexie, Paul Fletcher, Zhixiang Zhang, Weilan Liang, Virginia Marchman, & Twila Tardif. 2022. Early grammatical marking development in mandarin-speaking toddlers. *Developmental Psychology* 58(4). 631.
- Hubert, Lawrence & Phipps Arabie. 1985. Comparing partitions. *Journal of classification* 2(1). 193–218.
- Hyman, Larry M & Kemmonye C Monaka. 2011. Tonal and non-tonal intonation in shekagalagari. In *Prosodic categories: production, perception and comprehension*, 267–289. Springer.
- Jadoul, Yannick, Bill Thompson, & Bart de Boer. 2018. Introducing Parselmouth: a Python interface to Praat. *Journal of Phonetics* 71. 1–15. DOI: <https://doi.org/10.1016/j.wocn.2018.07.001>.
- Jaffe, Joseph, Beatrice Beebe, Stanley Feldstein, Cynthia L. Crown, Michael D. Jasnow, Philippe Rochat, & Daniel N. Stern. 2001. *Rhythms of dialogue in infancy: coordinated timing in development*. Vol. 66 (Monographs of the Society for Research in Child Development). Boston, MA: Published by: Wiley on behalf of the Research in Child Development.
- Jeong, Sunwoo. 2018. Intonation and sentence type conventions: two types of rising declaratives. *Journal of Semantics* 35(2). 305–356.
- Jiang, Ping & Aishu Chen. 2011. Representation of mandarin intonations: boundary tone revisited. In Zhuo Jing-Schmidt (ed.), *Proceedings of the 23rd north american conference on chinese linguistics (naccl-23)*, vol. 1, 97–109. Eugene: University of Oregon Press.
- Jingyao, Li. 2019. Hanyu ertong zaoqi fouding biaoda zai yuyi fanchou fangmian de xide [The acquisition of negation expressions by mandarin-speaking children]. *Jiaoxue fangfa chuangxin yu shijian* 2. 141–146.
- Johnson, Elizabeth K, Amanda Seidl, & Michael D Tyler. 2014. The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PloS one* 9(1). e83546.
- Keitel, Anne, Wolfgang Prinz, Angela D Friederici, Claes Von Hofsten, & Moritz M Daum. 2013. Perception of conversations: the importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology* 116(2). 264–277.

- Kendon, Adam. 1967. Some functions of gaze-direction in social interaction. *Acta psychologica* 26. 22–63.
- Kim, Yun Jung & Megha Sundara. 2021. 6-month-olds are sensitive to english morphology. *Developmental science* 24(4). e13089.
- König, Ekkehard. 2020. Clause types and speech act functions. In Bas Aarts, Jill Bowie, & Gergana Popova (eds.), *The oxford handbook of english grammar*. DOI: [10.1093/oxfordhb/9780198755104.013.20](https://doi.org/10.1093/oxfordhb/9780198755104.013.20).
- König, Ekkehard & Peter Siemund. 2007. Speech act distinctions in grammar. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 2, 276–324. Cambridge: Cambridge University Press.
- Kovács, Agnes Melinda, Tibor Tauzin, Ernő Téglás, György Gergely, & Gergely Csibra. 2014. Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy* 19(6). 543–557.
- Krifka, Manfred. 2011. Questions. In Klaus Von Heusinger, Claudia Maienborn, & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2, 1742–1785. Berlin: De Gruyter Mouton.
- Ladd, D Robert. 1981. On intonational universals. In *Advances in psychology*, vol. 7, 389–397. Elsevier.
- Ladd, D Robert. 2001. Intonational universals and intonational typology. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, & Wolfgang Raible (eds.), *Handbook on language typology and language universals*, 1380–1390. Berlin: De Gruyter.
- Ladd, D Robert. 2008. *Intonational phonology*. Cambridge University Press.
- Lammertink, Imme, Marisa Casillas, Titia Benders, Brechtje Post, & Paula Fikkert. 2015. Dutch and English toddlers’ use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology* 6. 495. DOI: [10.3389/fpsyg.2015.00495](https://doi.org/10.3389/fpsyg.2015.00495). <https://www.frontiersin.org/article/10.3389/fpsyg.2015.00495>.
- Landis, J Richard & Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics*. 159–174.
- Lausberg, H. & H Sloetjes. 2009. *Coding gestural behavior with the NEUROGES-ELAN system*. Version 6.0. Max Planck Institute for Psycholinguistics, Nijmegen. 841–849. DOI: [10.3758/BRM.41.3.841](https://doi.org/10.3758/BRM.41.3.841). <https://archive.mpi.nl/tla/elan>.
- Lee, Ok Joo. 2005. *The prosody of questions in Beijing Mandarin*. The Ohio State University dissertation.

- Lee, Thomas H-T. 1982. Ertong yuyan foudingci de fazhan [The development of negation in child language]. *Language Learning and Communication* 1(3). 233–344.
- Lee, Thomas H-T. 1989. Development of a Mandarin speaking child's comprehension of wh-questions. *Cahiers de Linguistique Asie Orientale* 18(1). 29–62.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Lewis, David. 1979. Scorekeeping in a language game. In *Semantics from different points of view*, 172–187. Springer.
- Li, Boya. 2006. *Chinese final particles and the syntax of the periphery*. Utrecht University dissertation.
- Li, Charles N. & Sandra A Thompson. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley & Los Angeles: University of California Press.
- Li, Hui, Mengguo Jing, & Eileen Chin Mei Wong. 2017. Predicting the development of interrogative forms and functions in early years: a corpus-based study of Mandarin-speaking young children. *Journal of Child Language* 44(1). 216–238. DOI: [10.1017/S0305000916000088](https://doi.org/10.1017/S0305000916000088).
- Li, Yuming & Qianduan Chen. 1997a. Ertong wenju lijie de qunan yu gean de bijiao yanjiu [comprehension of interrogatives by mandarin-speaking children: experimental and case studies]. *Yuyan Jiaoxue yu Yanjiu* 4. 119–130.
- Li, Yuming & Qianduan Chen. 1997b. Ertong wenju xitong lijie yu fasheng zhi bijiao [comprehension and production of interrogatives by mandarin-speaking children]. *Shijie Hanyu Jiaoxue* 4(1). 90–98.
- Li, Yuming & Zhidong Tang. 1991. *Hanzu ertong wenju xitong xide tanwei [acquisition of interrogatives by mandarin-speaking children]*. Shanghai: East China Normal University Press.
- Lidz, J., A. Omaki, & N Orita. 2012. Input in the lab causes 15-month-olds to learn an english non-adjacent dependency. In *Proceedings of the 37th boston university conference on child language*.
- Liebal, Kristin, Tanya Behne, Malinda Carpenter, & Michael Tomasello. 2009. Infants use shared experience to interpret pointing gestures. *Developmental science* 12(2). 264–271.
- Lin, Jing, Fred Weerman, Hedde Zeijlstra, et al. 2014. Mandarin shenme as a superweak npi.

- Liu, Fang & Yi Xu. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62(2-4). 70–87.
- Luchkina, Elena, David M. Sobel, & James L. Morgan. 2018. Eighteen-month-olds selectively generalize words from accurate speakers to novel contexts. *Developmental Science* 21(6). 1–11. <http://search.ebscohost.com.proxy-um.researchport.umd.edu/login.aspx?direct=true&db=asn&AN=132600863&site=ehost-live>.
- Ma, Joan K-Y, Valter Ciocca, & Tara L Whitehill. 2011. The perception of intonation questions and statements in cantonese. *The Journal of the Acoustical Society of America* 129(2). 1012–1023.
- MacWhinney, Brian. 2000. *The childe project: tools for analyzing talk. transcription format and programs*. Vol. 1. Psychology Press.
- Maria, Biezma & Rawlins Kyle. 2017. Rhetorical questions: severing questioning from asking. In *Proceedings of semantics and linguistic theory*, vol. 27, 302–22.
- Marklund, Ulrika, Ellen Marklund, Francisco Lacerda, & Iris-Corinna Schwarz. 2015. Pause and utterance duration in child-directed speech in relation to child vocabulary size. *Journal of child language* 42(5). 1158–1171.
- Menyuk, Paula & Nancy Bernholtz. 1969. Prosodic features and children’s language production. *MIT quarterly progress report* 93. 216–219.
- Miao, Xiaochun. 1986. Youer dui yinwenci de lijie [children’s comprehension of interrogative lexicon]. *Xinli Kexue Tongxun* 5. 1–5.
- Miao, Xiaochun & Manshu Zhu. 1992. Language development in. *Language processing in Chinese*. 237.
- Mintz, Toben H. 2006. Finding the verbs: distributional cues to categories available to young learners. *Action meets word: How children learn verbs* 31. 63.
- Mintz, Toben H. 2013. The segmentation of sub-lexical morphemes in english-learning 15-month-olds. *Frontiers in psychology* 4. 24.
- Moradlou, Sara, Xiaobei Zheng, Ye Tian, & Jonathan Ginzburg. 2020. Wh-questions are understood before polar-questions: evidence from English, German, and Chinese. *Journal of Child Language*. 1–27.
- Morissette, Paul, Marcelle Ricard, & Thérèse Gouin Décarie. 1995. Joint visual attention and pointing in infancy: a longitudinal study of comprehension. *British Journal of Developmental Psychology* 13(2). 163–175.

- Murray, Sarah E & William B Starr. 2018. Force and conversational states. *New work on speech acts*. 202–236.
- Murray, Sarah E & William B Starr. 2020. The structure of communicative acts. *Linguistics and Philosophy*. 1–50.
- Nazzi, Thierry, Isabelle Barrière, Louise Goyet, Sarah Kresh, & Géraldine Legendre. 2011. Tracking irregular morphophonological dependencies in natural language: evidence from the acquisition of subject-verb agreement in french. *Cognition* 120(1). 119–135.
- Nguyen, An & Colin Wilson. 2021. Learning the surface structure of wh-questions in english and french with a non-parametric bayesian model. In *Proceedings of the society for computation in linguistics 2021*, 403–405.
- Ni, Jinfu & Hisashi Kawai. 2004. Pitch targets anchor Chinese tone and intonation patterns. In *International conference on speech prosody*, 95–98.
- Ninio, Anat, Catherine E Snow, Barbara A Pan, & Pamela R Rollins. 1994. Classifying communicative acts in children’s interactions. *Journal of communication disorders* 27(2). 157–187.
- Ohala, John J. 1984. An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica* 41(1). 1–16.
- Onishi, Kristine H & Renée Baillargeon. 2005. Do 15-month-old infants understand false beliefs? *Science* 308(5719). 255–258.
- Orfitelli, Robyn & Nina Hyams. 2012. Children’s grammar of null subjects: evidence from comprehension. *Linguistic inquiry* 43(4). 563–590.
- Pearl, Lisa. 2020. Modeling syntactic acquisition.
- Perkins, Laurel. 2019. *How grammars grow: argument structure and the acquisition of non-basic syntax*. University of Maryland, College Park dissertation.
- Perkins, Laurel, Naomi H Feldman, & Jeffrey Lidz. 2022. The power of ignoring: filtering input for argument structure acquisition. *Cognitive Science* 46(1). e13080.
- Perkins, Laurel & Jeffrey Lidz. 2020. Filler-gap dependency comprehension at 15 months: the role of vocabulary. *Language Acquisition* 27(1). 98–115.
- Perkins, Laurel & Jeffrey Lidz. 2021. Eighteen-month-old infants represent nonlocal syntactic dependencies. *Proceedings of the National Academy of Sciences* 118(41).

- Perner, Josef & Johannes Roessler. 2012. From infants' to children's appreciation of belief. *Trends in cognitive sciences* 16(10). 519–525.
- Peterson-Hicks, J. 2006. The impact of function words on the processing and acquisition of syntax. *Unpublished dissertation, Northwestern University*.
- Phillips, Ann T, Henry M Wellman, & Elizabeth S Spelke. 2002. Infants' ability to connect gaze and emotional expression to intentional action. *Cognition* 85(1). 53–78.
- Pika, Simone & Katja Liebal. 2006. Differences and similarities between the natural gestural communication of the great apes and human children. In *The evolution of language*, 267–274. World Scientific.
- Pinker, Steven. 1984. *Language learnability and language development*. Boston, MA: Harvard University Press.
- Platzack, Christer & Inger Rosengren. 1997. On the subject of imperatives: a minimalist account of the imperative clause. *The Journal of Comparative Germanic Linguistics* 1(3). 177–224.
- Portner, Paul. 2004. The semantics of imperatives within a theory of clause types. In *Semantics and linguistic theory*, vol. 14, 235–252.
- Portner, Paul. 2018. *Mood*. Oxford: Oxford University Press.
- Povey, Daniel, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagenra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, & Karel Vesely. 2011. The kaldi speech recognition toolkit. In *Ieee 2011 workshop on automatic speech recognition and understanding*. IEEE Catalog No.: CFP11SRW-USB. Hilton Waikoloa Village, Big Island, Hawaii, US: IEEE Signal Processing Society.
- Povinelli, Daniel J, James E Reaux, Donna T Bierschwale, Ashley D Allain, & Bridgett B Simon. 1997. Exploitation of pointing as a referential gesture in young children, but not adolescent chimpanzees. *Cognitive Development* 12(4). 423–461.
- Prieto, Pilar & Gemma Rigau. 2007. The syntax-prosody interface: catalan interrogative sentences headed by que. *Journal of Portuguese Linguistics* 6(2).
- Rand, William M. 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association* 66(336). 846–850.
- Reimchen, Melissa & Melanie Soderstrom. 2017. Do questions get infants talking? infant vocal responses to questions and declaratives in maternal speech. *Infant and Child Development* 26(3). e1985 ICD-14-0159.R2, e1985. DOI: <https://doi.org/10.1002/icd.1985>.

<https://onlinelibrary.wiley.com/doi/abs/10.1002/icd.1985>.

- Repacholi, Betty M & Alison Gopnik. 1997. Early reasoning about desires: evidence from 14-and 18-month-olds. *Developmental psychology* 33(1). 12.
- Rialland, Annie. 2007. Question prosody: an african perspective. *Tones and tunes* 1. 35–64.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*, 281–337. Springer.
- Rizzi, Luigi. 2001. On the position “int (errogative)” in the left periphery of the clause. In *Current studies in italian syntax*, 287–296. Brill.
- Roberts, Craige. 1996. Information structure in discourse: toward an integrated formal theory of pragmatics. In Jae Hak Yoon & Andreas Kathol (eds.), *OSU working papers in linguistics 49: papers in formal semantics*, vol. 49, 91–136. The Ohio State University, Department of Linguistics.
- Roberts, Felicia, Alexander L Francis, & Melanie Morgan. 2006. The interaction of inter-turn silence with prosodic cues in listener perceptions of “trouble” in conversation. *Speech communication* 48(9). 1079–1093.
- Rossano, Federico, Penelope Brown, & Stephen C Levinson. 2009. Gaze, questioning and culture. *Conversation analysis: Comparative perspectives* 27. 187–249.
- Rowland, Caroline F, Julian M Pine, Elena VM Lieven, & Anna L Theakston. 2003. Determinants of acquisition order in *wh*-questions: re-evaluating the role of caregiver speech. *Journal of Child Language* 30(3). 609–635.
- Rudin, Deniz. 2018. *Rising above commitment*. UC Santa Cruz dissertation.
- Rudin, Deniz. 2019. Embedded rising declaratives and embedded quotation. In *Semantics and linguistic theory*, vol. 29, 1–21.
- Sacks, Harvey, Emanuel A Schegloff, & Gail Jefferson. 1978. A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*, 7–55. Elsevier.
- Sadock, Jerrold M & Arnold M Zwicky. 1985. Speech act distinctions in syntax. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 1, 155–196. Cambridge: Cambridge University Press.
- Sakkalou, Elena, Kate Ellis-Davies, Nia C Fowler, Elma E Hilbrink, & Merideth Gattis. 2013. Infants show stability of goal-directed imitation. *Journal of experimental child psychology* 114(1). 1–9.

- Santelmann, Lynn M & Peter W Jusczyk. 1998. Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition* 69(2). 105–134.
- Schulze, Cornelia & Michael Tomasello. 2015. 18-month-olds comprehend indirect communicative acts. *Cognition* 136. 91–98.
- Scott, Rose M & Renée Baillargeon. 2009. Which penguin is this? attributing false beliefs about object identity at 18 months. *Child development* 80(4). 1172–1196.
- Scott, Rose M & Renée Baillargeon. 2017. Early false-belief understanding. *Trends in Cognitive Sciences* 21(4). 237–249.
- Searle, John R. 1969. *Speech acts: an essay in the philosophy of language*. Vol. 626. Cambridge university press.
- Searle, John R. 1975a. A taxonomy of illocutionary acts. In *Language, mind and knowledge*, vol. 3, 344–369. Minneapolis: University of Minnesota Press.
- Searle, John R. 1975b. Syntax and Semantics. In Peter Cole & Jerry L. Morgan (eds.), vol. 3, chap. Indirect speech acts, 59–82. New York: Academic Press.
- Searle, John R. 1976. A classification of illocutionary acts. *Language in Society* 5(1). 1–23.
- Searle, John R. & Daniel Vanderveken. 1985. Speech Acts and Illocutionary Logic. In Daniel Vanderveken (ed.), *Logic, Thought and Action*, 109–132. Dordrecht: Springer Netherlands. DOI: [10.1007/1-4020-3167-X_5](https://doi.org/10.1007/1-4020-3167-X_5). https://doi.org/10.1007/1-4020-3167-X_5.
- Seidl, Amanda, George Hollich, & Peter W Jusczyk. 2003. Early understanding of subject and object wh-questions. *Infancy* 4(3). 423–436.
- Senju, Atsushi & Gergely Csibra. 2008. Gaze following in human infants depends on communicative signals. *Current biology* 18(9). 668–671.
- Shatz, Marilyn. 1978a. Children’s comprehension of their mothers’ question-directives. *Journal of Child Language* 5(1). 39–46.
- Shatz, Marilyn. 1978b. On the development of communicative understandings: an early strategy for interpreting and responding to messages. *Cognitive Psychology* 10(3). 271–301.
- Shatz, Marilyn. 1979. How to do things by asking: form-function pairings in mothers’ questions and their relation to children’s responses. *Child Development* 50(4). 1093–1099. <http://www.jstor.org/stable/1129336>.

- Shatz, Marilyn & Laura McCloskey. 1984. Answering appropriately: a developmental perspective on conversational knowledge. In *Discourse development*, 19–36. Springer. https://doi.org/10.1007/978-1-4613-9508-9_2.
- Shen, Xiaonan Susan. 1991. Question intonation in natural speech: a study of Changsha Chinese. *Journal of the International Phonetic Association* 21(1). 19–28.
- Shi, Rushen. 2014. Functional morphemes and early language acquisition. *Child Development Perspectives* 8(1). 6–11.
- Shi, Rushen, Elsa Santos, Jun Gao, & Aijun Li. 2017. Perception of similar and dissimilar lexical tones by non-tone-learning infants. *Infancy* 22(6). 790–800.
- Shi, Rushen, Janet F Werker, & James L Morgan. 1999. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72(2). B11–B21.
- Shipley, Elizabeth F, Carlota S Smith, & Lila R Gleitman. 1969. A study in the acquisition of language: Free responses to commands. *Language* (2). 322–342.
- Soderstrom, Melanie, Deborah G Kemler Nelson, & Peter W Jusczyk. 2005. Six-month-olds recognize clauses embedded in different passages of fluent speech. *Infant Behavior and Development* 28(1). 87–94.
- Soderstrom, Melanie, Kenneth Wexler, & Peter W Jusczyk. 2002. English-learning toddlers' sensitivity to agreement morphology in receptive grammar. In *Proceedings of the 26th boston university conference on child language*, vol. 2, 643–652. Somerville, MA: Cascadilla.
- Soderstrom, Melanie, Katherine S White, Erin Conwell, & James L Morgan. 2007. Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. *Infancy* 12(1). 1–29.
- Song, Hyun-joo & Renée Baillargeon. 2008. Infants' reasoning about others' false perceptions. *Developmental psychology* 44(6). 1789.
- Song, Hyun-joo, Kristine H Onishi, Renée Baillargeon, & Cynthia Fisher. 2008. Can an agent's false belief be corrected by an appropriate communication? psychological reasoning in 18-month-old infants. *Cognition* 109(3). 295–315.
- Stalnaker, Robert. 1978a. Assertion. In *Pragmatics*, 315–332. Brill.
- Stalnaker, Robert. 1978b. Assertion. In Peter Cole (ed.), *Syntax and semantics*, vol. 9, 315–332. New York: Academic Press.
- Stalnaker, Robert. 2002. Common ground. *Linguistics and philosophy* 25(5/6). 701–721.

- Starr, William B. 2014. Mood, force and truth. *ProtoSociology* 31. 160–181.
- Steffensen, Margaret S. 1978. Satisfying inquisitive adults: some simple methods of answering yes/no questions. *Journal of Child Language* 5(2). 221–236. DOI: [10.1017/S0305000900007431](https://doi.org/10.1017/S0305000900007431).
- Steinley, Douglas. 2004. Properties of the hubert-arable adjusted rand index. *Psychological methods* 9(3). 386.
- Stivers, Tanya. 2010. An overview of the question–response system in american english conversation. *Journal of Pragmatics* 42(10). 2772–2781.
- Stolcke, Andreas, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, & Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics* 26(3). 339–373.
- Stromswold, Karin. 1995. The acquisition of subject and object wh-questions. *Language Acquisition* 4(1-2). 5–48.
- Surian, Luca, Stefania Caldi, & Dan Sperber. 2007. Attribution of beliefs by 13-month-old infants. *Psychological science* 18(7). 580–586.
- Tamir, Lois. 1980. Interrogatives in dialogue: case study of mother and child 16–19 months. *Journal of Psycholinguistic Research* 9(4). 407–424.
- Tardif, Twila. 1993. *Adult-to-child speech and language acquisition in Mandarin Chinese*. Yale University dissertation.
- Tardif, Twila. 1996. Nouns are not always learned before verbs: evidence from mandarin speakers' early vocabularies. *Developmental psychology* 32(3). 492.
- Tice, Marisa & Tania Henetz. 2011. Turn-boundary projection: looking ahead. In *Proceedings of the annual meeting of the cognitive science society*, vol. 33.
- Truckenbrodt, Hubert. 2009. On rises and falls in interrogatives. *Proceedings from Interface, Discours, Prosodie (IDP)*. 9–17.
- Tyack, Dorothy & David Ingram. 1977. Children's production and comprehension of questions. *Journal of Child Language* 4(2). 211–224.
- Vaish, Amrisha, Özlem Ece Demir, & Dare Baldwin. 2011. Thirteen-and 18-month-old infants recognize when they need referential information. *Social Development* 20(3). 431–449.

- Vouloumanos, Athena, Alia Martin, & Kristine H Onishi. 2014. Do 6-month-olds understand that speech can communicate? *Developmental Science* 17(6). 872–879.
- Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: the pragmatics of fall-rise intonation. *Language*. 747–776.
- Warneken, Felix & Michael Tomasello. 2007. Helping and cooperation at 14 months of age. *Infancy* 11(3). 271–294.
- Wong, Wai Yi P, Marjorie K M Chan, & Mary E Beckman. 2005. An autosegmental-metrical analysis and prosodic annotation conventions for cantonese. *Prosodic typology: The phonology of intonation and phrasing*. 271.
- Woodward, Amanda L. 1998. Infants selectively encode the goal object of an actor’s reach. *Cognition* 69(1). 1–34.
- Xiao, Ling, Xiao Cai, & H-T Thomas Lee. 2006. *The development of the verb category and verb argument structure in children before two years of age*.
- Yang, Xiaolu, Rushen Shi, & Kailin Xu. 2018. Grammatical aspect in early child mandarin: evidence from a preferential looking experiment. *Journal of Psycholinguistic Research* 47(6). 1301–1320.
- Yang, Xiaolu & Cheng Yang. 2015. Control in mandarin-speaking children’s early naturalistic production. *Lingua* 163. 1–22.
- Yang, Yang. 2018. *The two sides of wh-indeterminates in Mandarin: a prosodic and processing account*. Leiden University dissertation.
- Yang, Yang, Stella Gryllia, & Lisa Lai-Shen Cheng. 2020. Wh-question or wh-declarative? Prosody makes the difference. *Speech Communication* 118. 21–32.
- Yang, Yu’an. 2021. Representing an issue as open: Mandarin discourse particle *ba*. In Joseph Rhyne, Kaelyn Lamp, Nicole Dreier, & Chloe Kwon (eds.), *Proceedings of the 30th Semantics and Linguistic Theory conference*, 523–541.
- Yang, Yu’an, Daniel Goodhue, Valentine Hacquard, & Jeffrey Lidz. 2022. Do children know *whanything?* acquisition of *wh*-ambiguity in Mandarin. *Language Acquisition*.
- Ying, Yuanfan, Xiaolu Yang, & Rushen Shi. 2021. Structures of function words guide Mandarin-learning 19-month-olds in backward syntactic categorization. *Proceedings of BUCLD 45*.
- Yoon, Jennifer MD, Mark H Johnson, & Gergely Csibra. 2008. Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*

105(36). 13690–13695.

- Yu, Yue, Elizabeth Bonawitz, & Patrick Shafto. 2019. Pedagogical questions in parent–child conversations. *Child Development* 90(1). 147–161.
- Yuan, Jiahong. 2006. Mechanisms of question intonation in mandarin. In *International symposium on chinese spoken language processing*, 19–30.
- Yuan, Jiahong. 2011. Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America* 130(6). 4063–4069.
- Yuan, Jiahong, Chilin Shih, & Greg Kochanski. 2002. Comparison of declarative and interrogative intonation in Chinese. *Speech Prosody 2002*.
- Zaitsu, Anissa, Jad Wahbe, Valentine Hacquard, & Jeffrey Lidz. 2020. Clause types and speech acts in speech to children. University of Maryland, College Park.
- Zeng, Xiao-Li, Philippe Martin, & Georges Boulakia. 2004. Tones and intonation in declarative and interrogative sentences in Mandarin. In *International symposium on tonal aspects of languages: with emphasis on tone languages*.
- Zhang, Zhao, Rushen Shi, & Aijun Li. 2014. Grammatical categorization of nouns and verbs in mandarin-learning infants. *Proceedings of BUCLD 38*.
- Zhang, Zhao, Rushen Shi, & Aijun Li. 2015. Grammatical categorization in mandarin-chinese-learning infants. *Language acquisition* 22(1). 104–115.
- Zhu, Dexi. 1982. *Yufa jiangyi [lectures on Chinese grammar]*. Commercial Press.
- Zwicky, Arnold M & Geoffrey K Pullum. 1983. Deleting named morphemes. *Lingua* 59(2-3). 155–175.