

# Automating the Corpus-based Study of the Acquisition of Pragmatic Markers

**Sandrine Zufferey**  
sandrine.zufferey@eti.unige.ch

**Andrei Popescu-Belis**  
andrei.popescu-belis@issco.unige.ch

University of Geneva, Switzerland  
School of Translation and Interpretation (ETI)

## Theory

### A. Theoretical background for the acquisition of pragmatic competencies

- most of the studies are speech act oriented
- in speech act terms, the first pragmatic competencies are acquired around age 2 (comprehension of simple acts), then acquisition continues until over 10 (e.g. 'promise')
- relatively little attention has been paid to the acquisition of pragmatic markers

### B. Factors in the acquisition of pragmatic competencies [Bek90]

- innate capacities
- child's participation
- role of the caregiver

### C. Use of pragmatic markers depends on:

- age of child
- presence of other persons (parents, peers) [KET99]
- type of activity (structured or not)
- geographical variation (dialect of English)

### D. Conclusions from the observations

- simple heuristics (filters) enable us to automate the extraction of the acquisition pattern
- the first uses of pragmatic markers appear between ages 2;6 and 3 ('well', 'I mean') but some markers are acquired only between ages 5 and 7 ('actually', 'basically')
- hypothesis: "Test New Words"
- towards a confirmation of the child's role

### Further work

- Study the acquisition of more complicated markers on older children ('since', 'therefore')
- Study the age of acquisition of the various pragmatic roles of the same marker

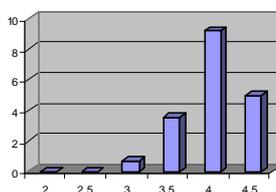
## Empirical study

### 1. Goals and Data

- study of the occurrence of various pragmatic markers. Only the results of 'well' are presented here.
- corpus: recordings of one child, from age 2;4 to 5;0. Available via CHILDES [Kuc76].
- data: 210 files containing each 30 minutes of spontaneous speech (interaction between child and parents).

### 2. Manual count

- distinguish pragmatic occurrences of 'well' from other ones
- compute frequency of 'well' depending on age
- group by 6-month periods (~26 hours of recording)



Occurrences of 'well' as a pragmatic marker, normalized per 30' session (vertical), depending on the age (horizontal).

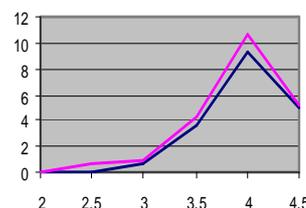
### 3. Automated count based on:

- position in the utterance (initial or not)
- patterns of collocation (e.g., 'well' not a pragmatic marker in 'as well', 'very well')

Compare

### 4. Validation and Results

Occurrences of 'well' as a pragmatic marker – difference between automatically detected (above, pink) and manually detected (below, blue) ones, depending on age (horizontal).



- filters detect a **constant proportion** of pragmatic uses
- characteristic acquisition pattern (↔ ↗ ↘)

### References

- [Bek90] Becker, J. *Processes in the acquisition of pragmatic competence*. In G. Conti-Ramsden & C. Snow: *Children's language*. Hillsdale: LEA, 1990, pp. 7-24.
- [KET99] Kyratzis, A. & S. Ervin-Tripp. The development of discourse markers in peer interaction. *Journal of pragmatics*. 1999, 31, pp. 1321-1338.
- [Kuc76] Kuczaj, S. *-ing, -s, and -ed: A study of the acquisition of certain verb inflections*. Unpublished doctoral dissertation, University of Minnesota, 1976.

## Automating the corpus-based study of the acquisition of pragmatic markers

Studies of the acquisition of pragmatic competencies have often difficulties in matching the stages hypothesized by theoretical analysis with the facts that can be derived from empirical data, such as recordings of children speech. In addition, it is often difficult to obtain quantitative results regarding these facts without manual counting of their occurrences in the data.

In this paper, we focus on a class of pragmatic competencies, namely pragmatic markers, and propose a method to automate their detection in children speech corpora. The method is validated by comparing manual and automated counts of occurrences of pragmatic markers. Finally, we discuss the factors that could explain the observed acquisition patterns, with emphasis on the child's own role in the acquisition process.

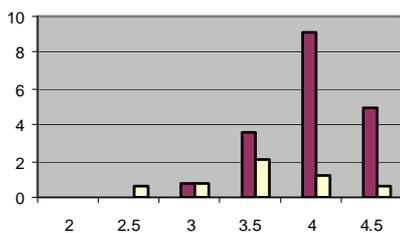
### 1. Approaches to the acquisition of pragmatic competencies.

Traditionally, the acquisition of pragmatic competencies has been studied from a speech-act point of view. However, more recent approaches (Papafragou 2000), have advocated relevance theory (Sperber & Wilson 1986) as a framework for acquisition studies. Our study focuses on the acquisition of pragmatic markers, which are not easy to account for in the speech-act framework. Moreover, pragmatic markers are used much earlier than phenomena related to speech act, such as indirect speech, thus bringing evidence for pragmatic competencies at earlier stages.

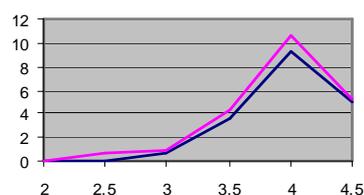
### 2. Towards automatic detection of pragmatic markers in the CHILDES corpus

Pragmatic markers in English (Andersen 2001:39) are devoid of semantic content, but serve significant pragmatic functions in conversation. However, most of the lexical items used as markers may also play non pragmatic roles in other contexts, which makes them difficult to detect automatically. We show however that a small number of heuristics help to identify pragmatic markers as such; moreover, the error margin when these heuristics are used alone is a constant proportion of the total number of pragmatic uses, so that automatic analysis *alone* allows us to draw conclusions about the emergence and evolution of such markers in children speech. This procedure compares favorably with attempts to automatically extract indirect speech – a particularly intractable problem.

Our work is based on excerpts from the CHILDES corpus. We use the CLAN tool (available with CHILDES) for manual extraction and a UNIX script for automated extraction. Two examples of results are shown in the figures below.



**Figure 1.** Occurrences of 'well' as pragmatic marker, depending on the age (*horizontal*), normalized per session (*vertical*).



**Figure 2.** Difference between automatically detected (*above*) and manually detected (*below*) occurrences of 'well' as a pragmatic marker, depending on age (*horizontal*).

### 3. Possible factors in the acquisition of pragmatic markers

Three main factors have been proposed to explain the acquisition of pragmatic markers: innate capacities, the child's participation, and of course the caretaker's influence (Becker 1990). Our study illustrates a phenomenon related to the child's role in the acquisition of pragmatics, namely the desire to "test" newly acquired markers: their acquisition curve peaks after the first occurrences, then slowly decreases. Conversely, the acquisition of new words contributes also to the decrease. Further studies should thus enable us to grasp more precisely the "test effect" and assess its precise role in the acquisition of pragmatic markers.