子どもに対する大人の助詞「ハ」の使用：言語獲得研究用コーパス CHILDES を用いた検討

橋本智也
大阪市立大学大学院
d05lc010@ex.media.osaka-cu.ac.jp

1 はじめに

助詞ハの基本的な特徴は，その直前に共起する語が文の主題であることを示すことである[1]．また助詞ハの直前に共起する語が主題として適切であるためには，話の流れ，発話場面の状況，常識などによって，聞き手または読み手がその語の指し示すものを特定できなければならない[2]．これまで助詞ハについて様々な記述や分析が行われているが，それらは大人の間での使用が対象であり，言語獲得初期の子どもに対する使用を対象としたものではない．言語獲得初期の子どもは大人と比べて語彙や知識が少なく認知発達も進んでいないため，大人であれば常識や場面などからわかることでも理解できない場合がある．そのため大人は助詞ハを使って子どもに話しかける際，指し示めされているものを特定しやすいように，助詞ハの直前に共起する語を選んでいる可能性がある．しかし，これまで子どもに対する大人の助詞ハの使用は検討されていない．そこで本稿では，言語獲得初期の子どもに対する大人の発話を対象に，助詞ハの直前に共起する語のカテゴリーと発話場面の詳細，そしてそれらの月齢による変化を検討する．

2 方法

2.1 CHILDES[3]

CHILDES(Child Language Data Exchange System；チャイルズ)は言語獲得研究のためのデータ共有を目的として Brian MacWhinney と Catherine Snow によって 1984 年に設立された．それまでの言語獲得研究では，研究者個人だけがデータを所有し，公開されることは少なかった．またデータの収集には時間がかかるため研究の数は増えにくく，さらに発話を文字に転写する方法が研究者によって様々であったため研究ごとの比較が難しかった．CHILDES では，そのような問題を解決するために，データがウェブ上で公開され，全言語に共通のデータフォーマットである CHAT(Codes for the Human Analysis of Transcripts；チャット)とデータ解析プログラムである CLAN(Child Language Analysis；クラン)が提供されている．発話データには約 30 ヶ国の言語が含まれ，日本語のデータは 6 コーパスが公開されている．現在では，発話を書き起こした文字情報だけではなく，音声や動画を含めて分析が行えるようになり，加えてウェブ上で分析が行える環境が整いつつあるなど，発展を続けている．

2.2 使用データ

本稿では，CHILDES に含まれる日本語データのひとつである Ishii コーパス[4]を使用した．Ishii コーパスを選択した理由は，他のコーパスよりも低い月齢の発話データが含まれているため，また動画データが公開されているので発話場面を確認しやすいためである．Ishii コーパスの主な発話者は男児 Jun(第三子)とその父親である．Jun が生まれたときに京都市から滋賀県草津市へ引っ越したが，家族は京都方言を話す．分析は Jun の 12，18，24，30，36 ヶ月齢での父親の発話を

対象とした．父親の 5 時点の総発話数は 2246，平均 449.2(*SD*=249.5)であった．

2.3 形態素の付与と曖昧性の解消

CLAN に含まれる，形態素付与プログラム MOR と曖昧性解消プログラム POST を使用した．MOR は文脈を考慮せずに可能な全ての品詞を付与するため，形態素を完全には決定することができず，複数の解析候補を併記する．そのため決定できないものについては曖昧性が残る．POST は CHILDES で公開されている各言語の語順のデータベースを利用して，その曖昧性を解消する．POST でも解消できない曖昧性は手動で決定した．最終的に，形態素の付与が正しく行われているかを全発話で確認した．以下にデータの例を示す．

*FAT: basu tooru ka ?
%mor: n|basu=bath^n|basu v:c|toor-PRES=pass ptl:coo|ka^ptl:conj|ka^ptl:fina|ka^n|ka =mosquito ?
%pos: n|basu v:c|toor-PRES=pass ptl:fina|ka ?

2.4 分析

まず各月齢における助詞ハの頻度および総発話数に対する割合を算出した．

次に助詞ハの直前に共起する語を「指示詞」,「幼児語」,「固有名詞・親族名称」,「普通名詞」の 4 カテゴリーに分類し，各カテゴリーの頻度の割合を算出した．指示詞は語の指し示すものが最も明確なカテゴリーである．幼児語は子どもにとって理解しやすいと大人が考えるカテゴリーである．固有名詞・親族名称は子どもにとって身近な存在である者を指し示すカテゴリーである．普通名詞は，これらの中で，子どもにとって語が指し示すものが最も理解しにくいと大人が考えるカテゴリーである．

またベースラインとして，助詞ハとの共起と無関係に各カテゴリーの頻度を求め，それらの総頻度に対する割合を各月齢において算出した．

最後に，発話場面を明らかにするため，助詞ハが使われている発話を「平叙文」と「疑問文」に分類し，その割合を算出した．

3 結果

3.1助詞ハの頻度および発話数に対する割合

12，18，24，30，36 ヶ月齢での助詞ハの頻度は，それぞれ 9，14，18，27，33 であり，発話数に対する割合はそれぞれ 4.5%，6.5%，4.3%，5.1%，3.7%であった．

3.2 助詞ハの直前に共起するカテゴリーの割合

助詞ハの直前に共起する語は, 全月齢を通して, 8 例[1]を除き 4 カテゴリーのいずれかであった．各カテゴリーの頻度の割合を月齢ごとに見ると，12 ヶ月齢で最も割合が高かったのが幼児語＋ハ(66.7%)であり，次いで指示詞＋ハと固有名詞・親族名称＋ハ(ともに 16.7%)であった．普通名詞＋ハは見られなかった．18 ヶ月齢以降は，全ての月齢において指示詞＋ハの割合が最も高かった(18, 24, 30, 36 ヶ月齢でそれぞれ 100%, 72.2%, 52.2%，72.7%)．結果を図 1 に示す．

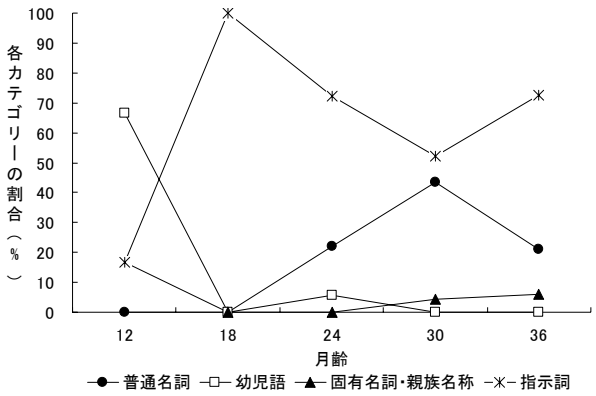

図 1: 助詞ハの直前に共起するカテゴリーの割合

3.3 各カテゴリーそのものの頻度の割合

12 ヶ月齢と 18 ヶ月齢において，助詞ハの共起

[1]形式助詞ノ 3 例，挨拶用語バイバイ 2 例，基数詞ロク 1 例，擬態語クルクル 1 例，格助詞デ 1 例.

と無関係に最も割合が高かったカテゴリーは，助詞ハとの共起で最も割合が高かったカテゴリーと同じであった. 12 ヶ月齢で最も割合が高かったのは幼児語(54.5%)であった．指示詞，固有名詞・親族名称，普通名詞の割合はそれぞれ 20.5%，18.2%，6.8%であった．18 ヶ月齢で最も割合が高かったのは指示詞(41.2%)であった．幼児語，固有名詞・親族名称，普通名詞の割合はそれぞれ 24.3%，14.7%，19.9%であった．したがって，12 ヶ月齢での幼児語＋ハと 18 ヶ月齢での指示詞＋ハの割合の高さは，それぞれのカテゴリーの頻度が高さから生じている可能性がある．

24 ヶ月齢以降では，助詞ハの共起と無関係に最も割合が高かったカテゴリーは全て普通名詞であり(24, 30, 36 ヶ月齢でそれぞれ 42.1%, 61.3%, 50.7%)，助詞ハと最も高い割合で共起するカテゴリー(指示詞)とは異なっていた. 指示詞, 幼児語, 固有名詞・親族名称の割合はそれぞれ，24 ヶ月齢では 28.6%, 23.2%, 6.2%, 30 ヶ月齢では 30.0%, 1.6%，7.0%，36 ヶ月齢では 31.0%，2.2%，16.1%であった．このことから，24 ヶ月齢以降で見られた，指示詞＋ハの高い割合は，指示詞そのものの割合が高いために生じているのではないと言える．結果を図 2 に示す.

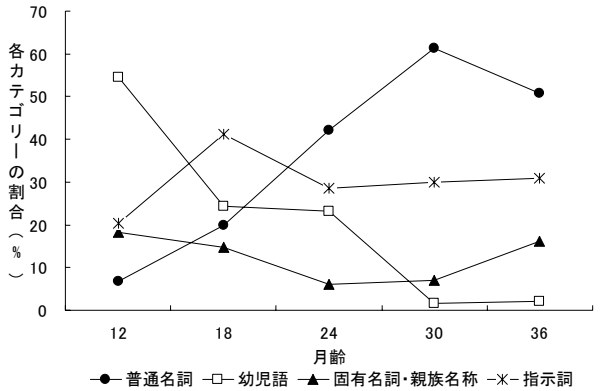

図 2：助詞ハの共起と無関係のカテゴリーの割合

3.4 「平叙文」と「疑問文」

全月齢で「疑問文」の割合が 50%よりも高かったが，月齢とともに「平叙文」が増加傾向を示した．12，18，24，30，36 月齢での「疑問文」の割合はそれぞれ, 88.9%, 92.9%, 66.7%, 51.9%, 60.6%であった．結果を図 3 に示す．また「疑問文」は助詞ハが発話の末尾に来る割合が高く, 12, 18, 24, 30, 36 月齢での割合はそれぞれ, 75.0%, 76.9%，91.7%，57.1%，70.0%であった.

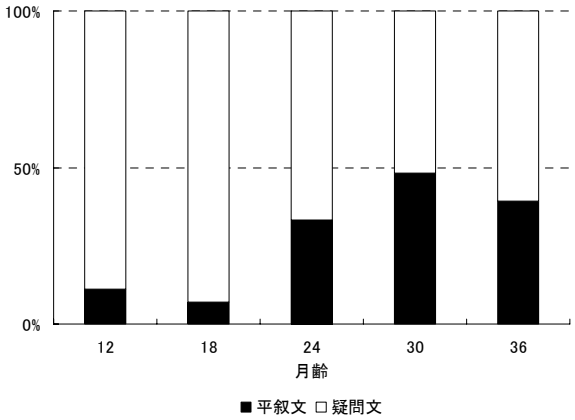

図 3：平叙文と疑問文の割合

助詞ハの直前に共起する割合が高いカテゴリーは 12 ヶ月齢を除き指示詞であるという結果と，全月齢で「疑問文」の割合が高いという結果を踏まえ，「疑問文」で指示詞＋ハがどの程度使用されているかを調べるために，「疑問文」における指示詞＋ハの割合を見た．12，18，24，30，36 月齢でそれぞれ，0%，92.3%，58.3%，35.7%，60.0%であった．指示詞そのものの頻度の割合と比べると，12 ヶ月齢を除き，「疑問文」における指示詞＋ハの頻度の方が高かった．結果を図 4 に示す.

4 考察

まず，子どもが理解しやすいと考える語を選んで助詞ハの直前に共起させている可能性につい

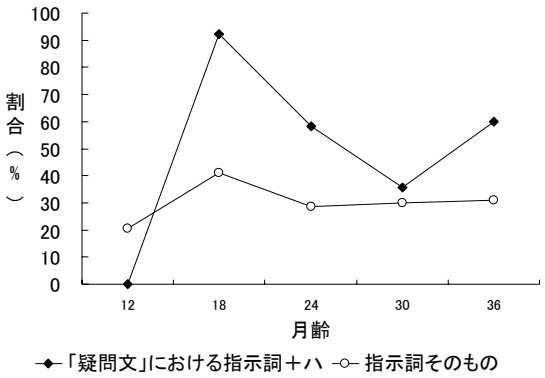

図 4: 疑問文における指示詞＋ハの割合

ては，全月齢を通してそのことを示唆する結果が得られた．それは 12 ヶ月齢において幼児語＋ハの割合が最も高く, 18 ヶ月齢以降において一貫して指示詞＋ハの割合が高かったことから示される．幼児語は普通名詞よりも子どもにとって理解しやすいと大人が考えるカテゴリーであり，指示詞は指し示されるものが最も明確なカテゴリーである．12 ヶ月齢では，幼児語＋ハが指示詞＋ハよりも高い割合であったことから，より低い月齢において助詞ハを使用する場合，指示詞よりも幼児語を使う方が，何を主題にしているかが子どもにとって理解しやすいと考えている可能性がある．ただし，12 ヶ月齢と 18 ヶ月齢についてはカテゴリーそのものの頻度が高いことが，助詞ハとの高い頻度での共起を生じさせている可能性もある．

次に，助詞ハを使用する場面については，主題について何かを述べるのではなく，主題を明確にしながらも，その名称を言わずに命名を求める場面が多いことが明らかとなった．これは「疑問文」が一貫して 50%を超える割合であったことと，「疑問文」における指示詞＋ハが，12 ヶ月齢を除き，指示詞そのものよりも高い割合であったことから示される．主題について述べるためには，子どもがその主題の名称を知っていることが前提となる．主題の名称を尋ねるのは，その前段階と言え，大人が子どもの理解の程度によって助詞ハを使用していることが示唆される．このことは月齢が高くなるにつれて「平叙文」の割合が高くなり，指示詞＋ハの割合も低くなっていたという結果とも一致する．12 ヶ月齢は「疑問文」における指示詞＋ハの使用は見られなかったが, これは 12 ヶ月齢が，まだ命名を求めても答えられない時期であるためと考えられる．

5 まとめ

本稿では，子どもは大人と比べて語や場面の理解が難しいということを踏まえ，大人が子どもに対して話しかける際に，助詞ハの直前に共起させる語や助詞ハを使用する場面を選んでいるのか，また共起させる語や場面を月齢によって変化させているのかについて検討を行った．その結果，大人は子どもの理解の程度に応じて，指し示されるものが最も明確な指示詞を主題に選び，また命名の場面で助詞ハを使用しているという結果が得られた．

文献

[1] 野田尚史:「は」と「が」, くろしお出版, 1995.

[2] 益岡隆志，田窪行則: 基礎日本語文法―改訂版―, くろしお出版, 1992.

[3] B. MacWhinney: *The CHILDES project : Tools for analyzing talk* (3rd ed.), Lawrence Erlbaum Associates, 2000.

[4] T. Ishii: *The JUN Corpus*, Unpublished, 2003.