

統語的曖昧性を解消する イントネーションとジェスチャーの共起*

岡久 太郎
静岡大学

白勢 彩子
東京学芸大学

okahisa-taro@inf.shizuoka.ac.jp

shirose@u-gakugei.ac.jp

概要: 本稿では、統語的曖昧性を解消する韻律的特徴がジェスチャーとどのように関わっているのかを明らかにする。日本語共通語では、発話の統語構造に対応するイントネーションパターンが存在し、統語的に曖昧な文がイントネーションにより区別されることが知られている。また、統語構造の曖昧性を解消するためのジェスチャーパターンが存在することも先行研究において指摘されている。本研究では、発話のマルチモーダル性に着目し、2つの実験を通して、統語的曖昧性を解消するジェスチャーが韻律にどのような影響を与えるかを検討した。実験1では、韻律による曖昧性の解消にジェスチャーが影響を与えるかを確かめるため、実験参加者はジェスチャー無し、ジェスチャー有りの2つの実験条件で統語的曖昧文を一方の意味が伝わるように発話した。その結果、ジェスチャー有りて発話した条件の方がジェスチャー無しで発話した条件よりも有意に韻律による曖昧性解消の効果が小さくなった。この実験の結果から、統語的曖昧性を解消するイントネーションとジェスチャーについて以下の2つの仮説が考えられた: (I) 一般的な認知リソースを必要とするため、認知リソースの不足から両者を同時に産出することができなくなった、(II) 認知的に同一の言語的領域を占めるものであるため、発話プロセスにおいて曖昧性を解消するジェスチャーがイントネーションに置き換わった。いずれの仮説が妥当かを確かめるため、実験2では、手を無意味に動かしながら、ないしは、3つの単語の単語を暗記しながら統語的曖昧文を音読する2つの実験条件を追加した。その結果、いずれの追加条件でもイントネーションによる曖昧性の区別は行なわれており、これは仮説 (II) を支持するものと考えられた。

キーワード: 統語的曖昧性、イントネーション、ジェスチャー、日本語共通語、ダウンステップ

* 本研究は JSPS 科研費 (研究課題番号: JP18J14784) の助成を受けたものである。

1 はじめに

自然言語の発話はその構成要素を線形順序に並べる必要があることから、統語的曖昧性を有することがある。しかしながら、我々のコミュニケーションにおいて話者の意図を伝達するための手段は単語・形態素の配列にとどまらない。当該発話の持つ韻律的情報や共起するジェスチャーのような視覚的情報を使用することによって、言語表現のみでは伝達しきれない内容を表すことが可能である。

統語的曖昧性を持つ発話において、日本語では意図された統語構造に応じて発話のイントネーションパターンが異なることが報告されており、これはダウンステップの生起と弱化として説明される (e.g., Selkirk and Tateishi 1991, 窪園 1995)。一方、ジェスチャーでも統語的曖昧性を解消する方略が報告されている (Kashiwadata et al. 2020, 岡久 2022)。

韻律的特徴とジェスチャーがそれぞれ発話の統語的曖昧性を解消する要素となることが分かっている一方で、それらの間の相互作用については十分な検討がなされていない。そこで、本稿では、統語的曖昧性を解消するイントネーションとジェスチャーの関係性を明らかにすることを目指す。

なお、挨拶のような一語文に伴うジェスチャーやビートと呼ばれる汎用的なジェスチャーは発話の韻律的特徴が変化することが報告されている (Bernardis and Gentilucci 2006, Krahmer and Swerts 2007)。しかしながら、これらの慣習的なジェスチャーは、本稿で対象とする統語的曖昧性を区別することが志向されたジェスチャーとは性格が異なる可能性が高い。本稿では、この点を考慮し、2つの実験を行うことで統語的曖昧性を区別するイントネーションとジェスチャーの位置付けについて検討する。

2 統語的曖昧性を持つ発話に伴うマルチモーダル要素

2.1 日本語におけるダウンステップと統語構造の関係性

日本語において、発話の統語構造を反映した韻律的特徴としてダウンステップという現象が存在する。ダウンステップとは、起伏式アクセントを持つ語句に続く語句の基本周波数 (fundamental frequency: f_0)¹の最大値が、先行する語句の f_0 最

¹ 基本周波数については、伝統的に“ F_0 ”という表記が用いられており、本章で取り上げている先行研究内においても“ F_0 ”によって基本周波数が表されていることが多い。しかし、基本周波数の“ F ”は、第1フォルマント周波数を表す“ F_1 ”や第2フォルマント周波数を表す“ N_2 ”とは異なり、フォルマント周波数を表さないことから、これらと区別し、小文字の f と oscillation の頭文字を用い、“ f_0 ”と表

大値よりも大きく下降する現象である。

例えば、(1)では「'」で表した位置がアクセント核を表しており、第1要素である「旨い」と第2要素である「豆が」はそれぞれ起伏式アクセントで発音されることが分かる。このため、第1要素「旨い」と第2要素「豆が」との間、さらに第2要素「豆が」と第3要素「ありました」との間においてダウンステップが生じる。

(1) 旨い豆がありました

ウマ'イ マメ'ガアリマ'シタ

(五十嵐・小磯 2012:15)

図1は、(1)の発話の波形とピッチ曲線を示したものであり、図下側のピッチ曲線が描かれている縦軸の上端は150Hz、下端は70Hzを表している。これを見ると、順に f_0 最大値が下降していることが確認できる。

一方、先行語が平板式アクセントの場合は、後続語におけるダウンステップは見られなくなる。

(2) 旨い飴がありました

ウマ'イ アメガアリマ'シタ

(五十嵐・小磯 2012:15)

(2)では、第1要素は(1)と同様「旨い」であるため、起伏式アクセントであるが、第2要素が「飴が」となっており、平板式のアクセントとなっている。そのため、第1要素と第2要素の間では(1)同様にダウンステップが生じるが、第2要素と第3要素の間においてはダウンステップが生じない。図2を見ると、実際に「旨い」と「飴が」の間は大きく f_0 最大値が下降しているのに対し、「飴が」と「ありました」の f_0 最大値を比べるとほぼ変化がないことが分かる。

ダウンステップは、先行語が起伏式のアクセントであっても、続く要素と文法的に結合可能であるにも関わらず、それを飛ばしてさらに先の要素と結びつく場合、ダウンステップが弱化・消失することが知られている(e.g., Selkirk and Tateishi 1991, 窪園 1995)。そのため、統語的曖昧文の発話において、意図されている意味に応じて、ダウンステップの生起の有無が異なる場合が存在する。

例えば、(3)の第1要素「黄緑の」は、「ひまわりの」を修飾するか、「ひまわりの模様」全体を修飾するかという統語的曖昧性を有する。

(3) 黄緑の ひまわりの 模様

記述することが近年多くなってきた(Titze et al. 2015)。このような状況を踏まえて、本稿では基本周波数を表す表記として、 f_0 を使用する。

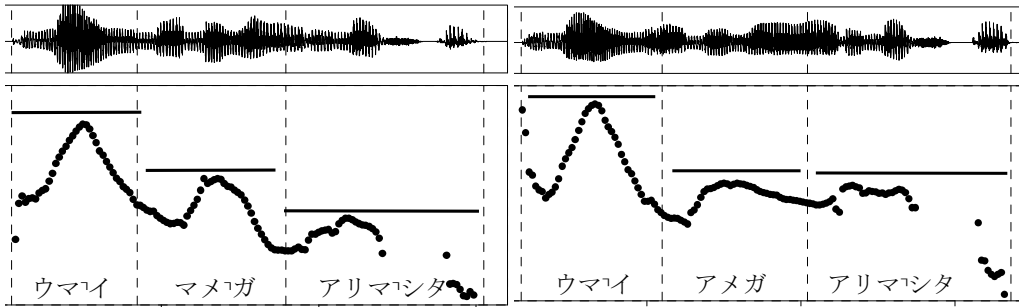


図1 「旨い豆がありました」のピッチパターン (五十嵐・小磯 2012:16 より)

図2 「旨い餡がありました」のピッチパターン (五十嵐・小磯 2012:16 より)

- a. [[黄緑のひまわりの] 模様]
[[キミ'ドリノヒマ'ワリノ]モヨウ]
- b. [黄緑の[ひまわりの模様]]
[キミ'ドリノ[ヒマ'ワリノモヨウ]]

(Venditti 1994:200)

(3) では、第1要素「黄緑の」と第2要素「ひまわりの」が起伏式のアクセント型を持つため、これらの間においてダウンステップが観察されるかによって、いずれの統語構造を意図しているのかを判断することができる。

(3a) の意味では、形容詞句「黄緑の」と名詞「ひまわりの」が結合し、名詞句「黄緑のひまわりの」が構築され、それが名詞「模様」と結合している。一方、(3b) の意味では、形容詞句「黄緑の」と直後の名詞「ひまわりの」は直接結合せず、「模様」を主要部とする「ひまわりの模様」全体と結合する。

よって、(3a) では、全ての要素の間でダウンステップが生じるが、(3b) では、「黄

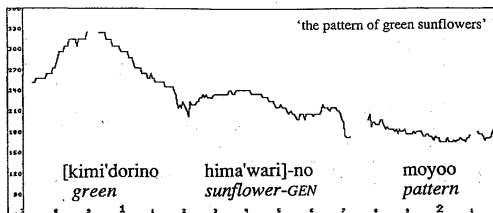


図3 (3a) の発話のピッチ曲線 (Venditti 1994:202)

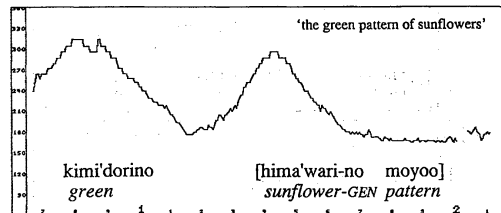


図4 (3b) の発話のピッチ曲線 (Venditti 1994:202)

緑の」と「ひまわりの」の間において、ダウンステップがリセットされることが予測される。図3と図4は、(3)を2つの統語構造を意図して発話した際のピッチ曲線であるが、上述の予測通り、「黄緑の」と「ひまわりの」の間のダウンステップの有無が異なっていることが分かる。

以上のように発話の韻律的特徴であるダウンステップは、発話の統語構造に対応してその振る舞いを変えることが古くから指摘されてきた。しかしながら、ダウンステップと発話の統語構造の関係性については、発話のみに集中させた条件での実験によって明らかになったものである。しかしながら、実際の発話ではイントネーションだけでなく、ジェスチャーにおいても統語構造に応じたパターンが少数だが観察されている。次節では、統語的曖昧性を持つ発話に伴うジェスチャーパターンに関する研究について概観する。

2.2 統語的曖昧性を解消するジェスチャーパターン

統語的曖昧性を有する発話に伴うジェスチャーについては、ジェスチャーの生起タイミングと内容に関して統語構造に対応したパターンが報告されている。

ジェスチャーの生起タイミングについては、Kashiwadata et al. (2020) が(4)のような統語的曖昧表現の発話に伴うジェスチャーのストローク²³生起時間に着目している。

- (4) 黒い尻尾の大きな猫
- a. [黒い [[尻尾の大きな] 猫]]
 - b. [[黒い 尻尾の] [大きな 猫]]
 - c. [[黒い [尻尾の 大きな]] 猫]

Kashiwadata et al. (2020) では、事前調査において理解がされやすかった(4a)と(4b)の2つの解釈に対応するイラストの一方のみを赤い枠で囲んだ刺激を呈示し、(4)の表現をジェスチャーと共に音読させた。統計的分析の結果、統語境界を

² ここでのストロークとは、Kendon (1980, 2004), McNeill (1992, 2005) 等において提案されているジェスチャーの記述法における呼称である。その記述法においては、1まとまりのジェスチャーをジェスチャー句 (Gesture Phrase) と呼び、ジェスチャー句は複数のジェスチャーフェーズ (gesture phase) から構成されている。ジェスチャーフェーズは準備 (preparation)、ホールド (hold)、ストローク (stroke)、復帰 (recovery/retraction) に大別される。ストロークは、ジェスチャーの意味が表現されるジェスチャーの中心であり、1つのジェスチャー単位に必須の要素であるとされる。

³ なお、Kashiwadata et al. (2020) には、ジェスチャーのストロークを分析対象とした旨は明記されていないため、当箇所についてはその論文の共著者の1人である小林春美氏 (東京電機大学 教授) との私信による情報である。

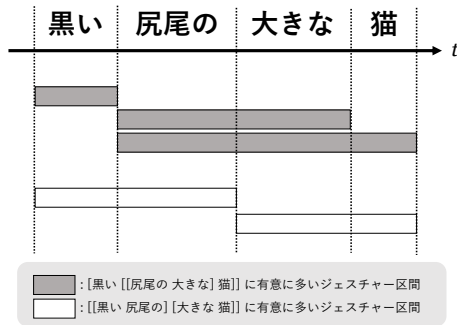


図5 Kashiwadate et al. (2020) の実験結果の概要

跨がないようにジェスチャーが実行されていることが分かった。(4a)の解釈が意図された場合、第1要素「黒い」の発話区間のみ、第2要素「尻尾の」から第3要素「大きな」にかけての区間、第2要素「尻尾の」から第4要素「猫」にかけての区間において、ジェスチャーのストロークが観察されることが、(4b)の解釈を意図した場合よりも有意に多かった。一方、(4b)の解釈が意図された場合、第1要素「黒い」から第2要素「尻尾の」にかけての区間、ないしは、第3要素「大きな」から第4要素「猫」にかけての区間にジェスチャーのストロークが行われることが(4a)の解釈を意図した場合よりも有意に多かった。

これらの結果を筆者らが整理して示したものが図5である。図中では(4a)の解釈において(4b)よりも有意に多く観察されたジェスチャーのストローク区間を灰色の棒で示し、(4b)の解釈において(4a)よりも有意に多く観察されたジェスチャーのストローク区間を白の棒で示している。

一方、ジェスチャーの内容については、岡久(2022)において(5)のような「Adj. N₁ の N₂」型の統語的曖昧表現を用いた実験を行った。

(5) 黄色い運動靴の紐

- a. [[黄色い運動靴の]紐]
- b. [黄色い[運動靴の紐]]

実験では、aとbの意味に対応したイラスト・写真の一方に黒い枠で囲まれた画像を参加者に呈示し、その意味が聞き手に伝わるようにジェスチャーと共に発話を行わせた。その結果、N₁を表すジェスチャーのみを行うパターン(N₁単独型)とN₁を表すジェスチャーを行った後にN₂を表すジェスチャーを続けて行うパターン(N₁→N₂型)がaの解釈を意図した場合に有意に多く、N₂を表すジェスチャー

のみを行うパターン (N_2 単独型) が b の解釈を意図した場合に有意に多かった。

以上の結果は、それぞれの文法的主要部をジェスチャーによって表現していると解釈できる。 a の解釈においては、最初に Adj. と N_1 が結合し、 N_1 が主要部となるが、次に [Adj. N_1] が N_2 と結合し、 N_2 が主要部となる。そのため、最初の Adj. と N_1 の結合に着目した場合、 N_1 のみが主要部となるが、表現全体に着目した場合、その主要部は N_1 から N_2 へと遷移する。一方、 b の解釈においては N_1 と N_2 の結合においても、それに続く Adj. と [$N_1 N_2$] の結合においても、主要部は常に N_2 である。意図された解釈に応じたジェスチャーパターンの違いは、上述の主要部の変遷を反映したものであると考えられる。

なお、岡久(2022)では(5)のように Adj. が色を表す統語的曖昧表現だけでなく、「四角いリュックのポケット」のように Adj. が形状を表す形容詞についても別に検討している。その結果、形状を表す形容詞はジェスチャーによってその内容を伝えることが可能であるため、形状形容詞が用いられる場合は必ずしも主要部に対応したジェスチャーが行われるわけではなかった。以上のように、統語的曖昧表現に共起するジェスチャーの生起タイミングや内容は言語表現の意味内容とも関わりながら、一定のパターンを有していることが考えられる。

2.3 韻律的特徴とジェスチャーの相互作用

前節まで見てきたように、統語的曖昧文に共起するイントネーションとジェスチャーは各々、意図された統語構造に対応したパターンを持っていることが分かっているが、それらの間の相互作用については管見の限りこれまで分析がなされていない。しかしながら、統語構造とは直接的に関係しない韻律的特徴については、ジェスチャーとの相互作用が考えられている。本節では、ジェスチャーが共起することによって韻律的特徴がどのような影響を受けるかを検討した研究を紹介する。

共起するジェスチャーが発話に対応したものである場合に限り、韻律的特徴が変化するということが Bernardis and Gentilucci (2006) において指摘されている。彼らは、イタリア語話者に“CIAO,” “NO,” “STOP” という一語文をジェスチャー無し、各々に対応したジェスチャー有り、無意味な手の動き有りの3条件の下、発話させた。その結果、ジェスチャー無しの時と比べ、各々に対応したジェスチャーを行った場合、 f_0 と第2フォルマント (N_2) の平均値が有意に高くなった。一方、無意味な手の動きと共にこれらの単語を発話させた場合、発話のみを行わせた場合と比べて韻律的特徴に有意差は見られなかった。また、擬似語 (pseudo-word) で

ある“LAO”という発話に対して、“CIAO,” “NO,” “STOP”のそれぞれに対応するジェスチャーを伴わせた場合、発話のみの場合と比較しても韻律的特徴に有意差は見られなかった。この結果は、発話に対応したジェスチャーが f_0 値と N_2 値を上昇させたということを意味している。Bernardis and Gentilucci (2006) は特に N_2 値の上昇に着目し、これは話者がジェスチャーを伴っていない場合と比較し、より協調的な発話を行った結果であると考察している。なぜなら、 N_2 値が高いということは、舌がより前に出ていることを意味しており、これは親密なインタラクションを行う場合の特徴であると考えられるためである⁴。

また、手、指、頭、眉といった部分を上下や前後に細かく動かすジェスチャーであるビート (McNeill 1992, McNeill et al. 2015, Theune and Brandhorst 2010) の共起が韻律的特徴に影響を与えることが Krahmer and Swerts (2007) において報告されている。彼らはオランダ語の「Amanda は Malta に行く」(“Amanda gaat naar Malta”) という文の最初の語 (“Amanda”) か最後の語 (“Malta”) の発話時に、ビートを伴わせ、最初の語と最後の語それぞれにおける強勢を持つ/a/の音響特性の差を分析した。その結果、 f_0 値と F_1 の値については、ビートが伴っているか否かによる変化は見られなかったが、 N_2 の値については、ビートが伴った語における N_2 値がそうではない語の N_2 値よりも有意に低くなった。この結果は、先述の Bernardis and Gentilucci (2006) とは対照的なものであるが、これにはオランダ語固有の特徴が関係していると考えられる。オランダ語においては、/a/等の母音にアクセントが置かれる場合は N_2 値は下がるという事実が指摘されており、Krahmer and Swerts (2007) の実験において、ビートの共起による N_2 値の減少の効果量はアクセントを置くことによる N_2 値の減少の効果量と同等であった。このことから、ビートはアクセントを置くこととほぼ同等の働きをしていると考えられる。

以上の結果は、ジェスチャーが共起することによりコミュニケーションに関与した発話の韻律的特性が変化することを示唆している。ここで重要な点は、ジェスチャー共起による発話の変化は必ずしも言語普遍的なものではなく、話されている個別言語の特徴に基づいたものであるという点である。すなわち、ジェスチャーと韻律的特徴の関係性は、生理的な動機付けによる人類共通の特徴ではなく、あくまで特定の言語コミュニティ内において形成されたものであると考えることができるだろう。

以上のように、ジェスチャーが発話の韻律的特徴に与える影響に関しては、一

⁴ このような特徴は、人間だけではなく、猿の毛繕い行動に伴うリップスマッキング (lip-smacking) における舌の前方向への運動にも見られるという (Van Hooff 1962, 1967)。

語文のジェスチャーやビートのようなコミュニケーション的機能を果たすものに着目した先行研究が多かった。本稿では、統語的曖昧文に共起するジェスチャーが発話のイントネーションに与える影響を検討することで、ジェスチャーの言語的性質に着目する。

3 実験 1: 統語的曖昧性を区別する韻律的特徴とジェスチャーの共起について

本節では、ジェスチャーが共起することによりダウンステップの有無による統語構造の区別が阻害されるかを検討する実験を行う。本実験では、統語的曖昧文を一方の意味で発話のみに集中させて音読させた条件と意味を区別するジェスチャーを共起させた状態で音読させた条件を比較することで、ジェスチャーによるイントネーションへの影響を見ていく。

3.1 実験に用いる刺激文

統語的曖昧文のタイプとして、以下の3タイプを用意した。

(6) 形容詞型

マリは 小さい バッグの ポケットを探った

- a. [マ'リハ [[[チイサイ バ'ッグノ] ポケ'ットヲ] サグ'ッタ]]
- b. [マ'リハ [[チイサイ [バ'ッグノ ポケ'ットヲ]] サグ'ッタ]]

(7) 副詞型

マミコは こっそり 話している 人を注意した

- a. [マ'ミコハ [[[コ'ソ'リ ハナ'シテイル] ヒトヲ] チュ'ウイシタ]]
- b. [マ'ミコハ [コ'ソ'リ [[ハナ'シテイル ヒトヲ] チュ'ウイシタ]]]

(8) 与格型

コウキは ケイコに 上司が書いた メモを見せた

- a. [コ'ウキハ [[[ケ'イコニ [ジョ'ウシガ カ'イタ]] メ'モヲ] ミ'セタ]]
- b. [コ'ウキハ [ケ'イコニ [[[ジョ'ウシガ カ'イタ] メ'モヲ] ミ'セタ]]]

いずれの文タイプにおいても、下線の引かれた第2要素(「小さい」、「こっそり」、「ケイコに」)の修飾する部分がaとbの解釈に応じて異なる。aの解釈の場合は第3要素から始まる直後の部分(「バッグの」、「話している」、「上司が書いた」)を修飾している。一方、bの解釈の場合は、aの解釈における修飾箇所よりも長い

部分（「バッグのポケットを」、「話している人を注意した」、「上司が書いたメモを見せた」）を修飾対象としている。すなわち、bの解釈はaの解釈では存在しなかった統語境界の左端が第2要素と第3要素の間に増えている。

本稿では以上3タイプの異なる統語構造を持つ文をターゲット文として使用するため、それぞれのaとbの解釈について、先行研究で多く使われている左枝分かれ・右枝分かれという用語によって指すことは適切ではない⁵。そこで、aの解釈では形容詞、副詞、与格が、bの解釈よりも小さい単位と結合していることから、本稿ではaの解釈を短単位修飾、bの解釈を長単位修飾と呼ぶこととする。

3.2 条件1の設定と予測される結果

実験1で検討する問題は以下の通りである。

- (9) ダウンステップの有無による統語構造の区別は、ジェスチャー共起に関わらず、当該の統語構造において義務的に使用されるものであるのか。

この問題を考えるために、以下の2つの条件を設定した。

- (10) a. 発話のみ条件
b. ジェスチャー共起条件

1つの条件において、各タイプ2文ずつ、合計6文を条件間での重複のないようにプログラムを用いて無作為に選んだ。(10a)の発話のみ条件は、事前に黙読した文脈に合う意味でジェスチャーをさせずにマイクに向かって統語的曖昧文を音読するよう実験参加者に求めた。(10b)のジェスチャー共起条件においては、カメラに向かって文脈に適した意味で統語的曖昧文をジェスチャー付きで音読するよう実験参加者に求めた。

これらの条件を比較することにより、先述の(9)の間を検討する。すなわち、統語構造に対応したダウンステップの有無がジェスチャー共起条件にも変わらず観察された場合、ダウンステップによる統語構造の区別は義務的なものであることが支持される。一方、ジェスチャー共起条件においてダウンステップによる区別が減少した場合、イントネーションによる統語構造の表示は選択的なものであることが示唆される。

⁵ そもそも形容詞型について左枝分かれ・右枝分かれという用語を用いることに関しても、統語構造を発話の線形順序に並び替えた時に生じるものであるため、純粋に統語的解釈について考える場合には適切な用語とは言い難いだろう(著者の発表における藤田耕司氏(京都大学教授)のコメント)。

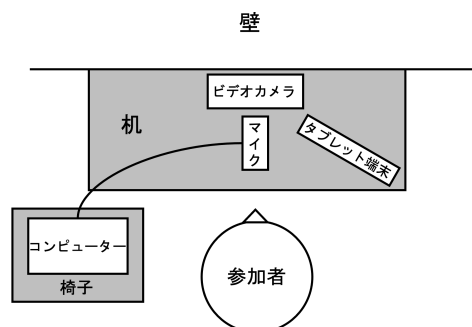


図6 実験環境の模式図

3.3 参加者と実験手続

実験は2018年12月から2019年1月にかけて、東京学芸大学において実施した。実験参加者は、実験2と共通の東京学芸大学の大学生20人であり、各人に全ての実験に参加した謝礼として1,000円分の図書カードを渡した。

実験環境は図6に模式的に示した。壁沿いに置いた机の上に、マイクロフォン(audio-technica製AT9934USB)とビデオカメラ(SONY製HDR-MV1)を設置し、実験参加者をカメラとマイクに向かって座らせた。なお、マイクは机の脇に置いたコンピューター(Apple製Macbook Pro 13inch)に接続し、音声録音・編集ソフトAudacityを用い、サンプルレート44100Hz、32bitでモノラル録音を行った。刺激呈示には、全てタブレット端末(Apple製iPad Pro 11inch)を用い、Pythonコードをタブレット上で実行することができるアプリケーションであるPythonista3を使用し、刺激呈示の管理を行った。実験1の実施手順は以下の通りである。

(11) 実験1の実施手順

1. タブレット画面に先行文脈とそれに対応した統語的曖昧文が表示される。参加者はそれらを黙読し、自身が音読する意味が理解できたら画面をタッチする。
2. 白い画面が5秒間表示され、参加者はその間待つ。
3. 自動的に統語的曖昧文のみが画面に表示される。参加者は、手順1において理解した方の意味でその文を音読する。
4. 音読を終えたら、手元の評価用紙に自身の発話がどの程度曖昧性を解消できていると思うかについて、1から7の7段階評価を行う。

発話のみ条件では、音声を聞いた第三者が意図された意味を理解できるようにマ

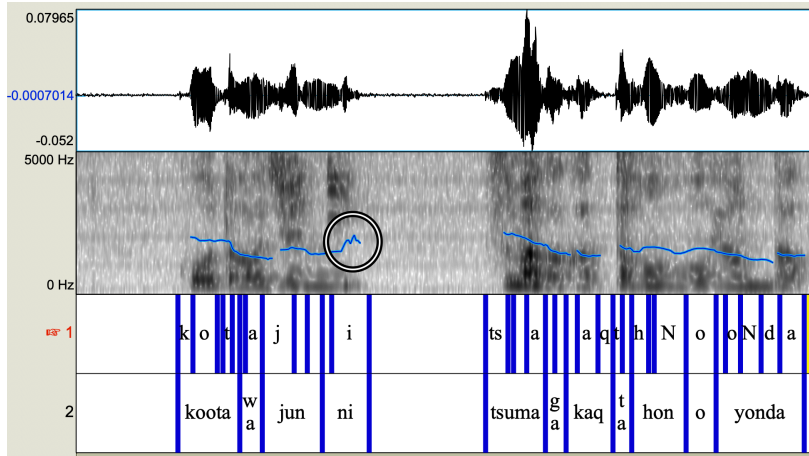


図7 「コウタはジュンに妻が買った本を読んだ」(長単位修飾)のPraatにおけるアノテーション画面。第2要素「ジュンに」は頭高型のアクセントパターンであるが、末尾のモーラ“ni”でピッチが急上昇している。

イクに向けて発話するように指示した。一方、ジェスチャー共起条件では、映像を視聴した第三者が意図された意味を理解できるようにカメラに向けてジェスチャーを共起させた発話をするように指示した。なお、実験参加者の意識はビデオカメラに向けているが、音声の録音は発話のみ条件と同一のマイクによって行っており、データ分析はその音声を用いた。本試行実施の前に発話のみ条件とジェスチャー共起条件の双方の練習を3文ずつ行った。本試行における両条件の実施順序は、無作為に決定した。

3.4 分析手法

本実験では、第2要素と第3要素との間にダウンステップが生起しているか否かを検討するため、それらにおける f_0 最大値を測定することを目的とし、分析手法を設定した。最初に、音素毎の時間長を認定するために、Julius (Lee et al. 2001, Lee and Kawahara 2009) の音素セグメンテーションキットを用いた。Julius の出力を Praat (Boersma and Weenink 2020) で扱える TextGrid ファイルに変換し、Praat 内で1ファイルずつ著者の1人が手作業で音素分割の誤抽出を修正した。今回の刺激文においては図7の丸印部分のように文末以外の文節末の最後の1モーラでピッチが急上昇し、その箇所でのピッチ最大値がアクセント核におけるピッチを上回ってしまう場合が見られたため、文節末の1モーラ以外の区間のピッチ最大値を Praat を用いて計測した。

今回の分析では、 f_0 最大値の差について議論するが、単純に2点間の f_0 [Hz] の差をとったのでは参加者の声の元々の高さに応じて、値の大ききの持つ意味が異なってしまう。そこで、本稿では、ピッチ分析において、Hzを対数変換した値であるセミトーン⁶を用いて、第2要素と第3要素のピッチ差について検討した。

統計手法は、線形混合効果モデルを使用し、従属変数として第2要素と第3要素のセミトーンピッチ最大値の差を、固定効果として意図された解釈(短単位修飾、長単位修飾)と条件(発話のみ条件、ジェスチャー共起条件)を、変量効果として実験参加者、ターゲット文、文タイプ(形容詞型、副詞型、与格型)を設定した。データ処理には、R(R Core Team 2018)を使用し、lme4(Bates et al. 2015)、lmerTest(Kuznetsova et al. 2017)、emmeans(Lenth 2021)のパッケージを用いた。さらに、各条件における統語構造に対応したダウンステップの有無を検証するために、第2要素と第3要素のセミトーンピッチ最大値の差についてチャンスレベル(=0、すなわち第2要素と第3要素が同じ値を取る場合)との比較をt検定によって行った。

3.5 結果

実験1の両条件における意図された解釈ごとの第2要素と第3要素の平均値は、図8の通りである。

第2要素は形容詞(e.g.「小さい」)、副詞(e.g.「こっそり」)、与格(e.g.「ケイコに」)のいずれかの要素であり、短単位修飾型の解釈では直後の部分(e.g.「バッグの」、「話している」、「上司が書いた」)を修飾し、長単位修飾型の解釈ではそれよりも後の部分まで(e.g.「バッグのポケットを」、「話している人を注意した」、「上司が書いたメモを見せた」)を修飾する。図8を見ると、図8aも図8bも短単位修飾の場合は第2要素と第3要素の間でピッチ最大値の大きな下降が見られ、長単位修飾の場合は両者のピッチ最大値の差が少なくなっている。しかしながら、長単位修飾型のピッチ最大値の変動について、図8aの発話のみ条件では第2要素から第3要素にかけて若干の上昇が見られるのに対して、図8bのジェスチャー共起条件では若干の下降が見られる。ここから、ジェスチャー共起条件の長単位修飾型においてダウンステップのリセットが発話のみ条件よりも弱いものになっていることが読み取れる。

上述の記述的分析は、統計分析からも支持される。線形混合効果モデルに

⁶ セミトーンへの変換においては、基準 f_0 値 (0 *semitone*) を 100Hz に設定し、以下の変換式を用いた (x は Hz を単位とする f_0 値)。

$$\text{semitone} = 12 * \log(x/100) / \log 2$$

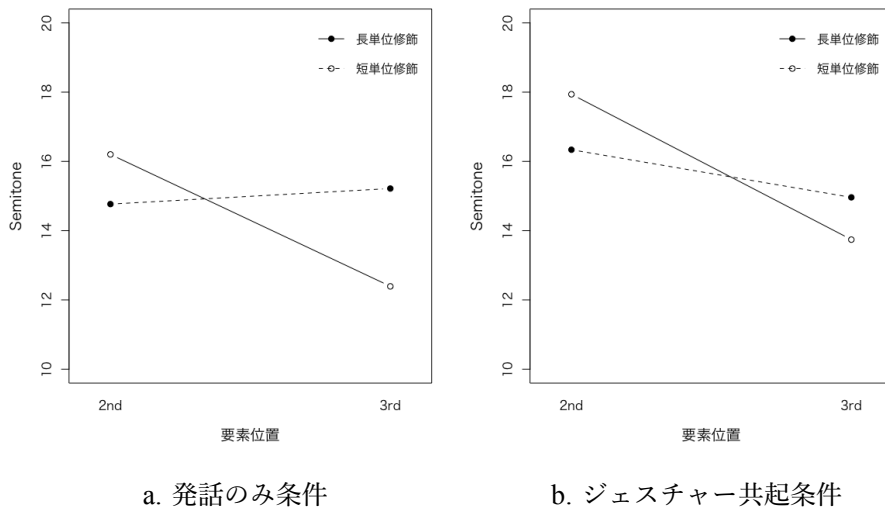


図8 実験1の各条件における第2,3要素のセミトーン最大値の平均値

よる分析の結果、条件の主効果 ($F(1, 192.47) = 10.68, p = 0.001284$)、意図された解釈の主効果 ($F(1, 198.89) = 88.94, p < 0.0001$)、条件と解釈の交互作用 ($F(1, 191.81) = 4.04, p = 0.045880$) がそれぞれ有意であった。

また、短単位修飾の解釈における条件の単純主効果について有意差は見られなかった (ジェスチャー共起条件 - 発話のみ条件: Estimate = 0.44, SE = 0.48, $t(194) = 0.91, p = 0.37$; p 値は Tukey HSD 法を用いて調整) が、長単位修飾の解釈における単純主効果には有意差が見られた (ジェスチャー共起条件 - 発話のみ条件: Estimate = 1.82, SE = 0.50, $t(191) = 3.67, p = 0.0003$; p 値は Tukey HSD 法を用いて調整)。すなわち、長単位修飾型の解釈において、発話のみ条件で見られた第2要素と第3要素間に見られるダウンステップのリセットが、ジェスチャー共起条件では阻害されていることが示された。

さらに、各条件・各解釈における第2要素と第3要素のピッチ最大値の差をチャンスレベル (=0) と比較した結果、発話のみ条件においては、短単位修飾型においては有意な下降が見られ ($t(58) = 12.70, p < 0.0001$)、長単位修飾においては有意な上昇・下降が見られなかった ($t(53) = -1.08, p = 0.283$)。一方、ジェスチャー共起条件においては、短単位修飾型で有意な下降が見られた ($t(54) = 11.47, p < 0.0001$) だけでなく、長単位修飾型でも有意な下降が見られた ($t(50) = 3.05, p < 0.003628$)。

3.6 考察

実験1の結果から、発話のみによって統語的曖昧文の意味を区別することを試みた場合とジェスチャーを共起させて同一の課題を行う場合とでは、韻律的区別に変化が見られることが示された。実験1で示したように、発話のみの場合では、短単位修飾の解釈においてダウンステップが生じるのに対して、長単位修飾の解釈においてはダウンステップがリセットされていた。一方、ジェスチャーが共起した場合は、長単位修飾型の解釈におけるダウンステップのリセットが阻害され、いずれの解釈においてもダウンステップが生じていることが分かった。この結果は、ダウンステップの有無による統語構造の区別は義務的なものではなく、選択的なものであることを示唆している。

この場合、統語構造を区別するイントネーションとジェスチャーの役割として以下の2つの可能性が考えられる。

- (12) a. 一般的な認知リソースを用いる発話の付加的要素
- b. 認知的に言語的領域を占める要素

(12a)は、イントネーションとジェスチャーを発話において認知的・運動的負荷をかける付加的要素の1つとして考える立場であり、ジェスチャーをすることでイントネーションの区別をする余裕がなくなったと捉える見方である。一方、(12b)は、統語的区別を行うイントネーションとジェスチャーが認知的に言語に関わる特別な位置を占めるものであると捉える立場であり、統語構造の区別においてはそれらの共存は実現されないと考える見方である。注意すべきは、前者はあくまで統語構造を区別するイントネーションとジェスチャーが一般的な認知リソースの割り振り先の1つとして想定されているに過ぎないのに対し、後者はそれらが言語的に特別な位置に置かれる要素であると捉え、単なる認知リソースの配分によって説明することが難しいと考えているという点である。

次節では、以上の点を考慮するため、実験1の条件に加えて、認知的・運動的負荷をかける条件を加えた比較を行う。

4 実験2: 認知的・運動的負荷の増大による統語構造の韻律的区別への影響について

前節の実験1の結果から、ダウンステップの有無による統語構造の区別はジェスチャーの介入により使用されにくくなるものであり、当該の統語構造において義務的なものではないことが示唆された。これを受け、統語構造を区別するイントネーションとジェスチャーが(12)の2つのいずれかであることが考えられた。

そこで、新たに2つの実験条件を加え、認知リソースの観点から実験1の結果を説明することができるかを検討した。

4.1 条件設定と予測される結果

実験2では、(12)の2つの可能性のいずれがより妥当であるかを検討するため、実験1で比較した2条件に2つの条件を加えた以下の計4条件を比較する。

- (13) a. 発話のみ条件
 b. ジェスチャー共起条件
 c. 手の動き条件
 d. 単語記憶条件

(13a)と(13b)の2条件は、実験1と同様のものであり、実験1のデータを用いた。(13c)の手の動き条件は、矢印で指示された無意味な形の手の動きをしながら、統語的曖昧文を音読する条件である。また、(13d)の単語記憶条件では、呈示された3つの単語を暗記した状態で統語的曖昧文を音読し、その後に記憶した単語を報告する課題である。暗記する単語は、天野・近藤(2000)の単語親密度における文字単語親密度5.5以上⁷のもののうち、漢字2文字で4モーラのもので無作為抽出した18語から3語ずつを組み合わせたものである。なお、3語の組み合わせについては、いずれの実験参加者に対しても同じ組み合わせを用いたが、呈示順についてはプログラムによってランダム化した。

手の動き条件は運動的負荷を、単語記憶条件は認知的負荷をそれぞれ発話時にかけることを意図して設定した。これらの条件のいずれか、ないしは両者において、発話のみ条件と比較してダウンステップの有無による統語構造の区別が減少していた場合、(12a)の仮説が支持される。一方、これらの条件においても発話のみ条件と変わらずにダウンステップの有無による区別が観察された場合、(12b)の仮説を支持していると考えられる。

⁷ 天野・近藤(2000)の単語親密度は、日本語の単語を音声のみ、文字のみ、音声と文字の3条件で呈示し、1-7の7段階で各単語の「なじみ」を評定させ、その平均値を採用したデータベースである。Amano and Kondo (1998)によると、単語親密度5以上の単語であれば、94%の成人が知っていることと推定されているため、本稿ではそれよりも高い5.5を閾値として暗記のための単語を選出した。

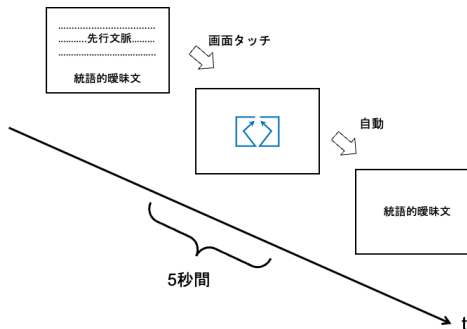


図9 手の動き条件の流れ

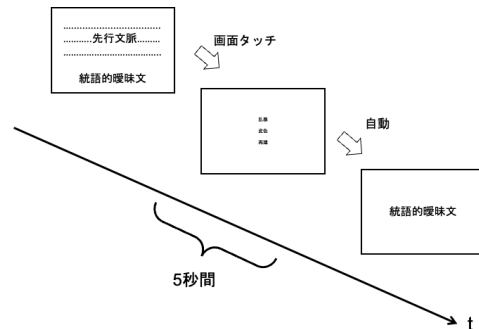


図10 単語記憶条件の流れ

4.2 参加者と実験手続き

実験は2018年12月から2019年1月にかけて、東京学芸大学において実施した。実験参加者は、実験1と共通の東京学芸大学の大学生20人であり、各人に全ての実験に参加した謝礼として1,000円分の図書カードを渡した。実験環境は、実験1と同一のものであり、全ての刺激をiPadで呈示し、その様子をマイク、カメラで録音、録画した。

発話のみ条件とジェスチャー共起条件は、実験1,3と同様のデータを用いた。今回の実験で追加した2条件は、発話のみ条件とジェスチャー共起条件の本試行が終了した後に、無作為に順序を決定し、実施した。各々の条件の具体的な手続きは以下の通りである。

(14) 手の動き条件の実施手順 (図9)

1. タブレット画面に先行文脈とそれに対応した統語的曖昧文が表示される。参加者はそれらを黙読し、自身が音読する意味が理解できたら画面をタッチする。
2. 両手の動きを指示する矢印が5秒間表示され、参加者はその矢印の通りに手を動かし始める。
3. 自動的に統語的曖昧文のみが画面に表示される。参加者は、手を動かし続けた状態で手順1において理解した方の意味でその文を音読する。
4. 音読を終えたら、手元の評価用紙に自身の発話がどの程度曖昧性を解消できているかについて、1から7の7段階評価を行う。

(15) 単語記憶条件の実施手順 (図 10)

1. タブレット画面に先行文脈とそれに対応した統語的曖昧文が表示される。参加者はそれらを黙読し、自身が音読する意味が理解できたら画面をタッチする。
2. 3つの単語が5秒間表示され、参加者はそれらの単語を暗記する。
3. 自動的に統語的曖昧文のみが画面に表示される。参加者は手順1において理解した方の意味でその文を音読する。
4. 音読を終えたら、手元の回答用紙に暗記した単語を思い出せる限り記入する。
5. 最後に、評価用紙に自身の発話がどの程度曖昧性を解消できていると思うかについて、1から7の7段階評価を行う。

手の動き条件も単語記憶条件も、解釈を一意に決定するための文脈を読んだ後に呈示される画面が実験1の2条件とは異なる。手の動き条件においては、図9のように、手の動かし方の指示が矢印で5秒間表示される。参加者はこの間に両手の運動を開始するよう事前に指示されており、5秒後に統語的曖昧文が表示されても手の動きを止めずに音読をするように求められた。単語記憶条件においては、図10のように、文脈の黙読画面の後に暗記する3単語が5秒間表示され、それらの単語を暗記した状態で統語的曖昧文の音読をするように指示された。単語記憶条件については、自信度を回答する前に暗記した単語を回答する時間を設けている点で他の3条件と異なる。

なお、手の動き条件と単語記憶条件に関しても、発話のみ条件とジェスチャー共起条件の時と同様に本試行の前に各条件3文ずつ練習試行を実施した。この練習試行で使用した統語的曖昧文は、発話のみ条件とジェスチャー共起条件で用いた3文と同じものである。手の動き条件と単語記憶条件双方の練習試行の終了後、これらの条件の本試行を無作為に決定した順序で実施した。

4.3 分析手法

分析の方針と手法は実験1,3と同様である。統計手法は、線形混合効果モデルを使用し、従属変数として第2要素と第3要素のセミトーンピッチ最大値の差を、固定効果として意図された解釈(短単位修飾、長単位修飾)と条件(発話のみ条件、ジェスチャー共起条件、手の動き条件、単語記憶条件)を、変量効果として実験参加者、ターゲット文、文タイプ(形容詞型、副詞型、与格型)を設定した。

なお実験1において見たように、ジェスチャー共起条件においては発話のみ条

件と異なり、短単位修飾型の解釈のみでなく、長単位修飾型の解釈においてもダウンステップが見られた。さらに、今回の実験において検討するのは、あくまでもジェスチャー共起条件と他の3条件であることから、今回の分析においては解釈ごとの単純主効果を検討するのではなく、上述の線形混合効果モデルのベースラインを長単位修飾型解釈のジェスチャー共起条件とし、他の3条件の係数を検討した。

さらに本実験では、ジェスチャー共起条件が認知的負荷を伴うことによりダウンステップによる区別が行われなくなるのかを考えるために、自信度についても比較を行った。分析方法は、従属変数として自信度(1-7)を、固定効果として条件(発話のみ条件、ジェスチャー共起条件、手の動き条件、単語記憶条件)を、変量効果として実験参加者、ターゲット文、文タイプ(形容詞型、副詞型、与格型)を設定した線形混合効果モデルを用いた。

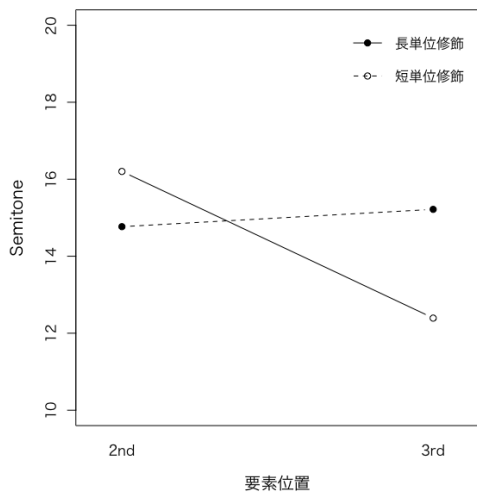
データ処理には、R(R Core Team 2018)を使用し、lme4(Bates et al. 2015)、lmerTest(Kuznetsova et al. 2017)、emmeans(Lenth 2021)のパッケージを用いた。さらに、各条件における統語構造に対応したダウンステップの有無を検証するために、第2要素と第3要素のセミトーンピッチ最大値の差についてチャンスレベル(=0、すなわち第2要素と第3要素が同じ値を取る場合)との比較を*t*検定によって行った。

4.4 結果

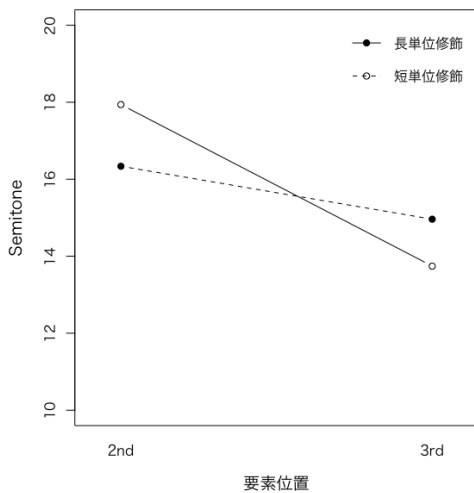
4条件それぞれで意図された解釈ごとの第2要素と第3要素のセミトーン最大値の平均は図11に示した通りである。グラフの概形を見ると、図11a-dいずれも短単位修飾型では第2要素と第3要素との間で大きなピッチ下降が見られ、長単位修飾型ではそれほどのピッチ下降は見られない。しかしながら、図11bのジェスチャー共起条件のみ、短単位修飾型と長単位修飾型のピッチ変化の差が少ないことが見て取れる。この点を確かめるために、各条件ごとに短単位修飾型、長単位修飾型それぞれにおける第2要素の最大セミトーンから第3要素の最大セミトーンを引いた値の平均値を図12に示す。

線形混合効果モデルによる分析の結果、主効果については、意図された解釈($F(1, 411.09) = 193.42, p < 0.0001$)と条件($F(3, 404.23) = 5.02, p = 0.001995$)の両者が有意であり、それらの交互作用は有意傾向であった($F(3, 407.42) = 2.11, p = 0.098396$)。

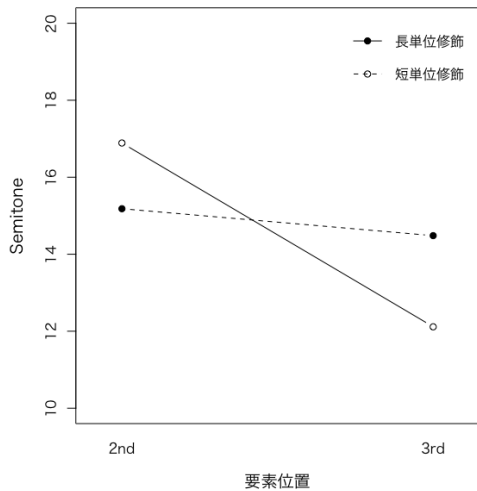
ベースラインを長単位修飾型解釈のジェスチャー共起条件とし、他の3条件(発話のみ条件、手の動き条件、単語記憶条件)の係数を検討した結果、発話のみ



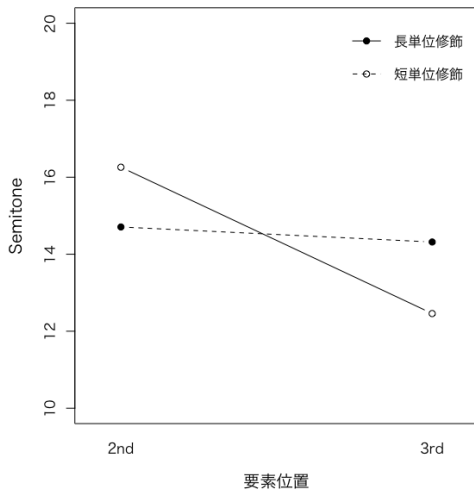
a. 発話のみ条件



b. ジェスチャー共起条件



c. 手の動き条件



d. 単語記憶条件

図 11 実験 2 の各条件における第 2, 3 要素のセミトーン最大値の平均値

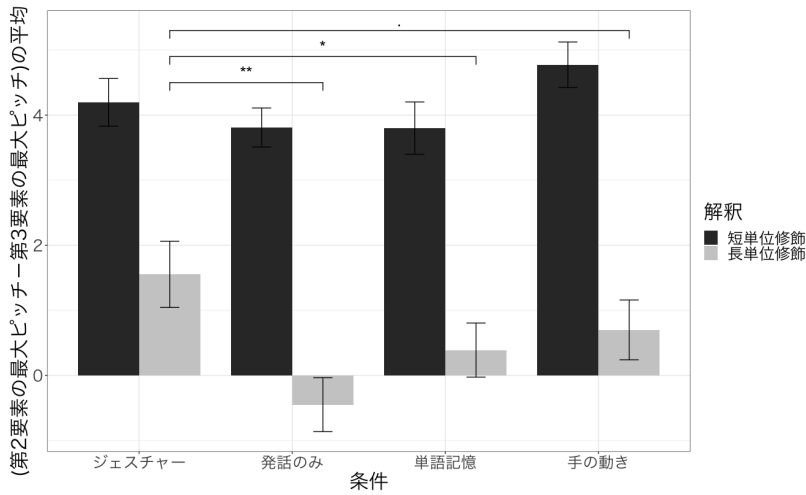


図 12 各条件における解釈ごとの第 2 要素と第 3 要素のセミトーン最大値の差の平均 (** $p < 0.01$, * $p < 0.05$, . $p < 0.10$)

条件 (Estimate = -1.62 , SE = 0.51 , $t(405.05) = -3.84$, $p = 0.000145$) と単語記憶条件 (Estimate = -1.25 , SE = 0.51 , $t(407.30) = -2.43$, $p = 0.015345$) については有意であり、手の動き条件については有意傾向であった (Estimate = -0.96 , SE = 0.51 , $t(405.11) = -1.88$, $p = 0.061215$)。

また、各条件における意図された解釈ごとの第 2 要素と第 3 要素のピッチ最大値の差をチャンスレベル (=0) と比較した結果は表 1 の通りであり、ジェスチャー

表 1 各条件・各解釈における第 2 要素と第 3 要素のピッチ最大値の差とチャンスレベルの比較 (** $p < 0.0001$, ** $p < 0.01$)

条件	解釈	df	t	
発話のみ	短単位修飾	58	12.70	***
発話のみ	長単位修飾	53	-1.08	
ジェスチャー共起	短単位修飾	54	11.47	***
ジェスチャー共起	長単位修飾	50	3.05	**
手の動き	短単位修飾	54	13.70	***
手の動き	長単位修飾	53	1.52	
単語記憶	短単位修飾	56	9.48	***
単語記憶	長単位修飾	53	0.94	

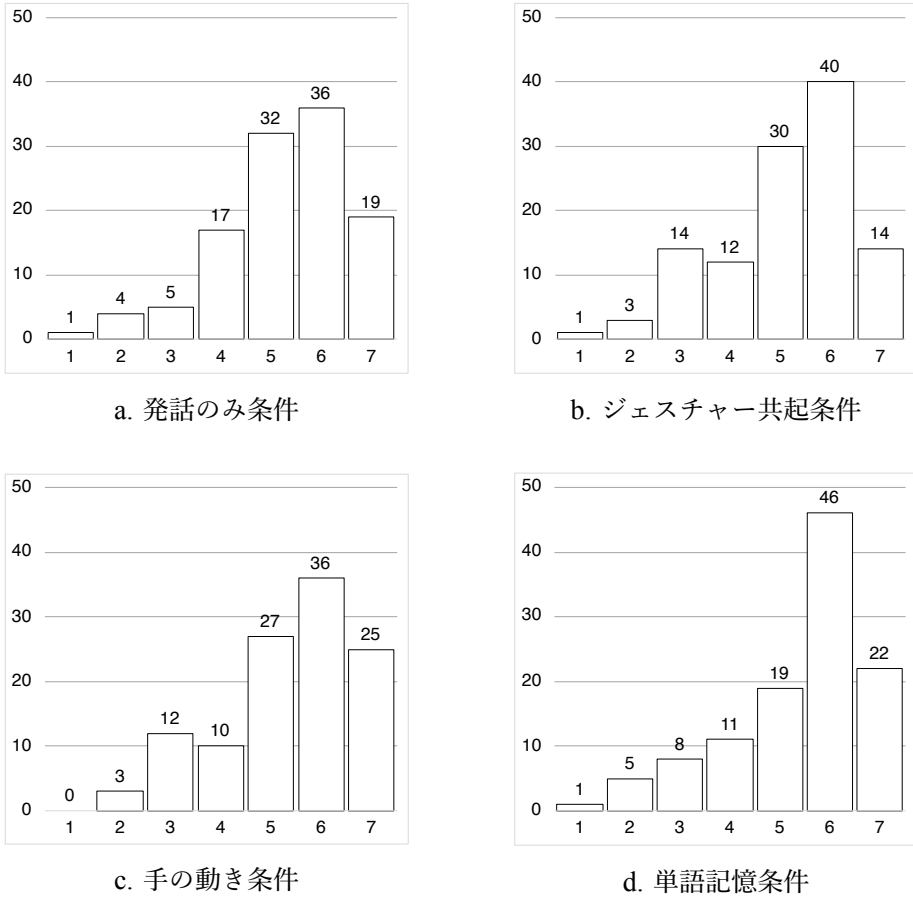


図 13 実験 2 の各条件における自信度の分布

共起条件以外の 3 条件では、短単位修飾型において有意な下降が見られ、長単位修飾型においては有意な上昇・下降が見られていないが、ジェスチャー共起条件においては両解釈において有意な下降が見られた。

また、各条件における自信度の分布については図 13 のようになっており、線形混合効果モデルによる分析の結果、条件の主効果は見られなかった ($F(3, 428.42) = 0.93, p = 0.42$)。この結果は、話し手自身が自覚できるレベルでの負荷がいずれの条件においてもかかっていないことを意味している。

4.5 考察

実験2の結果、長単位修飾型解釈のジェスチャー共起条件をベースラインとした時、第2要素と第3要素の f_0 最大値の下降幅に対して有意な係数となったのは、発話のみ条件と単語記憶条件であった。これは、ジェスチャー共起条件において見られなかった長単位修飾型解釈におけるダウンステップリセットが、単語記憶条件においては発話のみ条件と同様に観察されていることを示している。

また、手の動き条件についてはベースライン(長単位修飾型・ジェスチャー共起条件)に対する影響は有意傾向に留まった。この結果は、手を動かすという身体動作そのものがイントネーションに影響を及ぼしたことを示唆していると考えられることのできるものである。しかしながら、長単位修飾型の解釈かつ手の動き条件における第2要素と第3要素のピッチ最大値の差をチャンスレベル(=0)と比較した場合、その間に有意差が見られなかったことから、ダウンステップの弱化については生じていると解釈できる。このことを踏まえると、手の動き条件について有意差が見られなかったのは、サンプルサイズが小さいことによる可能性も高いと言えよう。先に言及した Bernardis and Gentilucci (2006) において、ジェスチャーは無意味な手の動きとは異なり、イントネーションにポジティブな影響を及ぼしていたことを考慮すると、今回の結果もジェスチャーと無意味な手の動きを単純に身体動作として一括りにするのは早計であると言えるだろう。

以上を踏まえると、統語的曖昧性を区別するジェスチャーを共起させるということは、少なくとも単語の暗記をしながら発話するという認知的負荷や無意味な手の動きが共起する運動的負荷とは異なる可能性が高いと考えられる。これは、統語構造を区別するイントネーションとジェスチャーの役割が認知的に言語的領域を占める要素(=(12b))であることを支持する。

興味深いのは、各発話に対して実験参加者が自己評価した自信度を見ると、ジェスチャー共起条件においても他の3条件と変化が見られなかった点である(図13)。この事実は、ジェスチャーによって統語構造を区別することが過度に認知的負荷をかけているわけではないという仮説を支持するものであると言える。仮に、手の動き条件と単語記憶条件が、極端に運動的・認知的負荷が低く、ジェスチャー共起条件はそれよりも大きな負荷がかかるものであった場合、その負荷の高さを発話者自身が自覚し、発話の自信度として現れることが想定される。しかしながら、上述のように、ジェスチャー共起条件においても自信度は他の3条件と同等の分布を示している。すなわち、ジェスチャーの共起により韻律的区別が行われなくなるのは、発話者自身も意識していないものであると言える。

5 総合考察

これまで、2つの実験を通して統語的曖昧文に伴うジェスチャーがイントネーションに与える影響について検討してきた。実験1の結果、発話のみに集中して統語的曖昧文を音読した場合、短単位修飾型の解釈では見られた第2要素と第3要素の間のダウンステップが長単位修飾型の解釈では見られなかった。一方、発話にジェスチャーを共起させた条件では短単位修飾型の解釈だけでなく、長単位修飾型の解釈においてもダウンステップが見られた。この結果は、ダウンステップによる統語構造の区別は義務的なものではなく、ジェスチャーに影響を受ける選択的なものであることを意味している。

実験1の結果を受け、ジェスチャーがイントネーションに影響を与えたのは、それらが一般的な認知リソースを用いる発話の付加的要素であるのか、認知的に言語的領域を占める要素であるのかを検証する実験2を実施した。その結果、無意味な手の動きや単語の暗記といった一般的な認知的負荷のかかる並行タスクを行った実験条件においてもイントネーションによる統語構造の区別は観察された。この結果は、統語構造を区別するイントネーションとジェスチャーが互いに言語的領域を認知的に占有するものであるため、同時に双方を使用することができなかったという仮説を支持するものであると言える。

実験1と実験2を通して、これまでの先行研究では明らかにされてこなかった統語構造に関わるイントネーションであるダウンステップにジェスチャーがどのような影響を与えるかを明らかにした。先行研究においては、2.3節で述べたようにコミュニケーション的機能を有する韻律的特徴と身体的動作の関係性に着目しており、それらは協調的な振る舞いを見せることが指摘されてきた。しかしながら、統語的解釈に関する機能を持つイントネーションとジェスチャーは共起しづらいことが明らかとなった。

このような結果は、言語習得研究の指摘とも符合するものであると言える。Tomasello (2003)によると、子供の言語習得は、単純に対象と音列を結びつけるのではなく、発話者のコミュニケーション的意図を読み取る能力を基礎としているという社会的・語用論的理論 (social-pragmatic theory) による説明が不可欠である。この説明に従えば、子供の言語習得はコミュニケーション的意図に根ざした言語単位から習得されていくことが想定される。そのような言語習得の最初期においては、イントネーションやジェスチャーがモダリティとして音素列と分化されず、一体として学習されることが予想される。そのため、コミュニケーション的機能を有するイントネーションやジェスチャーは一体となって学習され、互いに

協調的に生起するものとなったと考えられる。

一方、統語構造の区別を行うイントネーションやジェスチャーの場合、その前提として統語的解釈を行うことができる発達段階に到達している必要がある。統語的解釈を行えるようになるのは、(Tomasello and Brooks 1999)によると36ヶ月以降と言われており、一語文を習得する12ヶ月頃よりもかなり遅いと言われている。この点を考慮すると、統語構造を区別するイントネーションとジェスチャーは、語用論的機能を有するそれらのモダリティを含む言語単位とは言語習得の観点から見てもかなり質の異なるものであると考えられる。そのため、統語構造に対応するマルチモーダル要素はモダリティごとにそれぞれ別の言語的要素として習得され、使用されていると考えることができる。

以上のように、本研究の結果は、これまでの言語習得において指摘されてきた人間言語の特徴が他のモダリティにおいても観察されうることを示している。

6 おわりに

本稿では、統語的曖昧文の発話におけるイントネーションとジェスチャーの関係性を2つの実験を行うことによって検討してきた。その結果、発話のみに集中した場合に見られるダウンステップの生起と弱化というイントネーション上の区別は、ジェスチャーが共起することにより実現されなくなることが分かった。また、統語構造を区別するイントネーションとジェスチャーが単なる発話の付加的要素ではなく、認知的に同一の領域を占める言語的要素であることも示唆された。

以上の結果は、コミュニケーション的機能を持つイントネーションとジェスチャーに着目した先行研究とは異なるものであった。言語習得研究においてもコミュニケーション的機能を持つ言語要素と統語的機能を持つ言語要素の習得時期が大きく異なることが報告されており、それは言語以外のモダリティにも影響を及ぼすものであることを今回の結果は示していると言える。

今後の課題として、共起しているジェスチャーのタイプに応じてイントネーションが受ける影響に差が出るかをより詳細に検討することが考えられる。これは、定着しているジェスチャーパターンであれば、その処理が自動化しており、イントネーションの区別に影響を与えずに実行されるという仮説を検証するものである。また、統語的曖昧性を解消するイントネーションとジェスチャーが発話理解にいかなる影響を与えるかについても詳細に検討する必要がある。この点に関しては、現在研究を進めており、稿を改めて詳細に議論したい。

参考文献

- Amano, Nariaki and Kimihisa Kondo. 1998. Estimation of mental lexicon size with word familiarity database. *Proceedings of ICSLP'98* 5: 2119–2122.
- 天野成昭・近藤公久. 2000. 『日本語の語彙特性 第1巻』東京: 三省堂.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–48. DOI: 10.18637/jss.v067.i01.
- Bernardis, Paolo and Maurizio Gentilucci. 2006. Speech and gesture share the same communication system. *Neuropsychologia* 44(2): 178–190.
- Boersma, Paul and David Weenink. 2020. Praat: doing phonetics by computer [Computer program] Version 6.1.15, retrieved 20 May 2020. Technical report.
- 五十嵐陽介・小磯花絵. 2012. 「句末境界音調のピッチレンジに与える要因: 『日本語話し言葉コーパス』の分析」『第2回コーパス日本語学ワークショップ予稿集』: 15–24.
- Kashiwadata, Kei, Tetsuya Yasuda, Koji Fujita, Sotaro Kita, and Harumi Kobayashi. 2020. Syntactic structure influences speech-gesture synchronization. *Letters on Evolutionary Behavioral Science* 11(1): 10–14.
- Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the. In Key, Mary Ritchie (ed.), *The Relationship of Verbal and Nonverbal Communication*. 207–227. The Hague: Mouton.
- Kendon, Adam. 2004. *Gesture*. Cambridge: Cambridge University Press.
- Krahmer, Emiel and Marc Swerts. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57(3): 396–414.
- 窪蘭晴夫. 1995. 『語形成と音韻構造』東京: くろしお出版.
- Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13): 1–26. DOI: 10.18637/jss.v082.i13.
- Lee, Akinobu and Tatsuya Kawahara. 2009. Recent Development of Open-Source Speech Recognition Engine Julius. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*.
- Lee, Akinobu, Tatsuya Kawahara, and Kiyohiro Shikano. 2001. An Open Source Real-Time Large Vocabulary Recognition Engine. In *Proc. EUROSPEECH*. 1691–1694.

- Lenth, Russell V. 2021. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. URL: <https://CRAN.R-project.org/package=emmeans>.
- McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- McNeill, David. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.
- McNeill, David, Elena T Levy, and Susan D Duncan. 2015. Gesture in Discourse. In *The Handbook of Discourse Analysis Volume I*. 2nd edition. 262 Oxford: Wiley Blackwell.
- 岡久太郎. 2022. 「階層構造の外在化における言語外的特徴—統語的曖昧性を解消するジェスチャー」岡ノ谷一夫・藤田耕司(編)『言語進化学の未来を共創する』59–75. 東京: ひつじ書房.
- R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Selkirk, Elisabeth and Koichi Tateishi. 1991. Syntax and downstep in Japanese. In *Interdisciplinary Approaches to Language*. 519–543. Dordrecht: Springer.
- Theune, Mariët and Chris J. Brandhorst. 2010. To beat or not to beat: Beat gestures in direction giving. In Kopp, Stefan and Ipke Wachsmuth (eds.), *Gesture in Embodied Communication and Human-Computer Interaction*. Berlin and Heidelberg: Springer.
- Titze, Ingo R., Ronald J. Baken, Kenneth W. Bozeman et al. 2015. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America* 137(5): 3005–3007.
- Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- Tomasello, Michael and Patricia J Brooks. 1999. Early syntactic development: A construction grammar approach. In *The Development of Language*. 161–190. Hove: Psychology Press.
- Van Hooff, JARAM. 1962. Facial expressions in higher primates. In *Symposium of Zoological Society of London*. 8 97–125.
- Van Hooff, JARAM. 1967. The facial displays of the catarrhine monkeys and apes. In Morris, Desmond (ed.), *Primate ethology*. London: Weidenfield & Nicolson.
- Venditti, Jennifer J. 1994. The influence of syntax on prosodic structure in Japanese. *OSU Working Papers in Linguistics* 44.

Co-occurrence of intonation and gesture to disambiguate syntactic ambiguity

Taro Okahisa Ayako Shirose

We examine how prosodic disambiguation of syntactic ambiguity is related to hand gestures. It is well-known that there is a pattern of intonation to show the syntactic structure of utterances in Tokyo Japanese. In addition, some studies point out that native speakers of Japanese have some gesture patterns for syntactic disambiguation. In this study, focusing on multimodal interaction, we conducted two experiments to illustrate gesture's influence on prosody in syntactic disambiguation. To investigate if gestures interrupt prosodic disambiguation, in Experiment 1, the participants were asked to utter syntactically ambiguous sentences to convey one meaning with or without gestures. The results show that the effect of prosodic disambiguation is smaller in uttering sentences with gestures than without gestures. Based on the result of Experiment 1, we can postulate the following two hypotheses: (I) intonation and gesture for syntactic disambiguation require general cognitive resources and thus there are not enough cognitive resources to produce them simultaneously, (II) they occupy the same linguistic domain in our cognitive system and thus the gestural disambiguation is substituted for intonational disambiguation in the utterance production process. To investigate which hypothesis is the case, we carried out Experiment 2. In Experiment 2, we added two experimental conditions, under which the participants were asked to utter syntactically ambiguous sentences while moving their hands meaninglessly or memorizing three words. As a result, under the additional conditions, they can distinguish the intonation pattern to show the syntactic structure. This result supports the hypothesis (II).