

音声コーパスを用いた強形の can に関する研究

大井川 朋 彦*¹

A Corpus-based Study on “can” in Strong Form

Tomohiko OOIGAWA *¹

In order to confirm the importance of phonological research on *can* and *can't*, the present study examined how frequently *can* in strong form appeared in natural conversations using Buckeye Corpus (Pitt et al. 2007). The study is aimed to be a part of a research project to develop training programs to help language learners to correctly perceive and produce non-native speech sounds which are difficult to learn. Since English is globally used as a foreign/second language (e.g. Kachru 1992), it is important to learn its phonological features. It can be predicted that the English modal verb *can* and its negative contracted form *can't* are frequently used in everyday conversations because they are basic words. Especially in North-American-type English, *can* can be phonetically close to *can't* when *can* is realized as a strong form, and the auditory perception (i.e. identification) of the two words can be difficult for non-native speakers of English (e.g. Ooigawa 2018). According to the corpus analyses, *can* is a frequent modal verb. In a 30-to-60-minute conversation, both *can* and *can't* appear at least twice, and *can* in strong form appears from 0 to 20 times depending on the speaker. Non-native listeners may mis-identify the two words in actual conversations.

1. 導入

1.1 目的

先行研究によると、*can* が強形として実現した際、北米系英語の場合 *can't* と音声的に類似し、両語の聴取による弁別・同定が日本語母語話者などの非英語母語話者には困難になりうるとされる (Ooigawa 2014, 2018, Takahashi and Ooigawa 2012, 2016)。加えて、*can* 及び *can't* は基礎的な語であるため、会話において頻繁に使用されると考えられる。本研究では、音声コーパスを用いて *can* が会話の中でどの程度頻繁に用いられるか、そしてその内どの程度の割合で強形の *can* が出現するかを *can't* と比較しながら検討し、*can* 及び *can't* の正確な弁別・同定の重要性やそのための研究の重要性を検証した。本研究は、非母語の言

語の学習者向けの、母語に存在せず弁別・同定や産出が困難な音韻対立の音声を正しく弁別・同定や産出するための訓練プログラムの開発に、最終的につなげることを目指している。特に英語は、世界中で第二言語や外国語として使用されており (e.g. Jenkins 2007, Kachru 1992)、最も国際的に使用されている言語の1つであると考えられる。従って、その音韻的特徴の知識を持ち、その音声を正しく弁別・同定や産出することは重要であると考えられる。*can* 及び *can't* は会話において多用されると考えられ、加えて、取り違えると全く反対の意味になる重要な語であるのにも関わらず、両語の知覚・産出に関する研究は英語の /l/ 音及び /r/ 音などの他の音声・音韻的な研究と比べると、著者の知る限り、あまり行われていない。加えて、通訳業務などの実務において英語を運用

*1 日本大学国際関係学部国際教養学科 助教 Assistant Professor, Department of Liberal Arts, College of International Relations, Nihon University

する中で非英語母語話者が両語の区別に困難さを抱えているという報告もあり (Takahashi and Ooigawa 2012, 2016)、両語の音声学・音韻論的な研究を行い、それらの特徴を明らかにすることは、そのような国際的に活躍する実務家の抱える問題を解決することにつながる可能性もある。

1.2 can 及び can't

英語の法助動詞 can の否定形は cannot で、その縮約形は can't である。英語発音辞典 (Jones et al. 2011, Wells 2008) によると、米英語における can の強形の発音は [kæ̃n]、弱形は [kən]、[kŋ] などと記載されている。一方、can't は強形のみで、[kæ̃nt] と表記されている。法助動詞を含めた多くの機能語には強形と弱形の2種類の発音形式があるとされ、強形の語は強勢を伴い、弱形では伴わない (Collins and Mees 2008, Cruttenden 2008, Ladefoged 2006)。通常、can は文中では弱形もしくは強形で、文末では強形で発音されるのに対して、can't は常に強形で発音されるとされる (Collins and Mees 2008, Cruttenden 2008)。加えて、can't の /t/ が知覚可能な解放音を伴わない音声として実現することや、[t] や [tʰ] や [ʔ] などの /t/ の実現形を全く伴わないで実現する可能性も指摘されている (Ooigawa 2014, Shockey 2003, Takahashi and Ooigawa 2012, 2016, Wells 2008, 佐々木 1993)。表1はここまで述べられた米英語における can 及び can't の音声的な特徴を示している。強形の can 及び can't は共に [kæ̃n] として実現する可能性があり、聴取による弁別・同定が困難であることが予想できる。複数の研究で両語の弁別・同定が非英語母語話者にとって困難であることが示されており (Ernestus et al. 2017)、特に北米系英語の can が強形で実現した場合、顕著である (Ooigawa 2014, 2018, Takahashi and Ooigawa 2012, 2016)。しかしながら、実際の会話では、どのくらい両語がどの発音形式で出現し、これらを混同する可能性があるのかを示した研究は皆無である。そこで、本研究では音声コーパスを用いて北米系英語において強形の can がどの程度出現するかを検討した。

表1：米英語の can 及び can't の音声的特徴

	can	can't
弱形	[k(ə)n]	なし
強形	[kæ̃n]	[kæ̃n(t)]

1.3 コーパス

著者の知る限り、強形の can に着目し音声コーパスを用いた分析を行った研究は皆無である。本研究では、自然会話に近い会話が収録されていると考えられる Buckeye Corpus of Conversational Speech (以下、Buckeye Corpus) (Pitt et al. 2007) を用いて、can 及び can't が日常会話で頻繁に使用されるということを検証し、どの程度の割合で強形の can が出現するかを can't と比較しながら検討し、両語の弁別・同定の重要性について検討した。このコーパスは米オハイオ州出身の英語母語話者40名の30分から60分のインタビュー形式の会話音声収録されているデータベースであり、総語数に換算すると約30万語である。会話音声に加え、各語の正書法による表記、音声表記、文法的な役割を示すタグなどが記載されている。そのため、弱形・強形の can 及び can't の出現数、つまり、トークン数 (延べ語数) を検討するためのコーパスとして適切であると判断した。

2. 分析

2.1 コーパス全体の中の can

Buckeye Corpus には様々な種類の電子ファイルが含まれているが、その中で各語の話者、正書法による表記、音声表記、文法的な役割を示すタグを一度に確認できる modified_dot_words_files を分析に使用した。コーパスの音声は語または <LAUGH> (笑い) や <SIL> (無音) などの non-speech に分類されている。<LAUGH-can_be_hard> (笑いながら “can be hard” と言った) のような記録も存在したが、non-speech には音声表記や文法のタグ付けがなされていないため分析対象としなかった。

コーパスに収録されている中で non-speech に分類されなかった全ての語の総トークン数 (延べ語数) は 286,994 であった (表2参照)。その内、

can が 735 (0.256%)、can't が 425 (0.148%)、cannot が 5 (0.002%) であった。これら 3 つの語の合計は 1,165 で全体の約 0.4% であった。このことから、can 及び can't が頻繁に使用される語であるかは判断できないが、極端に cannot のトークン数が少なく、会話では主に can 及び can't が使用されることが判る。

次に can と同じ文法範疇である法助動詞 (法助動詞) の総数と比較した。このコーパスでは、法助動詞 (modal) は MD という文法タグで分類されるが、それは原形のもののみである。肯定の縮約形 (would の 'd) や否定の縮約形 (can't や shouldn't 等) は MD に含まれず、これらは他の種類の語と共に複数の別のタグに分類されているため、分析対象から除外した。原形の法助動詞 (would、can、should 等) の総トークン数は 2,794 で全体の約 1% であり (表 3 参照)、can 単独で 0.26% であるため、can は法助動詞の中では頻繁に使用されるものであることが判る。次に、他の法助動詞との比較を行った。

表 2 : コーパスの総トークン数と can を含む語のトークン数

	トークン数	%
全体	286,994	
can	735	0.256
can't	425	0.148
cannot	5	0.002
can系	1,165	0.406

表 3 : 法助動詞 (原形) の総トークン数と can のトークン数

	トークン数	%
全体	286,994	
法助動詞	2,794	0.97
can	735	0.26

2.2 他の法助動詞との比較

原形の法助動詞の総トークン数 2,794 の内、would が 903 (32.3%)、can が 735 (26.3%)、could が 395 (14.1%) であり、can は would に次いで 2 番目に多く使用された法助動詞であったことが示された (表 4 参照)。割合 (%) を見

ても、1 位の would と 2 位の can のトークン数の差 (168) と 2 位の can と 3 位の could のトークン数の差 (340) を比較しても、上位 2 語が圧倒的に多く使用されたことが判る。しかしながら、would には肯定の縮約形 ('d) が存在するため、実際には 1 位と 2 位の差は更に大きい可能性がある。加えて興味深いことに、would 及び can のみで全体の過半数であり、他の基礎的な法助動詞、should、will、might、may、must、shall よりもかなり頻繁に使用されていることが判った。このことから、can の重要性だけでなく would の重要性も示唆された。

表 4 : 法助動詞 (原形) のトークン数の比較

法助動詞	トークン数	%
would	903	32.3
can	735	26.3
could	395	14.1
should	261	9.3
will	231	8.3
might	153	5.5
may	54	1.9
must	33	1.2
ought	12	0.4
gotta	9	0.3
hafta	4	0.1
shall	2	0.1
usetta	2	0.1
合計	2,794	

2.3 話者 (インタビュー) ごとの比較

1.3 でも述べたように、Buckeye Corpus には米英語母語話者 40 名の 30 分から 60 分のインタビュー形式の会話音声収録されている。各話者、つまり、1 インタビューごとの can 及び can't のトークン数を分析した。can のトークン数では、最小値が 4 で、最大値が 43 で、平均値が 18.4、中央値が 15.5、標準偏差が 10.0 であった (表 5 及び図 1 参照)。一方、can't では最小値が 2、最大値が 35、平均値が 10.6、中央値 8.5、標準偏差が 7.8 であった。この結果から、どの話者もインタビュー中に 4 回は can を、2 回は can't を使用したということと、最小値と最

大値の差や、標準偏差が大きいことから、can は can't に比べ、使用される頻度が話者によって異なることも示された。更に、中央値から判断すると、米英語母語話者と30分から60分間会話をすると、can は15回前後、can't は8回前後、産出されるという可能性が高いことがうかがえる。

表5：話者（インタビュー）ごとのトークン数の比較

	can	can't
最大値	43	35
平均値	18.4	10.6
中央値	15.5	8.5
標準偏差	10.0	7.8
最小値	4	2

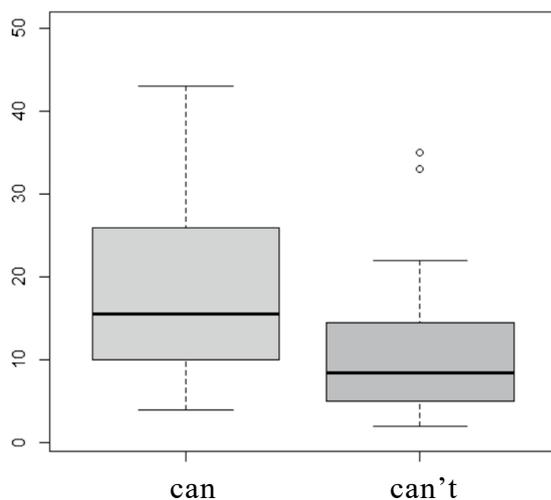


図1：話者（インタビュー）ごとのトークン数の比較

2.4 強形の can

2.4.1 強形の定義（can 及び can't の母音）

Buckeye Corpus には強勢の有無が示されていないため、母音の質で強勢の有無を判断した。表6は10トークン以上確認できた can 及び can't の母音（または音節主音的鼻音）を一覧にしたものである。1.2で述べたように、英語発音辞典（Jones et al. 2011, Wells 2008）によると、can の強形は [kæn]、弱形は [kən]、[kŋ] と記載されている。一方、can't は強形のみで、[kænt] と表記されている。can't のトークンのほとんどが [æ] を含み、[e] の場合もある程度存在するため、

can のトークンの内、[æ] または [e] を含んでいるものを強形とし、それ以外を弱形とした。尚、Buckeye Corpus において使用されている音声表記は国際音声記号（IPA）によるものとは異なるため、IPA（1999）及び Ladefoged（2006）を参考に IPA による表記も掲載した。

英語の [a] は強勢のない様々な母音を示すため、コーパス上では can の母音は様々な表記として現れたと考えられる。コーパス上の [ɪ]、[ʌ]、[ɪ] は発音辞典上の [ə] に相当すると考えられる。音節主音的鼻音の [ŋ] は辞典と同様であるが、[ŋ] に関しては後続する子音（調音場所が共通する [k] や [g]）の影響で同化現象を起こしたと考えられる。加えて、can の母音として [ʌ] ではなく、[ɪ] が最多であったことは大変興味深い。can の [k] が口蓋化している可能性が考えられる。

表6：can 及び can't の母音比較
(その他は10トークン以下のもの)

can			
Buckeye	IPA	トークン数	%
ih	ɪ	389	52.9
en	ɛ̃	116	15.8
eh	ɛ	76	10.3
ah	ʌ	57	7.8
ae	æ	45	6.1
eng	ɪ̃	13	1.8
ihn	ĩ	11	1.5
その他		28	3.8
合計		735	

can't			
Buckeye	IPA	トークン数	%
ae	æ	251	59.1
aen	æ̃	133	31.3
eh	ɛ	19	4.5
その他		22	5.2
合計		425	

2.4.2 強形の can の割合

2.4.1で示されたように、can のトークンの内、[æ] または [e] を含んでいるものを強形とし、それ以外を弱形とした。その定義の下でそれぞれの

トークン数を計算した (以後、弱形は can、強形は CAN とする)。表 7 が示すように、can は 614 (52.9%)、CAN は 121 (10.4%)、can't は 425 (36.6%) のトークン数だった。cannot を除き、法助動詞 can が含まれる語の内、10 回に 1 回は CAN であると言える。CAN は can や can't に比べてトークン数が少なかったが、同じく基本的な法助動詞である may、must、shall よりも多く出現した (2.2 参照)。

肯定形に限定してみると、83.5% は can、16.5% は CAN となる (表 8 参照)。つまり、肯定形の内、7 回に 1 回は CAN であったという計算になる。

更に、強形に限定して検討してみると、CAN は 22.2%、can't は 77.8% となる (表 9 参照)。つまり、強勢を伴った法助動詞 can を含む語の内、5 回に 1 回は CAN であったということが言える。

表 7 : can (弱形)、CAN (強形)、can't のトークン数の比較

	トークン数	%
can	614	52.9
CAN	121	10.4
can't	425	36.6
合計	1,160	

表 8 : can (弱形)、CAN (強形) のトークン数の比較

	トークン数	%
can	614	83.5
CAN	121	16.5
合計	735	

表 9 : 強形だけのトークン数の比較

	トークン数	%
CAN	121	22.2
can't	425	77.8
合計	546	

2.4.3 話者 (インタビュー) ごとの比較

1.3 でも述べたように、Buckeye Corpus には英語母語話者 40 名の 30 分から 60 分のインタビュー形式の会話音声収録されている。各話者、つまり、1 インタビューごとの can (弱形)、CAN (強形)、can't のトークン数を分析した。can のトークン数では、最小値が 2 で、最大値が 39 で、平均値が 15.4、中央値が 12.0、標準偏差が 9.0 であった (表 10 及び図 4 参照)。CAN では、最小値が 0 で、最大値が 20 で、平均値が 3.0、中央値が 2.0、標準偏差が 3.5 であった。can't では、最小値が 2 で、最大値が 35 で、平均値が 10.6、中央値が 8.5、標準偏差が 7.8 であった。中央値から判断すると、米英語母語話者と 30 分から 60 分間会話をすると、can は 12 回前後、CAN は 2 回前後、can't は 8 回前後、産出されるという可能性が高いことがうかがえる。最大値と最小値の差や標準偏差に着目すると、can の出現数は個人差が最も大きく、次いで can't で、CAN が最も小さいことが判る。しかしながら、全く CAN を産出しない話者もいれば、20 回産出する話者もいるため、絶対的に個人差が小さいとは言えない。

表 10 : can、CAN、can't の話者 (インタビュー) ごとのトークン数の比較

	can	CAN	can't
最大値	39	20	35
平均値	15.4	3.0	10.6
中央値	12.0	2.0	8.5
標準偏差	9.0	3.5	7.8
最小値	2	0	2

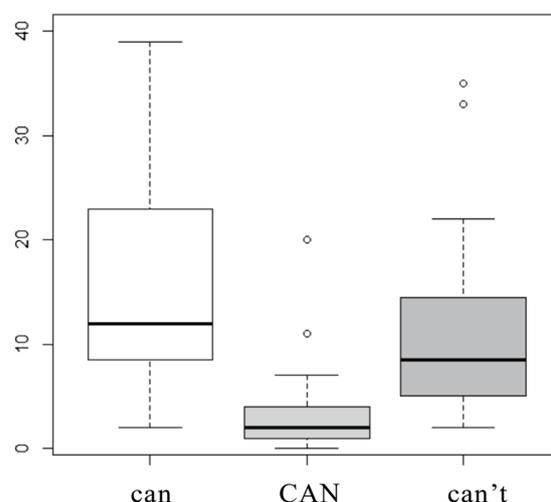


図 4 : can、CAN、can't の話者 (インタビュー) ごとのトークン数の比較

3. 考察

2.1 及び 2.2 によると、can は would に次いで 2 番目に頻繁に使用される法助動詞であり、他の基礎的な法助動詞の中でも特によく使用されるものであることが判った。加えて、cannot はほとんど使用されず、会話では基本的に can't として現れるということが判った。2.3 によると、米英語母語話者とある程度の長さの会話を行うと can 及び can't が両方とも使用される可能性が高いことが示唆された。2.4.2 によると、can 及び can't の出現数の内、半分は弱形の can であり、3 割強は can't であり、強形の can は 1 割程度であることが判った。肯定形に限定すると、8 割強が弱形の can、2 割弱が強形の can であることが判った。加えて、強勢を伴う、つまり強形として実現されるものに限定すると、約 8 割が can't であり、残りの約 2 割は強形の can であることが判った。例えば、日本語母語話者を聴取者として想定し、can't の t の実現形に関しては無視し、短く「キャン」と聞こえれば can として認識し、長く「キャン」と聞こえれば can't と認識する方略で聴取した場合、「キャン」の 2 割は can't ではなく強形の can なので、「キャン」が 10 回産出された場合、2 回は反対に取り違えてしまう可能性が考えられる。

更に、話者ごとに、つまり、1 インタビューごとに分析してみると (2.4.3 参照)、個人差も認められるが、30 分から 60 分の会話において、弱形の can が 12 回前後、強形の can が 2 回前後、強形の can't が 8 回前後、産出される可能性が高いことが示唆された。以上のことから、can 及び can't は比較的頻繁に使用される語であり、強形の can もある程度使用されるため、実際の会話中に非母語話者が強形の can と can't を混同するリスクがある程度存在することが明白となった。話者によって強形の can を使用しないということは、反対に混同のリスクを高める可能性も考えられる。従って、can 及び can't を正確に弁別・同定することやそのための研究は重要であると考えられる。

興味深いことに、今回の分析対象となった

286,994 トークン中、shall は 2 トークンのみであった (2.2 参照)。これは分析に用いたコーパスを作成するときに用いたインタビューというデータ収集方法に起因する可能性が考えられる。従って、データ収集方法が異なるコーパスでは法助動詞の頻度に変化する可能性があり、今後はこの研究で得られた結果が他のコーパスの分析結果においても一致するかどうかを検討する必要がある。

can は弱形で実現することが多く、can't は通常、強形で実現し、/t/ の音声解放音を伴わないことが多いため、「キャンは、弱ければ / 短ければ can、強ければ / 長ければ can't」といった聴取に関する指導法がある。しかしながら強形の can もある程度出現するため、その指導法に従うと、誤って同定してしまうことがあるということが今回のコーパス分析で明らかとなった。つまり、強形として聞こえるものを全て can't と同定した場合、その内 22.2% は CAN であるため、5 回に 1 回は誤って同定してしまうということである (2.4.2 参照)。今後は様々な指導法を比較・検討し、より高い確率で弁別・同定可能な指導法を模索していく必要がある。

加えて、本研究では法助動詞の would の重要性も示した (2.2 参照)。この語は最も頻繁に使用される法助動詞であり、今回の分析対象から外された 'd や否定形を含めると更に多く使用されていると考えられる。このことから、would は使用用途が広く、非母語話者も聴取する可能性が高いと思われるので、英語教育での同法助動詞のあり方を再考する必要があると考えられる。

4. 結論

can が強形として実現した際、北米系英語の場合 can't と音声的に近いものになり、両語の聴取による弁別・同定が日本語母語話者などの非英語母語話者には困難になりうる (Ooigawa 2014, 2018, Takahashi and Ooigawa 2012, 2016)。本研究では、Buckeye Corpus (Pitt et al. 2007) を用いて can が会話の中でどの程度頻繁に用いられるか、そしてその内どの程度の割合で強形の can

が出現するかを can't と比較しながら検討した。その結果、can は would に次いで 2 番目に頻繁に使用される法助動詞であり、can 及び can't は比較的頻繁に使用される語であり、強形の can もある程度使用されることが判明した。それにより、実際の会話時に非母語話者が強形の can と can't を混同するリスクがある程度存在することが示された。従って、can 及び can't の正確な弁別・同定やそのための音声学・音韻論的な研究は重要であると考えられる。

参考文献

- Collins, Beverley S. and Inger M. Mees. 2008. *Practical Phonetics and Phonology: A Resource Book for Students*. 2nd ed. New York: Routledge.
- Cruttenden, Alan. 2008. *Gimson's Pronunciation of English*. 7th ed. London, UK: Hodder Education.
- Ernestus, Mirjam, Huib Kouwenhoven and Margot van Mulken. 2017. "The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language." *Journal of Phonetics*, 62: 50-64.
- IPA. 1999. *Handbook of the International Phonetic Association*. Cambridge University Press.
- Jenkins, Jennifer. 2007. *English as a Lingua Franca: Attitude and Identity*. Oxford/New York: Oxford University Press.
- Jones, Daniel, Peter Roach, Jane Setter and John Esling. Eds. 2011. *Cambridge English Pronouncing Dictionary*. 18th ed. Cambridge, UK: Cambridge University Press.
- Kachru, Braj B. 1992. "Teaching World Englishes." In *The other tongue: English across cultures*. Ed. by Braj B Kachru. 2nd ed. Urbana and Chicago: University of Illinois Press.
- Ladefoged, Peter. 2006. *A Course in Phonetics*. 5th ed. Boston, USA: Thomson Wardsworth.
- Ooigawa, Tomohiko. 2014. "Perception of American English Utterance-Final "can" and "can't" by Japanese Listeners." *Sophia Linguistica*, 62: 31-43.
- Ooigawa, Tomohiko. 2018. "Perceptual Learning of Syllable-final Contrast: Perception and Training of Various "can" and "can't" by Japanese, Chinese and French Listeners." Doctoral thesis, Sophia University.
- Pitt, Mark, L. Dilley, Keith Johnson, Scott Kiesling, Raymond William, Elizabeth Hume and E. Fosler-Lussier. 2007. *Buckeye Corpus of Conversational Speech (2nd release)* [www.buckeyecorpus.osu.edu]. Columbus, OH: Department of Psychology, Ohio State University (Distributor) .
- Shockey, Linda. 2003. *Sound Patterns of Spoken English*. Oxford, UK: Blackwell.
- Takahashi, Kinuko and Tomohiko Ooigawa. 2012. "Perception of American English "can" and "can't" by Japanese professional interpreters." *Interpreting and Translation Studies*, 12: 249-262.
- Takahashi, Kinuko and Tomohiko Ooigawa. 2016. "Perception of American and Australian English "can" and "can't" by Japanese listeners: How to teach "can" and "can't"." *Proceedings of ISAPh2016 (the 1st International Symposium on Applied Phonetics) : Diversity in Applied Phonetics (ISBN: 978-4-9909245-0-8)* : 116-121.
- Wells, John C. Ed. 2008. *Longman Pronunciation Dictionary*. 3rd ed. Essex, UK: Pearson Education Limited.
- 佐々木めぐみ. 1993. "アメリカ英語における can と can't の音響的違いについて." *Sophia University Working Papers in Phonetics*: 3-10.