

六種類の可算性を用いた英語名詞句の可算／不可算の決定

3B-2

Francis Bond 小倉 健太郎

NTT 情報通信網研究所

1 まえがき

従来、機械翻訳等の英文生成では、可算性については辞書的情報としての名詞の可算性のみが考慮されてきたが、生成の精度に問題があった。我々は、運用上の可算性として、名詞句の可算性を新たに導入することにより精度の良い生成を可能にした。

本論文では、自然言語処理の観点から英語の一般名詞に関連した名詞句の可算性と名詞の可算性について述べる。まず、名詞句の可算性について名詞と決定詞の関係をj用いて定義する。次に、可算名詞句、不可算名詞句の表現の違いに着目して、名詞の可算性を六種類に分類し、その可算性を判定する方法を示す。最後に、日英機械翻訳システムに、これらを導入した効果について報告する。

2 名詞句の可算性

英語では、名詞句が指し示す概念を数える必要があるかどうか(名詞句の可算性)によって、名詞句の表現法が変わる。数える必要がある名詞句(可算名詞句)は、可算決定詞¹を付けて表現する。数える必要がない名詞句(不可算名詞句)は、名詞そのもので表現する。

名詞句の可算性に関する研究としてAllan^[1]がある。Allanの定義は以下の通りである。

I 名詞句の主要語が可算決定詞のスコープに入れば可算とする。

II 名詞句の主要語が複数であれば可算とする。

しかしIIの定義を使用すると可算性の概念がないのに可算名詞句になる場合がある。(例えば、*Elephants are mammals*の*elephants*は種全体のことを示し、可算性の概念がない。)これを防ぐために我々は、名詞句の可算性をAllanのIのみを用いて決定する。

名詞句の可算／不可算と数の関係について、以下のルールを提案する。

A 可算名詞句の主要語の数は、可算決定詞により決める。

B 不可算名詞句の主要語の数は、総称の*all*を付けた場合と同じにする。

(総称の*all*は*All elephants are mammals*や*All scissors have two blades*等の*all*である。)

3 名詞の可算性の分類

従来、可算性は、一般名詞の属性としての[±Count]が

¹可算決定詞(denumerators)は*one, two, three*等の自然数と*both, dozen, score*等の数量表現と、*a/an, each, another, every, either, neither, many, several, (a) few(er)*の数量詞がある。

Determination of whether an English noun phrase is countable or not using 6 levels of lexical countability.

Francis Bond and Kentaro Ogura.

NTT Network Information Systems Laboratories.

ある^[2]。しかし、多くの名詞は可算として不可算としても使える。例えば、*cake*は、*I eat cake*では不可算であるが、*I eat a cake*では可算である。可算性を名詞の2値属性とすると、意味が同じなのに、二つの名詞*cake[-Count]*と*cake[+Count]*を辞書に登録する必要がある。さらに、単数か複数の一方しかない名詞もあり、不定冠詞しかとれない名詞もある。そこで、我々は名詞の可算性を可算／不可算の二段階だけでなく、絶対可算、強可算、弱可算、不可算、絶対複数、半可算の六種類に分類した。(表1)

表1：名詞の六種類の可算性と名詞句表現

名詞の可算性	可算名詞句	不可算名詞句
絶対可算	an elephant	elephants
強可算	a cake, a slice of cake	cakes (cake)
弱可算	a cup of beer, a beer	beer (beers)
不可算	a piece of equipment	equipment
絶対複数	a pair of scissors	scissors
半可算	a knowledge, (2 kinds of ~)	knowledge

絶対可算名詞は可算決定詞を付けて可算名詞句を表現する。不可算名詞句の場合は複数形とする。(例えば、可算名詞句は*an elephant, two elephants*、不可算名詞句は*elephants*)

その反対は不可算名詞である。単複の区別がなく、いつも単数形である。不可算名詞は、可算決定詞を付けても、可算名詞句になれない。そのかわりとして、分類辞(classifier: *piece, pair*等)を使用することにより、可算名詞句を表現できる。(例えば、*a piece of equipment, two pieces of equipment*等)不可算名詞句の場合は、そのままの形で表現する。

ところで、多くの名詞は、可算／不可算両方の用法がある。例えば*cake*は、ケーキそのものを考えるとき不可算名詞として使うが、一個一個の個別のものとして考えると可算名詞として使う。その二つの使用と名詞句の二つの可算性を合わせると以下の四つの表現が可能である。(表2)

表2：可算／不可算両用名詞の名詞句表現

名詞	可算名詞句	不可算名詞句
可算として	2 cakes	cakes [are delicious]
不可算として	2 slices of cake	cake [is delicious]

Huddleston^[3]によると、全ての名詞は両方の使い方があり、可算か不可算かは名詞の可算性の程度(以下「可算度」と略す)と名詞句の決定詞と話者に依存する。可算度は連続的で、絶対可算名詞はただ可算度が高く、不可算名詞は可算度が低いものと考えることができる。しかし、可算度を連続量として扱うのは難しく、あまり細かく分類しても大きな結果は期待できないので、絶対可算と不可算の間に強可算と弱可算の二つのレベルを設定した。判定する情報がないなら、強可算名詞を可算名詞として、弱可算名詞を不可算名詞として扱う。

絶対複数と半可算の二つの種類がある。絶対複数名詞は不可算名詞のように、可算名詞句として使われる場合は分類辞が必要である。ただし、いつも複数で表現される。(例えば、

scissors, a pair of scissors)

最後に、半可算名詞はほぼ不可算名詞と同じであるが、可算名詞句を表現する場合、可算決定詞の *a/an* が付与できる。それ以外の可算決定詞なら分類辞を使うが、不定冠詞の場合はそのままいい。(例えば、*a knowledge, two kinds of knowledge*)

4 名詞の可算性の判定法

名詞の可算性は名詞の属性であり、辞書情報に付与する。その判定法を以下に述べる。

まず、次の三つの判定で非文であるかないかをテストし、絶対可算、強・弱可算、不可算、絶対複数と半可算に分ける。

表3: 名詞の可算性の判定法

名詞の可算性	(1) 複数になれる	(2) 'a/an' が付与できる	(3) 可算決定詞が付与できる
絶対可算	+	+	-
強・弱可算※	+	+	+
不可算	-	-	+
絶対複数	+	-	-
半可算	-	+	+

※ 強・弱可算名詞は両方とも可算と不可算の使用があるので文法的なテストだけでは判定できない。判定法はコーパスを使って、可算の使用が多いなら強にして、不可算の方が多いなら弱にすることが理想的であるが、ネーティブが判定することも可能である。

(1) 複数になれる ※ (*は非文を示す)

絶対可算	Elephants are dangerous
強・弱可算	Cakes are dangerous
不可算	* Equipments are dangerous
絶対複数	Scissors are dangerous
半可算	* Knowledges are dangerous

※ 単複が同じ形の名詞のために (*sheep* 等) 動詞との数の一致も考慮する。

(2) a/an が付与できる (*は非文を示す)

絶対可算	An elephant can be dangerous
強・弱可算	A cake can be dangerous
不可算	* An equipment can be dangerous
絶対複数	* A scissors can be dangerous
半可算	A knowledge can be dangerous

(3) 不可算決定詞が付与できる ※ (*は非文を示す)

絶対可算	* Much elephant can be dangerous
強・弱可算	Much cake can be dangerous
不可算	Much equipment can be dangerous
絶対複数	* Much scissors can be dangerous
半可算	Much knowledge can be dangerous

※ 不可算決定詞の *much, less, (a) little* は不可算、強・弱可算と半可算名詞にしか付けられない。

このテストを利用すると、言語学に詳しくない英語のネーティブでも、名詞の可算性が判断できる。それにより辞書の作成が簡単になる。

5 日英機械翻訳の英文生成への適用

英語を生成するとき決定詞、分類辞、数は、名詞句の可算性に依存する。以下、日英機械翻訳システム ALT-J/E^[4] におけ

る名詞句の可算性を用いた英文生成について述べる。

名詞句の可算性は名詞の修飾語によって決定できる。日本語の助数詞は、数量表現であるから、それがついている名詞句は可算として訳す必要がある。可算名詞は数字を付けてそのまま訳すが、不可算名詞などは分類辞を付けて訳す。接尾辞の「つ」や「個」は、一つ一つに分かれている意味があるので、強・弱可算名詞を可算として訳す。「切れ」等は大きなものを小さく分ける意味なので強・弱可算名詞を不可算として訳す。

「多くの」や「たくさんの」は特に可算/不可算の情報をもっていないので、強可算名詞は可算として訳すが、弱可算名詞は不可算として訳す。他の種類は数によって '*many*' か '*much*' を選ぶ。

可算性	日本語訳	英語
絶対可算	二頭の象	two elephants
強可算	二つのケーキ	two cakes
弱可算	二つのビール	two beers
不可算	二つの家具	two pieces of furniture
絶対複数	二つのハサミ	two pairs of scissors
半可算	二つの知識	two kinds of knowledge
強可算	二個のケーキ	two cakes
強可算	二切れのケーキ	two slices of cake
絶対可算	多くの象	many elephants
強可算	多くのケーキ	many cakes
弱可算	多くのビール	much beer
不可算	多くの家具	much furniture
絶対複数	多くのハサミ	many pairs of scissors
半可算	多くの知識	much knowledge

また、一般知識によって、可算/不可算の決定ができる。例えば、総称的名詞句は、クラスの概念を示しているので不可算名詞句として訳す。(例えば、象は動物だ→*elephants are animals*; 椅子は家具だ→*chairs are furniture*) また、だ文 (copula) では、主語が可算名詞句であれば、述部も可算名詞句とする。(例えば、全てのコンピュータはハードウェアだ→*every computer is a piece of hardware*)

ここで示した名詞の六種類の可算性と名詞句の可算性と、名詞句の機能的区別の導入によって、新聞記事の例文で、英語名詞句の数と冠詞の正解率を76%から93%にあげることができた。

6 あとがき

本論文は、英語の名詞句の可算/不可算表現を、六種類の名詞の可算性を用いて、決定する方法について述べた。また、名詞の可算性の判定法を示した。最後に、それらを日英機械翻訳に適用した場合の有効性について示した。

<参考文献>

- [1] Allan, Keith. 1980. Nouns and countability. *Language* 56. 541-67. Baltimore: Waverly Press.
- [2] Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge MA: MIT Press.
- [3] Huddleston, Rodney. 1984. *Introduction to the Grammar of English*. Cambridge: University Press.
- [4] 白井諭他. 1991. 前編集不要の日英機械翻訳の実現に向けて. *NTTR&D* vol 40, no 7. 897-904.