

# Building a Reference Lexicon for Countability in English

**Tibor Kiss**  
Ruhr-Universität Bochum  
Bochum, Germany  
tibor@linguistics.rub.de

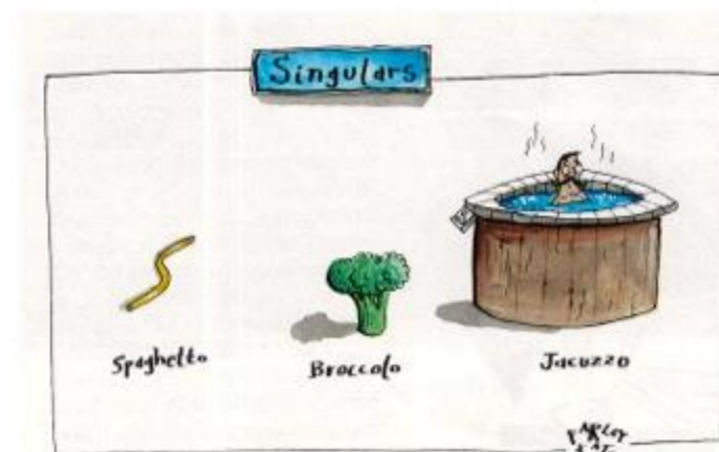
**Francis Jeffry Pelletier**  
Simon Fraser University  
Vancouver, Canada  
jeffpell@sfu.ca

**Tobias Stadtfeld**  
Ruhr-Universität Bochum  
Bochum, Germany  
stadtfeld@linguistics.rub.de

## Lexical Countability

Research on lexical countability of nouns usually considers only a very small set of prototypical nouns

→ neglects different readings and variations of the countability of a noun by reinterpretation or sorter/ container readings:



- a. This recipe requires an apple. (count)  
b. Put more apple into the parfait. (mass)

- a. He ordered a beer. (a container filled with beer)  
b. The waiter recommended two different wines. (two sorts of wine)

There seems to be more a “degree” of countability than an absolute distinction between count and mass, the dualism “countable – uncountable” is insufficient

**In contrast:** questionnaire-based approach focusing on a larger set of nouns to achieve a fine-grained distributional pattern of different classes of the countability of nouns

## Tests and Annotation

**Data** = set of nouns extracted from Open ANC with frequency >10 and senses provided by WordNet

**Annotators** = four native speakers of English (Canadian) judging the nouns with a set of countability pattern tests (list of nouns + senses)

~ 14,000 nouns in total

training phase of 1,000 nouns annotated by all four annotators to identify possible problems

→ other nouns mostly annotated by at least two annotators

## Countability Pattern Tests

restrictive patterns, main possible choices by the annotators: yes, no, na (not applicable) comments for problematic cases and possibility to mark „don't know“

→ results form a set of distinct countability classes

**Test I.1** NP1 VERB more NOUN[sg] than NP2.

- a. The boy ate more fruitcake than the girl.  
b. \*John bought more car than Bill.

**Test I.2** if I.1 is grammatical, determine mode of measurement

- a. intensity/ quantity etc. = not number  
b. number of items = number

**Test II.1** NP1 VERB more NOUN[p] than NP2.

**Testing semantic equivalence of:**

- a. He drank more **whiskeys** than her.  
b. He drank more **kinds/ glasses of whiskey** than her.

**Test II.2** NP1 VERB more CLASSIFIER of NOUN[sg] than NP2.

**Test III.1** [NPIND.DET.+NOUN[sg]] is {SOME PROPERTY OF NOUN}

- a. A car is a vehicle.  
b. \*Car is a vehicle.

**Test III.2** NOUN[sg] is {SOME PROPERTY OF NOUN}

- a. # A whiskey is a glass filled with whiskey.  
b. Whiskey is a drinkable liquid containing alcohol.

## Distribution of Countability Classes in BECL 2.0

Countability Class	Test I.1	Test I.2	Test II.1	Test II.2	Test III.1	Test III.2	Frequency
<b>regular count senses</b>							
235	no	na	yes	not equiv.	yes	no	8025
721	no	na	yes	not equiv.	yes	yes	7
371	na	na	yes	na	na	na	7 (group only)
73	no	na	yes	not equiv.	no	no	3
<b>regular mass senses</b>							
528	yes	not num.	na	na	no	yes	1866
519	yes	not num.	no	na	no	yes	25
531	yes	number	na	na	no	yes	12
<b>both mass and count</b>							
510	yes	not num.	yes	equiv.	no	yes	314
726	yes	not num.	yes	not equiv.	yes	yes	162
729	yes	number	yes	not equiv.	yes	yes	3 (LS&MJ only)
513	yes	number	yes	equiv.	no	yes	1(FW&MD only)
<b>neither mass nor count</b>							
523	no	na	na	na	no	yes	146
37	no	na	na	na	no	no	59
190	no	na	no	na	yes	no	11
514	no	na	no	na	no	yes	9 (group only)
199	no	na	na	na	yes	no	9
28	no	na	no	na	no	no	4
353	na	na	no	na	na	na	4

## An Interesting Gap

Unspecified	bias <sub>2</sub> fate <sub>1</sub> fate <sub>3</sub> tail <sub>2</sub>	a line or cut across a fabric that is not at right angles to a side of the fabric an event (or course of events) that will inevitably happen in the future your overall circumstances or condition in life the time of the last part of something
-------------	--	--

## Dual-life Nouns

The distribution of countability classes also shows not only a number of different „countability degrees“, but also a category of nouns that show different aspects of countability depending on sense and context:

**fruitcake #1 : countable vs. fruitcake #2 : substance-mass**

Dual-life nouns and how to include their different senses into the lexicon are one of the major points for further work on lexical countability

## Conclusion & Further Research

Results of the annotation and the class distribution show a deep analysis of the countability features of a large set of nouns

→ data could be also used as a gold standard for a supervised classification on the type level

**For further research: focus on an analysis on the token level**

Another interesting intermediate result: annotators show an unexpected degree of disagreement in case of **nominalizations** (Inter-annotator agreement with Krippendorfs  $\alpha$ )

	Test I.2	Test II.2	Test III.1	Test III.2
All data	0.731	0.695	0.773	0.761
Nominalizations	0.618	0.630	0.703	0.669