Supplementary Materials for:
Say *Anything*: Automatic Semantic Infelicity Detection in L2 English Indefinite Pronouns

Ella Rabinovich[1]   **Julia Watson**[1]   **Barend Beekhuizen**[2]   **Suzanne Stevenson**[1]

[1]Dept. of Computer Science, University of Toronto, Canada
[2]Dept. of Language Studies, University of Toronto, Canada
{ella,jwatson,suzanne}@cs.toronto.edu
barendbeekhuizen@utoronto.ca

A Supplementary Materials

A.1 Indefinite pronouns usage classes – distance matrix

	SP	NS	QU	CD	IN	CP	DN	FC
SP	0.0	0.18	0.25	0.25	0.6	0.9	0.82	0.95
NS	0.18	0.0	0.07	0.07	0.4	0.7	0.72	0.85
QU	0.25	0.07	0.0	0.0	0.2	0.43	0.55	0.68
CD	0.25	0.07	0.0	0.0	0.22	0.38	0.6	0.57
IN	0.6	0.4	0.2	0.22	0.0	0.2	0.28	0.53
CP	0.9	0.7	0.43	0.38	0.2	0.0	0.5	0.22
DN	0.82	0.72	0.55	0.6	0.28	0.5	0.0	0.7
FC	0.95	0.85	0.68	0.57	0.53	0.22	0.7	0.0

A.2 Indefinite Pronoun Function Annotation Guidelines

Indefinite pronouns are pronouns such as *somewhere*, *anybody*, and *nothing*. For our research, we would like to know how these pronouns are used. Per case, you follow the decision steps below to arrive at the category for the pronoun. Every category is illustrated with several examples on the next page. We define the *minimal clause* as the clause that contains the indefinite pronoun and that does not contain clauses itself that contain the indefinite pronoun. For instance, in *I'd advise you not to do that*, the clause *not to do that* is the minimal clause, and the clause *I'd advise you __* is superordinate to the minimal clause.

In our analysis, classes INDIRECT NEGATION, FREE CHOICE, and SPECIFIC/NON-SPECIFIC were grouped together into a class referred to in the paper as **MIXED**.

1. [examples 1 and 2] Does the minimal clause contain an explicit negative element (*no, not, -nt, no-, never*)? If so, **DIRECT NEGATION [DN]**, otherwise:

2. [examples 3 and 4] Does any clause superordinate to the minimal clause contain a negative element, explicit or implicit, or does the minimal clause contain an implicit negative element (e.g., *barely, few, doubt, deny, without, lack*)? If so, **INDIRECT NEGATION [IN]**, otherwise:

3. [examples 5, 6] Is the minimal clause a question? If so, **QUESTION [QU]**, otherwise:

4. [examples 7, 8] Is the minimal clause the antecedent of a conditional? If so, **CONDITIONAL [CD]**, otherwise:

5. [examples 9 and 10] Does the indefinite pronoun refer to a comparison class in a comparison (e.g. with *than*)? If so, **COMPARISON [CP]**, otherwise:

6. [examples 11 and 12] Does the pronoun refer to a class of entities, any one of which can be chosen? As a soft test: does changing the sentence into a future with *will* or *must* make the sentence sound structurally awkward? If so, **FREE CHOICE [FC]**, otherwise:

7. It is **SPECIFIC/NON-SPECIFIC [SP/NS]** if

    (a) [examples 13 and 14] Does the speaker intend to refer to a particular entity (person, thing, place, time,)? A good way to test this is to see if the pronoun can be replaced by the phrase *a certain* {*person, thing*}, or:

    (b) [examples 15 and 16] Are the indefinite pronouns used in cases where the speaker conveys a non-realis situation (e.g., with modals, adverbs like *apparently, allegedly*)?

- In case of doubt, mark the category you think fits best and a question mark next to it. - If you cannot decide between COMPARISON/FREE CHOICE and SPECIFIC/NON-SPECIFIC, mark the example as **UNCLEAR FUNCTION [UF]**.

- If none of the categories seem to match (e.g., in idioms, such as *'something something'*, or *'or something'*), mark the category as **OTHER [OT]**.

EXAMPLES

DIRECT NEGATION

[1] DIRECT NEGATION   You do<u>n't</u> talk to anyone about this but me.

[2] DIRECT NEGATION   There's <u>no</u>thing here.

In these two cases, the element -n't *in the contracted form* don't *and the element* no- *in* nothing *are explicit negative elements.*

INDIRECT NEGATION

[3] INDIRECT NEGATION   [I have<u>n't</u> seen [you eat anything all day]].

[4] INDIRECT NEGATION   He's <u>barely</u> said anything.

In [3], the explicit negative element n't is in the clause superordinate to the minimal clause (the square brackets indicate clause boundaries). In [4], the element barely *functions as an implicit negative element (meaning 'almost not').*

QUESTION

[5] QUESTION   Have I got something on my face, soldier?

[6] QUESTION   Does anyone have a visual on spiderman?

In both these cases, the minimal clause is an interrogative clause.

CONDITIONAL

[7] CONDITIONAL   Let me know if you ever need anything.

[8] CONDITIONAL   If anyone could have figured a way out, it would have been you.

In both these cases, minimal clause is a conditional antecedent clause.

COMPARISON

[9] COMPARISON   I need to know your thoughts better than anyone.

[10] COMPARISON   He loved you more than anyone in this world.

In these cases, the comparison is marked by than.

FREE CHOICE

[11] FREE CHOICE   I think you can be anything you wanna be.

(cf. *I think you <u>will/must</u> be anything you wanna be.)

[12] FREE CHOICE   They fear any water their horses can't drink.

(cf. *They <u>will/must</u> fear any water their horses can't drink.)

In both these cases, adding strong modality with will/must *makes the sentence unacceptable. The speaker intends to refer to a free choice ('any member of ...') from among the classes of 'things you want to be' and 'water which their horses cant drink')*

SPECIFIC/NON-SPECIFIC

[13] SPECIFIC   There's something you need to know.

(cf. There's <u>a certain thing</u> you need to know.)

[14] SPECIFIC   Someone put us on this rock, there's gotta be a way off it.

(cf. <u>A certain person</u> puts us on this rock, there's gotta be a way off it.)

In these contexts, the speaker presents a situation in which the something/someone is an identifiable entity. The test of replacing the indefinite pronoun by a certain thing/person *works here.*

[15] NON-SPECIFIC   Surely, <u>someone</u> must've questioned his sanity in the end.

[16] NON-SPECIFIC   See if there is <u>something</u> that we can use as a bandage

In these two cases, the indefinite pronoun refers to an entity that cannot (yet) be identified; the test with a certain thing/person *does not work and the indefinite pronoun does not refer to a free choice from the entire class (compare the use of* must *in [ns1]).*

A.3 Classification of sentences with IP into usage classes

Algorithm 1: Classification of a sentence with IP into usage classes

---

Input: A sentence with indefinite pronoun
Output: Usage class ('DN', 'QU', 'CD', 'CP', 'MIXED') assigned to the sentence

NEGATION_MARKERS = ["n't", 'cannot', 'never', 'hardly', 'barely', 'nor', 'nobody', 'not', 'nothing', 'rarely', 'none', 'almost'] ;
CONDITION_MARKER = 'if' ;
COMPARISON_MARKER = 'than' ;
QUESTION_MARKER = '?' ;

tokenize input sentence;
extract minimal clause (min_clause) embedding indefinite pronoun;

```
/* ind() denotes sequential index in a sentence                    */
```
if *any of NEGATION_MARKERS in min_clause and ind(marker) <ind(IP)* **then**
  sentence_usage_class := 'DN';
  break;
end
if *CONDITION_MARKER in min_clause and ind(marker) <ind(IP)* **then**
  sentence_usage_class := 'CD';
  break;
end
if *sentence ends with QUESTION_MARKER min_clause is the last clause of the sentence* **then**
  sentence_usage_class := 'QU';
  break;
end
if *COMPARISON_MARKER in min_clause and ind(marker) <ind(IP)* **then**
  sentence_usage_class := 'CP';
  break;
end

```
/* none of the above, we label the sentence as 'MIXED'             */
```
sentence_usage_class := 'MIXED';

A.4 Preferred Indefinite Pronoun Annotation Guidelines

Note on attached sample: A sample of 500 of the annotated sentences for learners is attached to the supplementary materials. Note that the confidence scores for the annotations ("manual conf") are not the raw percentage of annotators that agree; they also take into account the "trustability" of the annotators based on their performance on gold standard test questions we designed. The "trustability"-adjusted confidence scores were computed by the annotation platform we used, FigureEight.

Guidelines:

In this task we explore spontaneous lexical choice between "something" and "anything" in sentences produced by English learners.

For each sentence, please fill in the place holder (_____) with "something", "anything", or "other" when neither of the first two seems appropriate. Note that some sentences can use both alternatives, depending on speaker's intention. Yet, try to make a choice between the two based on what would sound like a more natural alternative in most cases.

Please note, that even though some sentences are ungrammatical or have spelling errors, a speaker's intention is clear enough to make a choice. In (the rare) case a sentence is very unclear or poorly constructed, thus preventing from making a choice, specify "other".

A.5 Idiomatic expressions excluded from analysis

1. than anything else

2. more than anything

3. or something like that

4. or something (followed by punctuation)

5. something something

6. if anything