Annotation Manual

Created by jdegen on 2/15/18

# Features to annotate

For each of the following features of the format

**Name:** *value1*, *value2*, …

DESCRIPTION

create a column in the Excel sheet with header **Name** and fill in column with *value*s. Throughout the annotation process, take into account the context that the utterance occurred in if you have any uncertainty whatsoever about whether the context might influence your judgment.

**Sentence\_type:** *declarative, interrogative, imperative*

The type of sentence that “or” occurs in.

**DE\_operator:** *negation, antecedent, relclause, negpredicate*

Examples of each downward-entailing operator:

* Negation (*negation*): “She did **not** trust Bill or John.”
* Antecedent of conditional (*antecedent*): “**If** Fred has failed the practical part or he has failed the theoretical part, then he has failed the entire test.”
* Relative clause (*relclause*): “At present, we will invite all passengers **who** need some extra help or are travelling with small children to board the aircraft.”
* Negative predicate (*negpredicate*):“I **doubt** she’ll take the train or the bus.”

**Numeral:** *num, nonum*

Are the disjuncts numerals or do they prominently contain numerals?

Example of *num*: “I’ve had it for **six or seven weeks**.”

**And\_test:** *change, nochange*

Does replacing “or” with “and” result in a change in (speaker intended) meaning?

Examples:

* *change*: “Well, we just started two **or** three weeks ago.” 🡪 “Well, we just started two **and** three weeks ago.”
* *nochange*: “They always like to be able to attract the, uh, Einsteins **or** the Professor Chou.” 🡪 “They always like to be able to attract the, uh, Einsteins **and** the Professor Chou.”

**Or\_both\_test:** *change, nochange, NA*

Does adding “or both” after the (relevant part of the) sentence (after the disjunction) result in a change in (speaker intended) meaning? Use *NA* if adding “or both” makes the sentence sound incoherent.

Examples:

* *change*: “I, I mean I could have had that or a down payment on a new car.” 🡪 “I, I mean I could have had that or a down payment on a new car, **or both**.”
* *nochange*: “And then the states that either don't adopt on a state wide basis or don't have as many kids, just have to follow along and take what Texas gets.” 🡪 “And then the states that either don't adopt on a state wide basis or don't have as many kids, **or both,** just have to follow along and take what Texas gets.”
* *NA*: “Well, I guess I have used it once or twice.” 🡪 “Well, I guess I have used it once or twice, **or both**.”

**Actually\_test**: *change, nochange*

Does inserting “actually” after “or” result in a change in (speaker intended) meaning?

Examples:

* *change*: “I, I mean I could have had that or a down payment on a new car.” 🡪 “I, I mean I could have had that or **actually** a down payment on a new car.”
* *nochange*: “And we intend to replace our other one with a much newer car in the next couple of years, or next year probably.” 🡪 “And we intend to replace our other one with a much newer car in the next couple of years, or **actually** next year probably.”

**Disjunct\_relation**: *A>B, A<B, A~B*

What is the relation between the two disjuncts A (first) and B (second)? If the first is stronger than the second according to some ordering relation (e.g., it entails the second or is “narrower” in meaning than the second, use *A>B*. If the reverse is true, use *A<B*. If neither is stronger than the other, use *A~B*.

Examples:

* *A>B*:
  + “And there was no such thing as **a job** or **any hope of a job**, unti-, well, unless,”
  + “Now, you know, I mean somebody, a s-, s-, **serial killer** or **somebody** goes in and machine guns, you know, fifty, fifty children in a school yard,”
* *A<B*:
  + “And we intend to replace our other one with a much newer car in the **next couple of years**, or **next year** probably.”
  + “You should, uh, **pursue that** I think, or **patent it**, you know.”
* *A~B*:
  + “We find it **hard to believe sometimes**, or **hard to understand** when, uh, we work for a computer company, and, uh,”
  + “You see all these people, yeah, these stately people, and well dressed people out walking their dogs with **their plastic bag** or **their piece of paper** for them,”

**Tense**: *past, future, present*

Is the sentence about something that happened in the *past*, for which there is already a fact of the matter? Is it instead about something that will or should occur in the *future*, for which there is not yet a fact of the matter? Is it instead about the *present* (including habitual occurrences)?

Examples:

* *past*: “Do you work for a big or a little place.”
* *future*: “You should, uh, pursue that i think, or patent it, you know.”
* *present*: “I think in some cases it doesn't matter if it's a deterrent or not.”

**Disjunct\_individuation**: *same, different*

Do the two disjuncts refer to the *same* object, action, event, fact, times, etc, or to two *different* ones? Assign *same* if you can paraphrase the sentence by replacing the disjunctive phrase with “a certain X, which can be characterized as A or B”.

Examples:

* *same*:
  + “And there was no such thing as **a job** or **any hope of a job**, unti-, well, unless,” 🡪 Paraphrase : There was no such thing as **a certain thing**, which can be characterized as « a job » or as « any hope of a job ».
  + “Well, I guess I have used it **once** or **twice**.” 🡪 Paraphrase : Well, I guess I have used it **a certain number of times**, which can be characterized as « once » or « twice »
* *different*:
  + “I, I mean I could have had **that** or **a down payment on a new car**.” 🡪 Paraphrase ***fail***: I mean I could have had **a certain thing**, which can be characterized as « that »  or as « a down payment »

**Notes**: anything you notice about this particular example that doesn’t fit

**Event Subtype Coding**

Apply the event subtype annotation scheme developed by Annemarie Silke Friedrich, States, events, and generics: computational modeling of situation entity types (Chapter 5).

(<https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/23722/1/scidok_final.pdf>)

Descriptions and examples below are taken from this paper.

For each sentence, look at the smallest possible clause that contains “or” (embedded or relative clause) and annotate genericity of main referent and habituality of clause. For questions, write down two possible answers and annotate the features according to these answers. For imperatives, make a declarative.

**Genericity of main referent:** *expletive, generic, non generic*

* + If the main referent refers to a kind or to arbitrary members of the kind, its generic, if it refers to some particular entity, its non-generic.
  + Non-generic main referents are particular entities (1), particular groups of entities (2), organizations (3), particular situations (4) or particular instantiations of a concept (5)
  + Cases such as (6), which make a statement about a kind or characterize arbitrary members of a kind, are labeled as generic.
  + Gerunds may occur as the subject in English sentences. They usually describe some process or kind of process, and thus also do not have a clear physical referent. When they describe a particular process such as in (7), we mark them as non-generic individuals. When they describe a kind of process as in (8), however, we mark them as concepts.
  + The label cannot decide may be used for cases where the annotator does not feel comfortable choosing one of the other labels.
  + **TEST:**  replace the main referent with an arbitrary member of a set

Examples:

1. Mary likes popcorn. (non-generic)
2. The students met at the cafeteria. (non-generic)
3. IBM was a very popular company in the 80s. (non-generic)
4. That she didn’t answer her phone really upset me. (non-generic)
5. Today’s weather was really nice. (non-generic)
6. Lions in captivity have trouble producing offspring. (generic)
7. Knitting this scarf took me 3 days. (non-generic)
8. Knitting a scarf is generally fun. (generic)
9. The students of Saarland University don’t mind eating at their mensa. (cannot decide
10. A lion has a bushy tail. (generic)
11. A lion escaped from the zoo. (non-generic)
12. Dinosaurs are extinct. (generic)
13. Dogs are barking in the garden. (non-generic)

**Habituality of clause:** *habitual, episodic, static*

* Habituality is a situation-related feature at the clause-level. Episodic clauses report particular events (31), and habitual clauses constitute generalizations over events and activities (32), or even states (33).
* In the case of clauses whose subject refers to kinds, habitual clauses include cases where a situation occurs regularly for (possibly different) members of the class (4).
* English clauses in past or present perfect such as (5) are static, as they focus on the post-state of an event rather than the event itself.
* Modalized (6) and negated sentences (7) tend to be static: they do not express information about a particular event, but refer to actual or possible states of the world.
* Negated sentences whose positive version is habitual (8) are considered habitual when they characterize behavior.
* Sentences containing modal verbs are marked as habitual if they have a strong implicature that an event has actually happened regularly (9), otherwise, they are marked as static (6)
* **Test:** reformulate the sentence by changing the positive form to a repeated negation.

Examples:

1. Mary cycled to work today. (episodic)
2. Mary usually cycles to work. (habitual)
3. Sloths sometimes rest on top of branches. (habitual)
4. Spiders die in autumn after producing an egg sac. (habitual)
5. Mary has made a cake. (static)
6. Mary can swim. (static)
7. Mary didn’t go swimming yesterday. (static)
8. John does not smoke. (habitual)
9. I had to eat an apple every day. (habitual)

**Universal quantifier**: *yes, no*

All, necessary, always, every time/moment/place, everybody, each

Examples:

* *yes*:
  + “I always sing or talk on the phone when I’m alone in the house”
* *no:*
  + “I don’t approve or support everything they say ”

**Negative quantifier:** *yes, no*

No, none, few, fewer, barely/hardly any, impossible, unlikely, improbably, never, rarely, infrequently, nobody, hardly/barely, anyone

Examples:

* *yes*:
  + “No chocolate or candy tastes as good as fresh vegetables.”
* *no*:
  + “I don’t eat chocolate or candy, nothing tastes as good as fresh vegetables to me. ”

**Existential quantifier:** *yes, no*

You can distinguish an existential quantifier from quantifiers in 4. because they have a literal meaning that makes reference to a 0 point. Eg, "some" means "more than 0"; "possibly" means "greater than 0 percent chance", etc.

somebody, someone, something, sometimes, occasionally, on occasion, several, a few, possible that, possibly, conceivable that

Examples:

* *yes*:
  + “I sometimes eat or drink while I talk on the phone.”

**Other quantifier:** *yes, no*

Frequently, often, usually, probably, likely, most, many, tons of, a bunch, a lot, exactly one third, about half, ca. 90%

Examples:

* *yes*:
  + “I will get many dogs and cats when I have my own place.”

**Embedding Predicate:** *the predicate*

say, mention, shout, declare, announce, remark, hear, doubt, deny

Examples:

* *believe*:
  + “I believe she skis or snowboards” 🡪 the smallest full sentence containing “or” is “she skis or snowboards”.
* *no embedding predicate*:
  + “I believe the doctor or the teacher” 🡪 “the doctor or the teacher ” is the direct object of “believe”
* *know:*
  + “He didn’t know whether to laugh or cry” 🡪 the smallest full sentence containing “or” is a question “weather to laugh or cry”.

**Modals:** *yes, no*

can, could, may, might, must, shall, should, will, would, have to

Examples:

* *yes*:
  + “I can try or give up”

# Previously on Corpus Or…

Everything below is obsolete and included only for the purpose of historical record-keeping. In particular, it contains the uses of “or” we were hoping to be able to annotate before giving up and switching to annotating features. We may come back to this once we’ve completed the feature annotation and performed some initial clustering of uses.

1. Boolean ‘or’ / Downward Entailing

Diagnostic: 1. Under negation: “Not A or B” → Not A and not B

2. In antecedents of conditionals

3. In relative clauses

i. “If Fred has failed the practical part or he has failed the theoretical part, then he has failed the entire test.”

ii. “She did not trust Bill or John.”

iii. “At present, we will invite all passengers who need some extra help or are travelling with small children to board the aircraft.”

2. Conditional:

- imperative in first clause

i. “Stand up, or I’m going to break your arm.”

ii. “Be quiet, or I’ll tell your parents.”

iii. “Kiss me, or go to hell.”

3. Conjunction:

Diagnostic: Replace “or” with “and” without (significant?) change in meaning

a. i. “Language is governed or, in other words, it follows regular patterns.”

ii. “A year has 12 months or 365 days.”

iii. “The gap between the door and the floor should be at least 1 inch, or 2.5 cm.”

b. With “free choice” modals

i. “You may pick from a dish from column A or from column B.”

ii. “You can have this candy or that one.”

4. Exclusive:

Diagnostic test to distinguish from 5.: can you felicitously respond with “Actually, both.” If yes, “exclusive”. If no, “exclusive world knowledge”.

i. “She went to the movies or had a drink with her friends.”

ii. “Wood is going to be sunbathing in the Bahamas soon or lining his pockets in Vegas.

5. Exclusive world knowledge:

i. “John is in Tokyo or in New York.”

6. Repair

Diagnostic: Second disjunct contradicts first disjunct?

Apply “or actually” – if legible, repair.

ii. “He and Kozinski talk a lot, … or whisper.”

7. Alternative offer

Diagnostic: Cannot replace “or” with “and” without change in meaning, but second disjunct provides more detail or a similar alternative to the first – but you can’t follow up with “or both”

i. “Let’s go for a drink, or, … let’s take a nap.”

8. True Alternative

Diagnostic: Cannot replace “or” with “and” without change in meaning, but can follow with “or both”

9. Exclusive analytic

-“or not”

10. Lexical aspectual class of main verb: *stative, dynamic*

* Treat as a property of verbs in context, as many verbs may be used both in dynamic and in stative senses. (1) and (2)
* Stative verbs disallow progressive constructions (3) while dynamic verbs allow only limited use of the simple present tense (4)
* English sentences with perfect aspect are usually considered to introduce states to the discourse but we are interested in the aspectual class before this transformation takes place. At the clause-level, (5) introduces a State, but the fundamental aspectual class of the verb “kiss” in this context is dynamic.

Examples:

1. She filled the glass with juice. (dynamic)
2. Water fills the pool. (stative)
3. \*I am owning a house.
4. \*I run to class now.
5. John has kissed Mary. (dynamic)
6. My name is John. (stative)
7. Your soul was made to be filled with God Himself. (stative/dynamic)

Proposed changes:

* Call category 1 “downward-entailing” instead of “Boolean” – diagnostics / markers: “or” under negation, antecedent of conditional, relative clause
* Add category “hedge” – ways of distinguishing it from “repair” and “alternative”?
* Add sentence type annotation (interrogative, imperative)
* Add explicit numeral annotation
* Barely any conditionals or free choice cases even though these are so widely discussed (make note in paper)
* How to distinguish “conjunction” from “alternative”? 🡪 Does second disjunct offer more detail, or say the similar thing (as opposed to contradicting or seeming incompatible with the first disjunct)
* **Special category for “or not”? 🡪 “Exclusive analytic“**
* Diagnostic for “Boolean”:
  + Under negation: “not A or B” 🡪 “not A and not B”
  + In antecedents of conditionals:
  + In relative clauses:
* In interrogatives, check for speaker intent and mark that (exclusive etc)
* **Conjunction: if you can replace “or” with “and” without (significant?) change in meaning**
* **Alternative: if you can’t replace “or” with “and” without change in meaning, but second disjunct provides more detail or a similar alternative to the first – but you can’t follow up with “or both”**
* **True alternative: if you can’t replace “or” with “and” without change in meaning, but you can follow up with “or both”**
* **Alternative vs. repair → Ask if speaker/context applies correction (subjective ):**
* **Boolean vs. alternative → If wavering, default to boolean as downward entailing**