CuRIAM Annotation Guidelines
v1.0.1

Michael Kranzlein, Nathan Schneider, Kevin Tobia

June 2023

Contents

Introduction

These guidelines establish and detail categories of metalanguage commonly found in legal documents. In this context, metalanguage describes language used to describe or reason about other language. A justice citing a dictionary definition in an opinion to support an argument about the meaning of a word is a common example of metalanguage. It's helpful to think of each of the categories in this schema as belonging to one of 3 groups according to what the category describes:

- General Metalanguage

- Quotes and Sources

- Interpretive Rhetoric

Be aware that the examples in this document do not include comprehensive annotation. They may only include an annotation or two to demonstrate a specific point. If a word is not annotated, don't assume it shouldn't be annotated (it may just not be relevant to what the example is trying to achieve). If an example is meant to show that something should not be annotated, the relevant span will be bracketed and colored like a normal annotation, but the category will be struckthrough.

Normal annotation: [FT this is an annotation.]
Do not annotate: [FT this is something that shouldn't be annotated.]

General Metalanguage

Focal term

Focal Term
Word or phrase used metalinguistically and/or whose meaning is under discussion.

It's helpful to think of two subtypes of focal terms. The first is a higher-level group of terms that are central to the discussion of the case. These words will often be repeated many times in a case, but not every instance will be a focal term. The second subtype includes focal terms that exist in a more local context. These are words or phrases that may only appear a few times, or even only once in an opinion, but they are used metalinguistically or in the context of a discussion of meaning. These are often easier to identify.

Mentions

For making the distinction between when to annotate and when not to annotate words or phrases that might be focal terms, first **consider whether the span is a use or a mention**. If the span is preceded by something like "the word" or "the term," this is one of the most direct clues that it is a mention. Mentions should be annotated as focal terms.

(1)   The word "[FT from]" is broad in scope, but context often imposes limitations.

Other times, a span is a mention, even without those preceding terms:

(2)   The use of "[FT arising under]" in §113(b) echoes Congress's more familiar use of that phrase...

Non-mentions

When a span is not a mention, consider whether the span is the focus of a discussion about meaning:

(3)   Is pollution that reaches navigable waters only through groundwater pollution that is "[FT from]" a point source, as the statute uses the word?

Here are two sentences that both include "facility." However, only the second sentence gets a focal term annotation. The first sentence includes quotation marks around "facility" but does not discuss what the word means. While the second sentence doesn't set out to define the word "facility," it is metalinguistic in that it involves discussion of whether something qualifies or counts as a "facility."

(4)   One of these classes encompasses any person who owns a "[FT facility]" where hazardous waste has "come to be located."

(5)   Because the landowners' properties qualify as "[FT facilit[ies]]" where...

Quotation marks are indicators sometimes

Focal terms often appear in quotation marks, and quotation marks can be helpful for finding focal terms. But the presence of quotation marks around a single token does not make that token a focal term. Instead, "metalinguisticness" and the presence or absence of a discussion of meaning are factors to consider. In the next example, "from" is a focal term, but "conveyance" is not, despite both tokens being inside quotations marks.

(6)   A pollutant is "[FT from]" a point source only if a point source is the last "[FT conveyance]" that conducted the pollutant to navigable waters.

Other considerations

This pattern—where a term is given with its dictionary definition—is one way in which lower-level or more local focal terms appear:

(7)   See generally 2 Van Nostrand's Scientific Encyclopedia 2600 (10th ed. 2008) (defining "[FT Hydrology]").

When a term is used conceptually (but without discussion of meaning), do not annotate it as a focal term:

(8)   the Ninth Circuit's "[FT fairly traceable]" test...

Within example uses, annotate the exemplified word or phrase as a focal term. In the context of this example, the hypothetical is given to demonstrate the behavior of the word "from":

(9)   A recipe might instruct to "add the drippings [FT from] the meat to the gravy"

For focal terms, if you aren't sure, err on the side of not annotating.

Do not include quotation marks in focal term annotations unless necessary, e.g.:

(10)   The meaning of [FT some fancy "term]"...

In other words, imagine the quotation marks weren't there and focus on just the words.

Definition

> **Definition**
> Succinct, reasonably self-contained description of what a word or phrase means. Need not be exhaustive. May also be negative—defining a word by what it's not.

(11)   ..the term "violation" referred to [D the "[a]ct or instance of violating, or state of being violated]." Webster's New International Dictionary 2846 (2d ed. 1949) (Webster's Second).

"i.e." can be an indicator of a definition:

(12)   When a liquid flows over the surface of land to navigable waters, the surface is a conveyance, i.e., [D a "means of carrying or transporting something" from one place to another].

Definitions can be short and sometimes less obvious:

(13)   The alternative way in which the statutory language could be interpreted—reading "from" to mean "[D originally from]" would lead to extreme results, as the court recognizes.

(14)   ...this Court interpreted the statutory term "final order of deportation" in the... to mean "[D all determinations made during and incident to the administrative proceeding" on removability].

(15)   ...an application is commonly regarded as having been "'filed'" if "[D it is delivered to, and accepted by, the appropriate court officer for placement into the official record]."

(16)   We understand a golden cup to be [D a cup made of or resembling gold].

Here, there are two individual definitions and the whole sentence will receive an example use annotation.

(17)   An "American flag" could literally encompass [D a flag made in america], but in common parlance it denotes [D the Stars and Stripes].

In this example, a translation of a Latin phrase functions as a definition:

(18)   ...see id., at 21 (invoking the "interpretive canon noscitur a sociis, [D a word is known by the company it keeps]" (internal quotation marks omitted)).

Do not include quotation marks in definition annotations unless necessary.

Do not include sentence-final punctuation in definition annotations.

Metalinguistic Cue

Metalinguistic Cue
Word or short phrase cueing nearby metalanguage.

Metalinguistic cue is a category that mostly comes down to annotating words from a set list. However, there are exceptions, where sometimes these words won't be metalinguistic cues. The lists below should also not be considered comprehensive. While they can help find many metalinguistic cues, keep your eyes peeled for others that aren't on these lists.

Almost always metalinguistic cues

- interpret (interprets, interpreted, interpretation, interpretations)

- mean (means, meant, meaning, meanings)

- language (linguistic)

- phrase (phrases, phrasing, phrased)

- read (reads, reading)

- term (terms, termed, terminology)

- text (texts, textual, textualist, atextual)

- unambiguous (unambiguously)

- word (words, wording, worded)

- understand (understanding, understood)

Exceptions to the list above will be uses of these words that aren't related to meaning. For example, "[MC reading] a book" vs. "we [MC read] this statute to mean"; "Under the [MC terms] of the contract" vs. "This [MC term] is clearly defined"; "The French [MC language]" vs. "The [MC language] of the statue."

"Understand" usually occurs in a context where it could be substituted with "interpret", or where "understanding" could be substituted with "interpretation." In these cases, annotate as a metalinguistic cue.

(19)   We [MC understand] this term to mean...
(20)   We [MC interpret] this term to mean...

Uses of "understand" that are **not related to language** are not metalinguistic cues, like in:

(21)   "Assuming we [MC understand] the situation correctly..."

Sometimes metalinguistic cues

- clear (clearly)

- context (contexts, contextual)

- explicit (explicitly)

- expressly

- refer (refers, referred, reference)

Words that describe the degree of clarity (e.g. clear, explicit, expressly, unambiguously) are metalinguistic cues when they refer specifically to the meaning of language.

"Context" is tricky. The key is to determine whether the "context" is a linguistic sort of context (e.g. the context is the surrounding sentences or the definition of a word) or some other kind of context, like a cultural context. If it's a linguistic context, annotate it as a metalinguistic cue. Otherwise, don't.

A reliable test for "refer" is whether "refers to" can be replaced with "means." If so, annotate "refer" as a metalinguistic cue. If not, consider whether what's being referred to is about language and meaning or something else, like a legal concept. In the former case, annotate "refer" as a metalinguistic cue. In the latter case, don't.

Not metalinguistic cues

For standardization, **do not** annotate these terms as metalinguistic cues:

- statute (statues, statutory)

- provision

- clause

- intent

- history

"Intent" and "history" are important terms for recognizing potential appeals to meaning or named interpretive rules, and these terms may very well be part of those annotations. There's an argument to be made that these words are metalinguistic cues as well, but this starts to overly broaden the category beyond its focus on meaning.

The exact boundaries of metalinguistic cues are not critical, but err on the short side if hesitant. For example, "[MC statutory text]" and "statutory [MC text]" are both fine. Note that "statutory" (on the not metalinguistic cue list) is only acceptable as part of the annotation because it is modifying "text."

If there are multiple metalinguistic tokens adjacent to each other, adjectival or adverbial modifiers are fine to include as part of a metalinguistic NP or VP, but break other constituents up. Use the following as an example:

(22)   This [MC unambiguous language] [MC clearly defines] the [MC term].

Quotes and Sources

Direct Quote

Direct Quote
Span of text inside quotation marks.

Direct quotes are a straightforward category. Annotate the quotation marks and the content inside.

(23)   The Clean Water Act requires a permit for [DQ "any addition of any pollutant to navigable waters from any point source."]

If quotes are nested, create a separate annotation for each layer of quotation. If a direct quote spans multiple sentences, create one annotation from the beginning of the direct quote in the first sentence to the end of the first sentence. Then create an additional annotation from the beginning of the second sentence to the end of the direct quote.

Legal Source

<div style="border: 1px solid black; background-color: #f8cccc; text-align: center;">

Legal Source
Citation or mention appealing to a legal document or authority.

</div>

Legal sources include citations to common law, Supreme Court precedent, legislative evidence (committee reports, legislative history), opinions from lower courts, other opinions from the current case, case materials like briefs and appendices, the Constitution, Federal Rules of Civil Procedure, legal textbooks, law review articles, and more.

Legal sources are most identifiable as a standalone sentence following another sentence.

(24)   Under its terms, once a plaintiff establishes a prima facie case of race discrimination through indirect proof, the defendant bears the burden of producing a race-neutral explanation for its action, after which the plaintiff may challenge that explanation as pretextual. [LeS Texas Dept. of Community Affairs v. Burdine, 450 U.S. 248, 257-258, 101 S.Ct. 1089, 67 L.Ed.2d 207 (1981)].

Note, in the actual opinion, the text is "Texas Dept. of Community Affairs v. Burdine, 450 U.S. 248, 257-258 (1981)". Harvard Caselaw Access Project data adds (and hyperlinks) citations for other reporters of Supreme Court data (in this case "101 S.Ct 1089" and "67 L.Ed.2d 207"). This can make it difficult to separate sources, but be aware of this pattern and look for clues like semicolons, a citation of a different volume of the U.S. Reports (the "450 U.S. 248" above), or a separate case name. Mention of multiple reporters for a citation for one case counts as one legal source, but citations to separate cases should involve separate annotations.
a specific part "of" a specific larger text would be one source right? "Section 145 of the Patent Act"

Coreference and legal sources

Think of full citations as the prototypical legal source (like "Texas Dept. of Community Affairs v. Burdine" above). However, there are other coreferent forms that should also be annotated. For example, annotate "Act" as a legal source when "Act" refers to a previous citation:

(25)   The [LeS Act] authorizes noncitizens to obtain direct "review of a final order of removal" in a court of appeals.

You should also annotate initialisms or other ways of referring to pieces of legislation, treaties, etc., like here with "Convention" in reference to the Convention Against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment:

(26)   The [LeS Convention] was sent to the Senate for its advice and consent in 1990.

Or here, with Title XII:

(27)   [LeS Title VII] was enacted in 1964...

This also extends to specific sections of legislation, as in:

(28)   Meanwhile, [LeS §1981] dates back to 1866...

Group multiple sections together as one legal source and do not annotate the term "section" as a legal source without a detailed specifier:

(29)   ...the Court's "precedents have ... construed [LeS §§1981 and 1982] similarly."

(30)   "Except as provided in [LeS subsections (a) and (h)] of this [LeS section]..."

Note how "section" in the prior example is not annotated as a legal source.

Annotate citations as legal sources if they use any of these types of terms: *supra*, *ibid.*, *id.*, *post*, *ante*. Do not include *cf.*, *see*, or similar in legal source annotations.

(31)   See [LeS ante, at 1171].

(32)   [Les Id., at 346-347, 133 S.Ct. 2517] (internal quotations marks omitted).

(33)   In Click-to-Call's view, which the dissent embraces, [Les post, at 1380-1387], the bar on judicial review applies only...

Do include "In re" as part of a legal source when present:

(34)   [Les In re Jurado-Delgado, 24 I. & N. Dec. 29, 31 (BIA 2006)].

Parentheticals

When a parenthetical describes that one source is quoting another, each source should be separately annotated:

(35)   [Les Ante, at 1349-1350] (quoting [Les American Well Works Co. v. Lane & Bowler Co., 241 U.S. 257, 260, 36 S.Ct. 585, 60 L.Ed. 987 (1916)])

When a parenthetical describes the purpose of the citation or talks about omitted quotation marks, emphasis, etc., do not include that as part of a legal source annotation unless it is necessary for finding the specific document (e.g. which opinion in a cited Supreme Court case). See the following examples:

(36)   [Les INS v. St. Cyr, 533 U.S. 289, 328-330, 121 S.Ct. 2271, 150 L.Ed.2d 347 (2001) (Scalia, J., dissenting)]

(37)   [Les Id. at 531, 125 S.Ct. 2641] (internal quotation marks omitted)

(38)   see [Les FRAP 4(a)(4)(A)(iv)] (A party's "time to file an appeal runs" from "the entry of the order disposing of the [Rule 59(e)] motion").

For the citation pattern involving the phrase "codified at," try to create one annotation:

(39)   Second, the so-called gatekeeping provision of the [Les Antiterrorism and Effective Death Penalty Act of 1996 (AEDPA), codified at 28 U.S.C §2244(b)], governs federal habeas corpus proceedings.

Legal sources in broader NPs

For standardization, if a legal source is part of a larger NP, do not annotate it if it is an adjectival modifier. Do annotate it if it is possessive. *'s* will be tokenized separately, and this token does not need to be included in the annotation.

(40)   This Court first adopted [Les Title VII] 's motivating factor test...

(41)   There, a plurality and two Justices concurring in the judgment held that a [Les Title VII] plaintiff...

Language Source

Language Source
Citation or mention appealing to an authority on language.

Language sources are almost always dictionaries:

(42)   And the meaning of "individual" was as uncontroversial in 1964 as it is today "A particular being as distinguished from a class, species, or collection." [LaS Webster's New International Dictionary, at 1267].

But there are some exceptions like academic textbooks or other reference materials, like the Diagnostic and Statistical Manual of Mental Disorders (DSM):

(43)   Nothing resembling what is now called gender dysphoria appeared in either [LaS DSM-I] (1952) or [LaS DSM-II] (1968).

If something would normally be annotated as a legal source, but it is functioning as a language source by defining a term, it should still just be annotated as a legal source.

Black's Law Dictionary is a language source, not a legal source.

Interpretive Rhetoric

Appeal to Meaning

Appeal to Meaning
An explicit argument, implicit value judgment, or other statement indicating how one should go about interpreting meaning (e.g., by appealing to common sense, ordinary meaning, or the language of another statute).

This is a rarer category, and exact boundaries are not critical, so err on the side of large annotation spans for this category. It is okay to include sentence-final punctuation in appeal to meaning annotations. An ATM passage should explicitly comment/reflect on the mode of inquiry or philosophy underlying the interpretation: e.g., textualist, or plain meaning, etc. Merely articulating a set of logical steps involving (e.g.) textual evidence or hypotheticals is not ATM. ATM is not limited to passages articulating the author's own viewpoint; it could be applied to a passage discussing another interpretative process (e.g., from another opinion that the author disagrees with). TODO: in a future version consider broadening ATM to Reflection on Interpretation (whether it is on the procedure for interpretation, or that an interpretation is correct/flawed/difficult/etc.) Maybe stipulate that it applies to entire sentences.

(44)   [ATM And where, as here, Congress has simultaneously chosen to amend one statute in one way and a second statute in another way, we normally assume the differences in language imply differences in meaning].

(45)   [ATM Consequently, "[s]licing a statue into phrases while ignoring... the setting of the enactment... is a formula for disaster."]

Multi-sentence annotations are disallowed, so instead, just create two annotations:

(46)   [ATM Many Justices of this Court, both past and present, have not espoused or practiced a method of statutory interpretation that is limited to the analysis of the statutory text.] [ATM Instead when there is ambiguity in the terms of a statute, they have found it appropriate to look to other evidence of "congressional intent," including legislative history.]

Appeals to meaning often invoke "plain meaning," "ordinary meaning," etc. This demonstrates a pattern of overlapping annotations since "plain meaning" and "ordinary meaning" are also part of the semantic canons of construction and therefore named interpretive rules. These terms can also be considered metalinguistic cues. Therefore, annotate "plain meaning" as both a metalinguistic cue and a named interpretive rule, and the broader context as an appeal to meaning.

(47)   [ATM Without strong evidence to the contrary (and there is none here), our job is to ascertain and apply the "ordinary meaning" of the statue.]

An appeal to meaning does not need to stand on its own. In this example, you might not be inclined to annotate this as an appeal to meaning, but the hypothetical a few sentences prior in this opinion (which is annotated as an example use of the phrase "discrimination because of sex") makes it clear that the following is an argument about the impact of that specific example on the Court's interpretation:

(48)   [ATM The attorney's concession was necessary, but it is fatal to the Court's interpretation, for if an employer discriminates against individual applicants or employees without even knowing whether they are male or female, it is impossible to argue that the employer intentionally discriminated because of sex.]

Named Interpretive Rule

Named Interpretive Rule
Mention of a well-established interpretive rule or test used to support an argument about the meaning of a word or phrase.

Before annotating, it's a good idea to skim the 33 semantic canons in the appendix of *Statutory Interpretation: Theories, Tools, and Trends*: https://crsreports.congress.gov/product/pdf/R/R45153

Named interpretive rules can be hard to spot, and some documents may have none. Look for latin phrases like *noscitur a sociis* and *ejusdem generis*. Focus primarily on instances where a canon is mentioned by name. However, if it's very clear that an argument is being made on the basis of an interpretive canon but the canon is not explicitly mentioned by name, consider annotating that argument as a named interpretive rule.

(49)   Without strong evidence to the contrary (and there is none here), our job is to ascertain and apply the "[NIR ordinary meaning]" of the statue.

(50)   ...a bit of [NIR surplusage] makes no difference in any event.

While the main focus of this category is the semantic canons, some of the substantive canons are related to meaning and ambiguity like the rule of lenity. When that's the case, annotate as named interpretive rules.

(51)   I write separately to elaborate on why the [NIR rule of lenity] does not apply here.

Example Use

> **Example Use**
> Intuitive, quoted, or hypothetical examples that demonstrate a word/term can or cannot be used in a certain way.

Example uses are another rare category. Favor long over short annotations for this category. It is acceptable to include sentence-final punctuation and to annotate entire sentences. Remember to annotate focal terms within example uses.

Two example uses for the word "from":

(52)   [EU When a point source releases pollutants to groundwater, one would naturally say that the groundwater has been augmented with pollutants from the point source.]

(53)   [EU A sign that asks all persons who arrive from Baltimore to speak to the desk clerk includes those who took a taxi from the train station.]