What Quotation Marks, String Literals and DNA Have in Common

First-Order Expressions, Second-Order Expressions, and the Delimiters Between Them

Guillaume Lethuillier
glthr.com

July 6, 2024 · updated November 15, 2026

Series: Delimiters in Languages — Part 1 of 5
Next: Special Tokens: The Quotation Marks of LLMs (glthr.com/special-tokens-the-quotation-marks-of-llms)

Abstract

I propose a common way to read quotation marks, string literals and CRISPR repeats: they mark a change between expressions that a system evaluates and expressions that it represents. I call these first-order and second-order expressions. This conceptual analogy connects speech, code and genetic memory; the questions it raises about LLMs are developed in the later articles of this series.

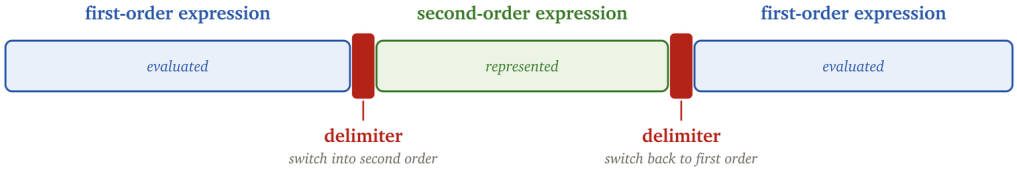
There seems to be a pattern in *languages* (in the broadest sense in the context of this article: a series of symbols that are evaluated or interpreted): **a duality between first-order expressions and second-order expressions that have different purposes but paradoxically use the same symbols.**

This observation is probably self-evident (it echoes the use–mention distinction in philosophy of language, discussed by Gottlob Frege and Willard Van Orman Quine, and Alfred Tarski's distinction between an object language and the metalanguage that talks about it), but I have not seen it explicitly formulated in these terms:

- A language primarily uses **first-order expressions**, which are intended to be evaluated. A human reader evaluates a human language to produce a representation of the world. An example of a first-order expression in English is: *It is raining outside.*
- A language also needs to represent information used in a capacity different from first-order expressions. This information is represented with **second-order expressions**. Example: *"And what is the use of a book," thought Alice, "without pictures or conversations?"* In this context, this expression has a different purpose: it represents what someone other than the narrator is thinking or saying.

Cite as: Guillaume Lethuillier, "What Quotation Marks, String Literals and DNA Have in Common", glthr.com, July 6, 2024. https://glthr.com/quotation-marks-string-literals-and-dna
Interactive figures, animations and videos of the web page are shown here as still images or addresses.

- First- and second-order expressions use **the same symbols** (for instance, the English alphabet).
- First- and second-order expressions are separated with **delimiters**. *Only delimiters identify whether an expression is of the first-order kind or of the second-order one.* In other words, they signal a transition between purposes.
- A language evaluator (*e.g.*, someone reading a sentence) needs the delimiters to **switch** between first-order and second-order purposes.



A delimiter switches into second order; the next one switches back.

Three Languages, One Mechanism

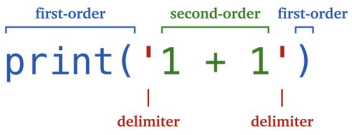
This observation can be applied to many forms of *languages*. I suggest three of them: natural language, artificial language, and genetic code.

- Since antiquity (for example, ḏd-mdw in Ancient Egypt), human language has used this mechanism to report what someone other than the narrator says. In English, the delimiters are the quotation marks.



In reported speech, the quotation marks are the delimiters.

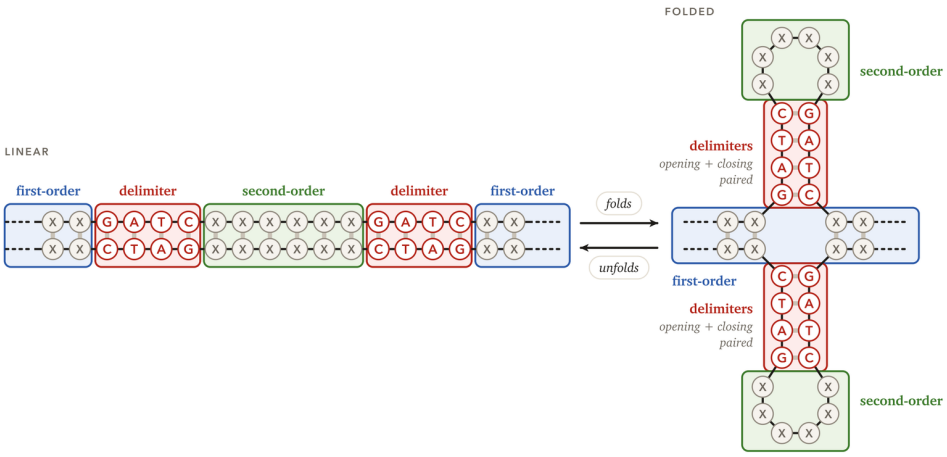
- Many programming languages use this approach to distinguish between *evaluated* and *represented* statements. Examples in Python: `print(1 + 1)` → `2` vs. `print('1 + 1')` → `1 + 1`.



In code, the quotes of a string literal do the same.

LISP uses an apostrophe (or `QUOTE`). Quine programs are characteristic of this approach: they usually duplicate and combine first-order symbols (the code evaluated by the computer) and second-order symbols (the code represented by the computer) to blur the line between them as they output the exact representation of the evaluated code. The delimiters are usually quotation marks, too.

- When viewing DNA as a language, a similar mechanism appears to be in place. In a CRISPR locus, short repeated sequences act as delimiters: the same thirty-odd bases return again and again, and set apart segments with a different purpose than the rest of the DNA. For instance, when a bacterium is infected by a virus, it can store part of the virus's DNA, in segments named "spacers" (second-order expression), into its own DNA (first-order expression). Precisely because they are second-order expressions, these spacers are not evaluated but ultimately (after their transcription) used as signatures by Cas proteins (Cas9 in the best-known systems) to detect known invading viruses. They act as a genetic memory (*i.e.*, they represent information). More generally, CRISPR embodies this duality.



In the scientific literature, analogies with human language focus mainly on palindromes (for example, this parallel with the Sator square, about the palindromic motifs that regulate genes). The "P" of CRISPR stands for palindromic because many of these repeats are partly palindromic, yet what delimits a spacer is the repetition: the same sequence before it and after it. *I believe a more meaningful analogy would be to compare them with quotation marks.* Because genetic code has a limited set of "symbols," it cannot express the delimiters with a special base: a fixed sequence of ordinary bases constitutes the delimiter. It is as if the English language did not use special characters but a fixed sequence of letters instead: `She told him xyyx It is raining outside xyyx and therefore . . .` (In this example, the same `xyyx` opens and closes the second-order expression, as straight quotation marks do.)

When a Delimiter Appears Inside: Escaping

Because delimiters are written with ordinary symbols, one can turn up *inside* a second-order expression, where it must not act. Languages need a way to quote the quotation mark itself: English alternates “ ” and ‘ ’, Python writes `'It\'s raining'`, and a program that forgets to do it with the text it receives is open to SQL injection. Programmers know this family of problems as code versus data, quoting and evaluating, in-band signalling, escaping.

Summary

The following table summarizes this idea.

Language	Symbols	First-order expression	Second-order expression	Delimiter
Human language	Alphabet, numbers, diacritics, *etc.*	Conveys an information It is raining outside	Represents an information She told him: "It is raining outside"	Quotation marks " "
Programming language	Subset of the human alphabet	Is evaluated LISP `(print (+ 1 1))` → 2	Is represented LISP `(print '(+ 1 1))` → (+ 1 1)	Special character LISP `
Genetic code	Adenine (A), Cytosine (C), Guanine (G), Thymine (T)	Is evaluated DNA of a bacteria	Represents an information Fragment of foreign DNA stored in bacteria DNA ("spacer")	Palindrome-like sequence 5′ **TAATTGAATTA** 3′ 3′ **ATTAACTTAAT** 5′

One duality, three languages.

Open Questions About AI

There are certainly more examples, and **I wonder in particular whether there is a connection between artificial intelligence and this duality**:

- Can models infer *by themselves* the existence of first- and second-order expressions and the role of delimiters in switching between them?
- Could this duality be leveraged to reinforce models' security by making them inherently distinguish pure knowledge representation from information evaluation during the training stage?

(Update November 2026) There does seem to be a connection. I explore it further in my later articles on special tokens in LLMs and XML tags in Claude.

Series and Companion Articles

Delimiters in Languages (glthr.com/page/delimiters-in-languages)

#	Article	Address
1	**What Quotation Marks, String Literals and DNA Have in Common** (this article)	glthr.com/quotation-marks-string-literals-and-dna
2	Special Tokens: The Quotation Marks of LLMs	glthr.com/special-tokens-the-quotation-marks-of-llms
3	Formulaic Delimiters in the Iliad and the Odyssey	glthr.com/formulaic-delimiters-in-the-iliad-and-the-odyssey
4	Why XML Tags Are So Fundamental to Claude	glthr.com/xml-fundamental-to-claude
5	Transitive Policies for Agents	glthr.com/transitive-policies-for-ai-agents

How to cite this article

> Guillaume Lethuillier, “What Quotation Marks, String Literals and DNA Have in Common”, glthr.com, July 6, 2024.
> https://glthr.com/quotation-marks-string-literals-and-dna

For BibTeX:

```
@misc{lethuillier2024quotation,
  author = {Lethuillier, Guillaume},
  title  = {What Quotation Marks, String Literals and DNA Have in Common},
  year   = {2024},
  month  = jul,
  url    = {https://glthr.com/quotation-marks-string-literals-and-dna},
  note   = {glthr.com}
}
```

This document was typeset with LaTeX from the article published at https://glthr.com/quotation-marks-string-literals-and-dna, which remains the reference version (as updated on November 15, 2026).