Writing Sentences for LLMs

How NLP systems and large language models read a sentence, why many sentences fall apart when they're lifted out of a page, and eight rules for writing ones that don't.

Teaching series · Lesson 6 · beginner to intermediate

A SENTENCE A MACHINE CAN LIFT WITHOUT LOSING MEANING

The Vatican Museums tour	lasts	three hours	and		
ENTITY · SUBJECT	VERB	VALUE			
includes	the Sistine Chapel	,	except on Sundays	.	
VERB	ENTITY · OBJECT		CONDITION		

Named subject, precise verbs, concrete value, condition attached. Illustrative example; check real opening days before using it.

50	8	1
minutes	sentence rules	lift test exercise

Ameneh Saeednia
Senior SEO and Generative Search Strategist

Before you start

What this lesson covers

AI search rarely uses your whole page. It **retrieves pieces**, often a few sentences, and builds an answer from them. Every sentence that matters has to make sense **on its own**.

By the end of this lesson you can

- Name the main steps an **NLP pipeline** takes to read a sentence.
- Explain how **retrieval-augmented** AI answers use passages from your page.
- Spot sentences that **break when lifted** out of context.
- Apply **eight rules** for writing sentences machines and people can use.
- Read the evidence from my own **retrieval study** without overclaiming.

Before this lesson

Lessons 2 and 5 help: **triples** explain sentence structure, and **entities** explain naming. Neither is required.

How to use it

Read Parts 1 to 7, do the **lift test on page 12**, then take the **quiz on page 13**. Answers are on the last page.

Part 1

How a machine reads a sentence

Classic NLP breaks reading into steps. LLMs learn most of these implicitly, but the steps still explain **what makes a sentence easy or hard to interpret**. [1]

Step	What it does	What it means for your writing
Tokenization	Splits text into tokens, often word pieces	Unusual spellings and invented brand terms split into fragments. **Spell names consistently.**
Part-of-speech tagging	Labels nouns, verbs, adjectives	Vague verbs like “provides” carry little meaning. **Verbs are relationships.**
Dependency parsing	Finds the main verb and what attaches to it	Long sentences with nested clauses hide **who does what to whom**.
Named entity recognition	Finds people, places, organizations, products	Entities you **name** can be recognized. “It” and “this” can't.
Coreference resolution	Links “it” or “the tour” back to a named entity	Works within a passage. **Breaks when the passage is cut** from its context.
Embedding	Turns the sentence into a vector of meaning	Sentences about one thing produce **clearer vectors** than sentences that mix topics.

The key idea

Most writing problems for machines are the **same problems readers have**, made worse by one thing: the machine may see your sentence **without the paragraph around it**.

Part 2

How LLM answers use your sentences

Many AI search features use **retrieval-augmented generation**: find relevant passages first, then generate an answer grounded in them. [2]

Stage	What happens	Where sentences fail
1. Chunking	The page is split into passages, often by headings or length	A claim and its condition land in **different chunks**.
2. Retrieval	Passages are matched to the query, by words and meaning	A passage that **never names the subject** matches poorly.
3. Selection	The most useful passages are chosen	A passage that needs the paragraph before it is **less useful alone**.
4. Generation	The model writes an answer from the selected text	Vague or hedged wording gets **paraphrased loosely** or dropped.

What Google says

Google says there are **no extra requirements** to appear in its AI features; normal SEO best practices apply. [3] Clear sentences are good writing first, and machine-friendly second.

Why it still matters

When a system lifts **two sentences** from your page, those two sentences are your whole argument. Write the important ones so they **stand alone**.

Part 3

Anatomy of a liftable sentence

A sentence that survives extraction usually has the same five parts.

1

Named subject

The entity, **by name**. Not “it,” “this option” or “our service.”

2

Precise verb

The relationship: **lasts, costs, includes, requires, is located in**.

3

Concrete object or value

What the verb points to, with **units**: three hours, €65, ages 6 and up.

4

Condition

The exception or scope, **in the same sentence**: except Sundays, for groups of up to 8.

5

Time or source

When it's true or who says so: **as of 2026**, according to the operator.

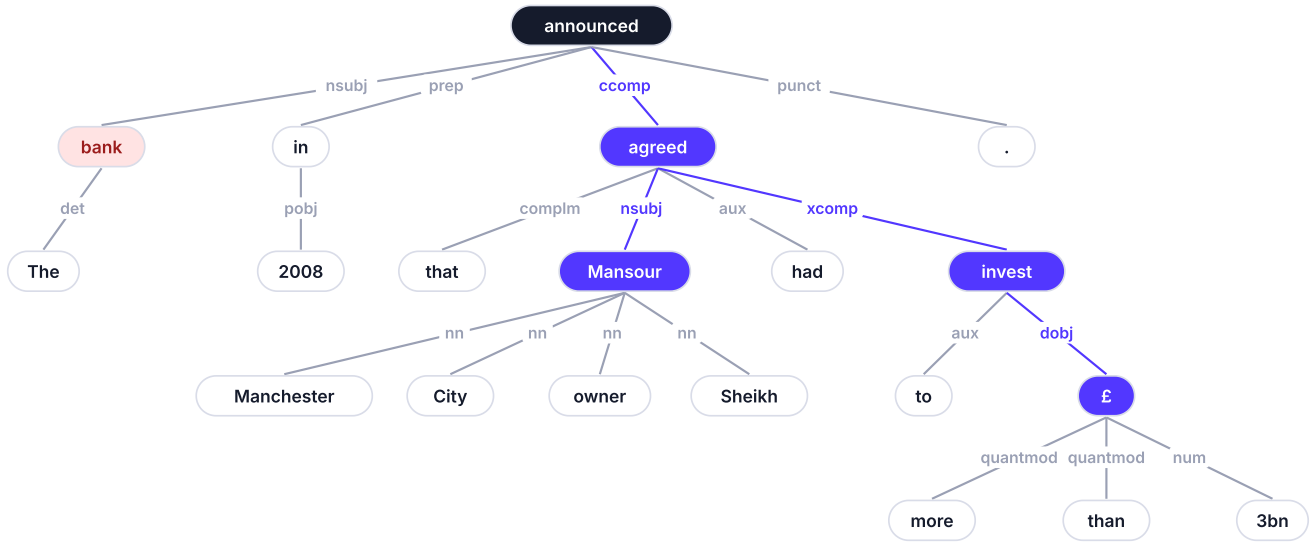
Built from the parts

“**The Vatican Museums tour** *(subject)* **lasts** *(verb)* **three hours** *(value)* and **includes the Sistine Chapel** *(object)*, **except on Sundays** *(condition)*.” Illustrative only; confirm real schedules before publishing a sentence like this.

Part 4

Reading a dependency parse

A dependency parser draws a sentence as a tree: the **root verb** at the top, and every other word attached to the word it depends on. Here's a real news-style sentence: “The bank announced in 2008 that Manchester City owner Sheikh Mansour had agreed to invest more than £3bn.”



What the tree shows

The **root is “announced”** and its subject is “bank.” So the main triple is *bank → announced → [something]*.

The actual news, **Mansour → agreed → to invest → £3bn**, sits two clauses down (ccomp, then xcomp), shown in violet.

Four words (Manchester, City, owner, Sheikh) all hang off **Mansour** as noun modifiers.

What it reveals for writing

“The bank” is unnamed (red). No entity to recognize or link.

Invest in what? The verb has money as its object but no “in [X].” The fact is incomplete.

The tool that drew the original tree showed the pound sign as **“$”**: parsers normalize symbols, so **write currencies as words or codes** where it matters.

Rewritten

“In 2008, **Sheikh Mansour, the owner of Manchester City, agreed to invest more than £3bn in [Bank name]**. The bank announced the investment the same year.” Now the root is **agreed**, the subject is a named person, and the investment has a target. [Bank name] is a placeholder; check the facts before reusing this example.

Part 5

Check your sentences with Google’s Natural Language demo

Google Cloud's free Natural Language demo shows how one NLP system reads your text. It's a **proxy**, not a view of Google Search, but it's the quickest way to see entities, sentiment and syntax for a sentence. [5]

Tab	What it shows	What to check in your sentence
Entities	Each entity with a **type** (person, organization, location, price, date…), a **salience** score from 0 to 1, and a Wikipedia link when it recognizes a known entity	Is the **main entity first** and most salient? Are names split or merged wrongly? Is anything important left as a common noun, like “the bank”?
Sentiment	A **score** from −1.0 (negative) to +1.0 (positive) and a **magnitude** for overall emotional strength, for the document and each sentence	Factual sentences should sit **near 0**. A score near 0 with **high magnitude** means mixed feelings, common in honest reviews.
Syntax	The **dependency tree**, part of speech and lemma for every word	Is the **root verb** the fact you meant? Is your subject a named entity? Do conditions attach to the right word?
Categories	Content categories with **confidence**	Does the category match the page's purpose? (See Lesson 1.)

How to run it

Paste a sentence or paragraph and select **Analyze**. The demo asks you to confirm you're not a robot before it runs.

Test **one sentence at a time** when you're checking structure, and a full section when you're checking entities and sentiment.

Reading sentiment carefully

Sentiment measures **tone, not truth or quality**. A glowing sentence can be wrong.

Magnitude grows with **text length**, so compare sentences of similar length.

Don't chase a positive score. For reviews, **balanced and specific** beats enthusiastic.

Part 6

Eight rules, with before and after

Rules 1 to 4: making the sentence stand alone.

BEFORE

It also includes skip-the-line entry.

AFTER · Rule 1: name, don't point

The Vatican Museums tour includes skip-the-line entry.

BEFORE

Our tours offer an unforgettable experience with flexible options and great value for families.

AFTER · Rule 2: one claim per sentence

Our family tours last two hours. Children under 6 join free.

BEFORE

We provide solutions for travel planning.

AFTER · Rule 3: precise verbs

We match travellers with a private local guide in their destination.

BEFORE

Cancellation is free. Some restrictions apply (see below).

AFTER · Rule 4: keep the condition with the claim

Cancellation is free up to 24 hours before the tour starts.

Part 6, continued

Eight rules, with before and after

Rules 5 to 8: making the sentence precise and consistent.

BEFORE

The tour is pretty long and fairly affordable.

AFTER · Rule 5: values with units

The tour lasts five hours and costs €65 per adult.

BEFORE

Prices recently went up.

AFTER · Rule 6: date what changes

As of March 2026, the adult price is €65.

BEFORE

The Old Town walk… this city stroll… our heritage experience…

AFTER · Rule 7: one entity, one name

The Old Town walking tour (same name everywhere on the page).

BEFORE

The jacket is waterproof. [Three paragraphs later:] Not recommended for heavy downpours.

AFTER · Rule 8: keep the claim and its limit together

The jacket is waterproof in light to moderate rain, not in heavy downpours.

Part 7

What my retrieval study found

In *Built to Be Retrieved*, I rewrote pages several ways and measured which passage a retriever picked in a small, controlled corpus. [4]

Rewrite tested	Result in my study
Short, self-contained sections	Retrieval scores improved for **84 to 92%** of questions, depending on method. The top passage contained the answer **20 to 29 percentage points** more often.
Answer-first opening	**No statistically significant** improvement over a plain paraphrase (p = .18 to .77).
Adding explicit entities only	**Smaller and mixed** effects on answer selection.
All combined	Similar to short self-contained sections.

What I take from it

The strongest change was making each section **answer on its own**: subject, claim and condition together. That's what the eight rules do at sentence level.

What it doesn't prove

It's a **small corpus**, and question-style headings may have helped matching. It doesn't prove an ideal length, and it says nothing about **rankings or citations** in live AI products.

Common mistakes

What doesn’t help

Some “AI-friendly” writing advice makes sentences worse.

1

Fragments without verbs

Bullet lists like “Family-friendly. 3 hrs. Rome.” lose the **relationships** between facts. Keep at least one full sentence.

2

Keyword repetition

Repeating a phrase doesn't make a passage clearer. It makes the **meaning noisier**.

3

Throat-clearing intros

“In this article, we will explore...” is often the **first sentence retrieved**, and it says nothing.

4

Over-hedging

“May potentially in some cases...” gets dropped or flattened. **State the condition** instead.

5

Splitting everything

Very short sentences can separate a claim from **its reason or limit**. Clarity beats shortness.

6

Writing for the machine only

If a sentence reads badly to a person, it's **not better for a model**. Both want the same thing.

Exercise

The lift test

Time: 15 minutes. Take one important page: a product, service, tour or policy page.

Step	What to do
1	Pick the **five sentences** a customer most needs: price, duration, what's included, conditions, location.
2	Copy each one **alone** into a blank document. Read it with no context.
3	Mark every **pronoun, missing condition, missing unit and vague verb**.
4	Rewrite each sentence with the **five parts from Part 3**.
5	Check the rewritten sentence uses the **same entity name** as the rest of the page.
6	Run each rewritten sentence through the **Natural Language demo**. Check the **root verb**, the **top entity** and that sentiment is near neutral for factual claims.
7	Optional: paste the page into an AI assistant and ask the five questions. Note which **sentences it quotes or paraphrases**.

What good looks like

Each sentence answers **one customer question correctly on its own**, with no “it,” no missing condition and no missing unit.

Discussion question

Which of your five sentences was **most wrong** when lifted alone? What would a customer have misunderstood?

Check your understanding

Quick quiz

1. Which sentence survives being lifted out of a page best?

a) It also includes a free tasting.
b) The Trastevere food tour includes a free gelato tasting.
c) This one is great too.

2. Coreference resolution breaks most often when:

a) The passage is cut from its context
b) Sentences name their subject
c) The page has headings

3. Where should a condition like “except Sundays” go?

a) In a footnote
b) In the same sentence as the claim
c) In the FAQ only

4. In my retrieval study, answer-first openings:

a) Showed no statistically significant improvement
b) Doubled retrieval scores
c) Were not tested

5. Google says appearing in its AI features requires:

a) Special AI markup
b) No extra requirements beyond SEO best practices
c) Short sentences only

6. In the parse on page 6, the root verb of the original sentence is:

a) invest
b) announced
c) agreed

7. A sentiment score near 0 with high magnitude usually means:

a) The text is neutral and short
b) The text has mixed strong feelings
c) The text is factually wrong

Answers and sources

Answer key, glossary and further reading

Answer key

1: **b**. Named subject, precise verb, concrete object.
2: **a**.
3: **b**.
4: **a**. Short self-contained sections did better.
5: **b**.
6: **b**. The news itself sits two clauses down.
7: **b**.

Glossary

Token: a word or word piece a model processes.
Dependency parsing: finding how words relate to the main verb.
Coreference: linking “it” to what it refers to.
Chunk: a passage split from a page for retrieval.
Root verb: the top of a dependency tree.
Magnitude: strength of emotion, positive or negative.
Liftable sentence: one that stays true and clear alone.

01 **Speech and Language Processing, 3rd edition draft. Jurafsky and Martin**
Standard reference for tokenization, tagging, parsing, NER and coreference.
https://web.stanford.edu/~jurafsky/slp3/

02 **Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Lewis et al., 2020**
The paper that named retrieval-augmented generation.
https://arxiv.org/abs/2005.11401

03 **AI features and your website. Google Search Central**
No additional requirements for AI features beyond SEO best practices.
https://developers.google.com/search/docs/appearance/ai-features

04 **Built to Be Retrieved. SEO Girl**
My controlled passage retrieval study, and its limits.
https://www.seogirl.ca/research/built-to-be-retrieved

05 **Natural Language API: sentiment, entity and syntax analysis. Google Cloud**
Score, magnitude, salience and dependency output used in Part 5.
https://cloud.google.com/natural-language/docs/basics

Lesson by Ameneh Saeednia, prepared with AI assistance. Tour and product sentences are illustrative. Related reading on seogirl.ca: *NLP for SEO*, *Part of Speech and SEO* and *What Happens When Your Best Paragraph Loses Its Context?*