Discourse Cohesion

Topic-wise study notes | Natural Language Processing

Covers: definition of cohesion, lexical cohesion, grammatical cohesion, and cohesive devices. Each topic includes characteristics, a conceptual architecture, a worked example, and applications.

Learning outcomes

• Define cohesion and distinguish it from coherence.

• Identify lexical and grammatical links across sentences.

• Explain reference, substitution, ellipsis, and conjunction.

• Analyse a paragraph step by step and describe uses in NLP.

Guide to the notes

Section	Content
1	Definition and characteristics of discourse cohesion
2	Lexical cohesion: repetition, semantic relations, collocation
3	Grammatical cohesion: reference, substitution, ellipsis, conjunction
4	Cohesive devices: classification and selection
5	Complete paragraph analysis and NLP workflow
6	Revision questions, answers, and reading list

What does “architecture” mean here?

Cohesion is a linguistic concept, so it has no single fixed computer architecture. The diagrams in these notes show conceptual structures and illustrative analysis workflows; they are not a prescribed neural network model.

1. Definition of Cohesion

Definition

Cohesion is the network of linguistic links that connects parts of a text. A cohesive relation occurs when interpreting one expression depends on another expression in the discourse. These links can occur within a sentence and across sentence boundaries.

Example: “Ravi bought a book. He read it.” Here, **He** links to Ravi and **it** links to a book.

Characteristics

• Creates ties between words, phrases, clauses, and sentences.

• Uses vocabulary relations and grammatical resources.

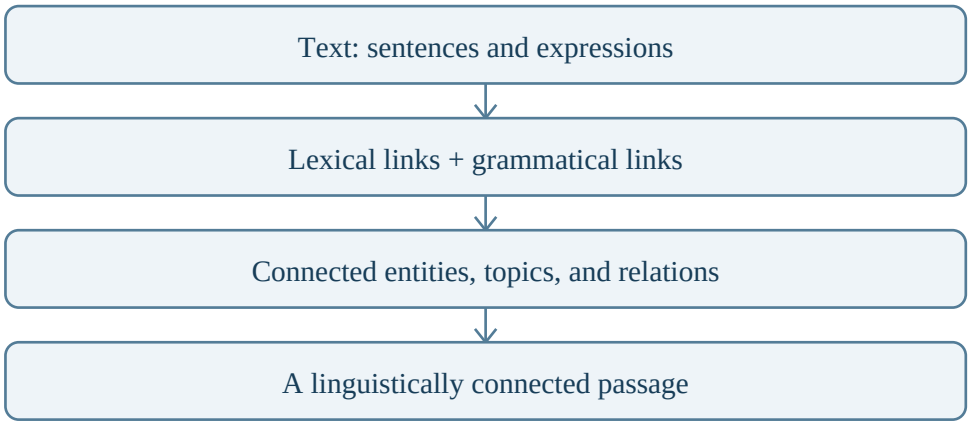
• Supports continuity of people, objects, actions, and topics.

• May link to earlier text (anaphora) or later text (cataphora).

• Can connect nearby or distant parts of a passage.

• Helps interpretation, but does not by itself guarantee meaningful or logical writing.

Conceptual architecture



Step-by-step example

• **Step 1:** Read: “Meena planted a tree. The plant needs water. She waters it daily.”

• **Step 2:** Identify entities: Meena and a tree.

• **Step 3:** Link “the plant” to the tree: a general lexical term plus contextual reference.

• **Step 4:** Resolve “She” as Meena and “it” as the tree.

• **Step 5:** Connect “water” and “waters”: related forms supporting topic continuity.

Applications

Cohesion supports reading comprehension, paragraph writing, discourse analysis, reference resolution, summarization, and the evaluation of connected text.

1.1 Cohesion and Coherence

Cohesion	Coherence
Linguistic links expressed through wording and grammar.	Overall meaningfulness and logical unity of a text.
Observed through pronouns, repetition, omission, and connectors.	Interpreted through context, world knowledge, purpose, and relations.
Example: “Asha bought a laptop. She uses it.”	The two events make sense as a connected account.

Cohesion without strong coherence

“Ravi bought a car. The car was blue. Therefore, the moon is made of cheese.” Repetition connects the first two sentences, and “therefore” signals a relation. However, the final conclusion does not logically follow. A connector cannot create a valid argument by itself.

Coherence with few explicit cohesive devices

“The doorbell rang. Asha went to the door.” No pronoun or explicit causal connector is required. Readers infer that the ringing prompted her action.

Important terms

Term	Meaning and example
Cohesive tie	A link between expressions: “a bicycle ... it”.
Antecedent	Expression referred back to: “Ravi” in “Ravi arrived. He smiled.”
Anaphora	Backward reference: “Sita arrived. She sat down.”
Cataphora	Forward reference: “Before she spoke, Sita checked the microphone.”
Exophora	Reference outside the text: “Put it there,” while pointing. It relies on the situation and is not a text-internal cohesive tie.

Exam-ready definition

Discourse cohesion refers to lexical and grammatical relations that connect parts of a text and help readers interpret them as a linked whole. Major resources include lexical repetition and semantic relations, reference, substitution, ellipsis, and conjunction.

2. Lexical Cohesion

Definition

Lexical cohesion connects a text through relationships among content words. It is commonly described through **reiteration** (repetition and related lexical expressions) and **collocation** (words associated in a field or situation).

Characteristics

• Depends on vocabulary and semantic associations.

• Maintains the topic through repeated or related words.

• Can create lexical chains extending across several sentences.

• Includes exact repetition and relations such as synonymy and class membership.

• Is context-sensitive: related words do not automatically refer to the same entity.

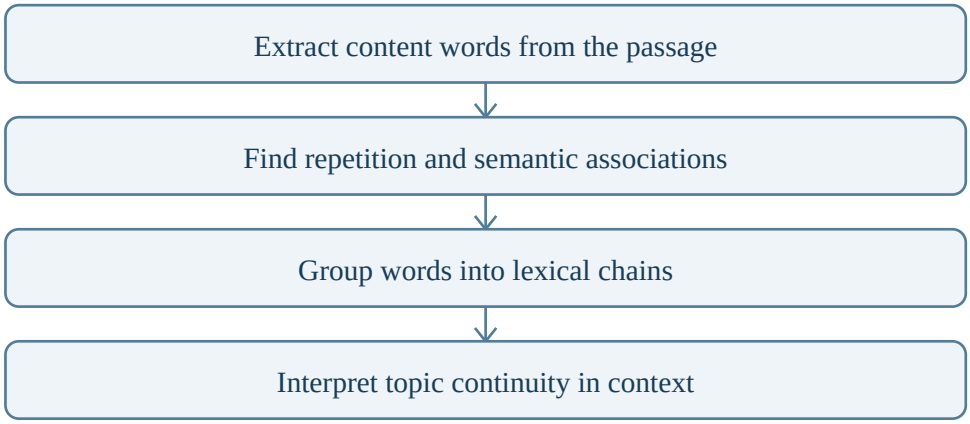
Main forms with examples

Form	Definition and example
Repetition	Repeating a word: “The river flooded. The river damaged nearby farms.”
Synonym / near-synonym	Related meaning: “The child was joyful. The happy youngster laughed.”
Superordinate / hyponym	General and specific categories: “A rose bloomed. The flower attracted bees.”
General word	Broad expression: “A truck arrived. The vehicle stopped outside.”
Antonymy	Contrast in meaning: “Some days are hot; others are cold.”
Part-whole relation	Associated whole and component: “The car stopped. Its engine had failed.”
Collocation	Words associated with a setting: “doctor, patient, diagnosis, treatment”.

Classification note: Extended teaching taxonomies often list antonymy and part-whole relations separately. In broader traditional accounts, these can be discussed among lexical relations or collocational associations.

2.1 Lexical Cohesion: Working

Conceptual architecture



Step-by-step example

Passage: “The school opened a new library. The library contains many books. These volumes help students learn. The reading room also has journals.”

• **Step 1 - Extract words:** school, library, books, volumes, students, reading room, journals.

• **Step 2 - Find repetition:** “library” repeats in sentences 1 and 2.

• **Step 3 - Find related naming:** “volumes” is a contextual alternative for books.

• **Step 4 - Find associated vocabulary:** library, books, reading room, and journals belong to the same reading/study setting.

• **Step 5 - Form a chain:** library - books - volumes - reading room - journals.

• **Step 6 - Interpret:** The lexical chain keeps the passage focused on library resources. “These” and “also” add grammatical links.

Applications

• **Topic tracking:** lexical chains help identify continuing subjects.

• **Summarization:** repeated and related terms can help locate central information.

• **Information retrieval:** synonym and category relations can support relevant search matches.

• **Text segmentation:** changes in lexical patterns can suggest topic boundaries.

• **Teaching writing:** controlled repetition and accurate alternative wording improve continuity.

Common mistake

Do not replace every repeated word with a synonym. Technical terms often need consistent repetition. Also, “car” and “vehicle” may be related words without referring to the same object in every context.

3. Grammatical Cohesion

Definition

Grammatical cohesion connects parts of a text through grammatical resources: reference, substitution, ellipsis, and conjunction. Conjunction is often included here in introductory teaching, although its main function is to express semantic relations between clauses or messages.

Characteristics

• Often uses pronouns, demonstratives, comparison, replacement forms, and connectors.

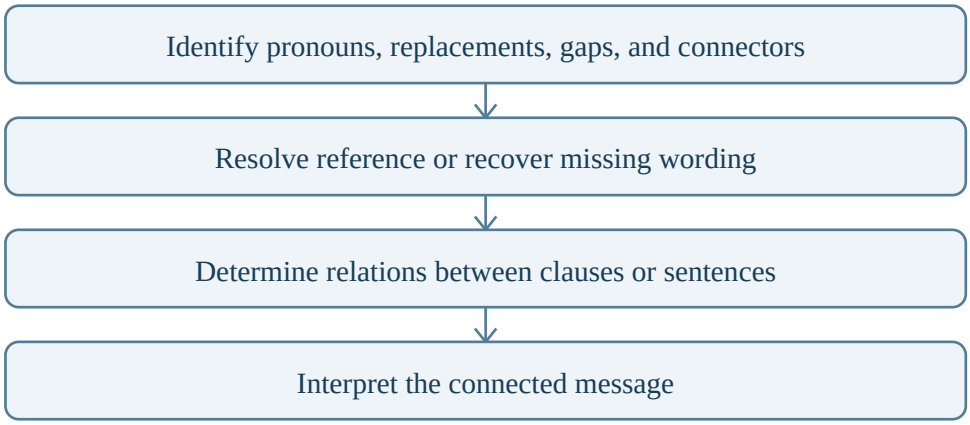
• Reduces unnecessary repetition while preserving recoverable meaning.

• Requires context to identify referents or omitted expressions.

• Expresses relations such as addition, contrast, cause, and sequence.

• Ambiguous references and unclear omissions can weaken a passage.

Conceptual architecture



The four major types

Type	Function and example
Reference	Points to an entity or comparison: “Anu has a pen. She likes it.”
Substitution	Replaces a linguistic item: “I need a pen. Do you have one?”
Ellipsis	Omits recoverable wording: “Ravi ordered tea; Sita, coffee.”
Conjunction	Signals a relation: “It rained. Therefore, the match was cancelled.”

Applications

Grammatical cohesion is useful for coreference resolution, dialogue understanding, translation, question answering, text generation, and the teaching of clear sentence connections.

3.1 Reference and Substitution

A. Reference: definition and types

Reference directs the reader to an entity or comparison needed to interpret an expression. Personal reference uses forms such as he, she, it, and they; demonstrative reference uses this, that, these, those, and context-dependent “the”; comparative reference uses forms such as same, different, and better.

Reference: step-by-step example

“Priya borrowed a laptop. She used it for her project. This device was faster than her old one.”

• **Step 1:** Introduce Priya and a laptop.

• **Step 2:** Resolve “She” as Priya.

• **Step 3:** Resolve “it” as the borrowed laptop.

• **Step 4:** Resolve “This device” as the borrowed laptop.

• **Step 5:** Interpret “faster” as a comparison between the borrowed laptop and Priya’s old laptop; “one” substitutes for laptop.

Characteristic: Reference concerns what an expression points to. Ambiguity is possible: “Ravi met Arun. He was tired” does not clearly identify who was tired.

B. Substitution: definition and types

Substitution replaces a word, phrase, or clause with a substitute form rather than repeating the original wording. Common forms include nominal “one/ones”, verbal “do”, and clausal “so/not”.

Type	Example
Nominal	“I prefer the red pen to the blue one.” One substitutes for pen.
Verbal	“Ravi plays chess, and Anu does too.” Does substitutes for plays chess.
Clausal	“Will the train arrive soon?” “I think so.” So substitutes for the proposition that it will arrive soon.

Substitution: step-by-step example

“I need a notebook. Do you have one?” (1) Identify notebook as the earlier noun. (2) Locate one. (3) Replace one mentally with a notebook. (4) Interpret the question as asking whether a notebook is available; it need not be the same physical notebook.

3.2 Ellipsis and Conjunction

A. Ellipsis

Definition: Ellipsis is the omission of wording that can be recovered from context. It can affect nominal, verbal, or clausal material. It is sometimes described as substitution by zero.

Characteristics: It avoids repetition, depends on recoverability, and is common in dialogue and coordinated structures.

Ellipsis: step-by-step example

“Ravi ordered tea, and Sita coffee.”

• **Step 1:** Identify the complete clause: Ravi ordered tea.

• **Step 2:** Notice that the second clause lacks an explicit verb.

• **Step 3:** Recover “ordered” from the parallel structure.

• **Step 4:** Expand it: “Ravi ordered tea, and Sita ordered coffee.”

Other examples: “How many pens do you need?” “Two [pens].” “Have you finished?” “Yes, I have [finished].” Brackets show recovered material and are not part of the original utterance.

B. Conjunction

Definition: Conjunction connects clauses or messages by indicating a relation such as addition, contrast, cause, or time. Some cohesive connectors, such as “however”, are adverbs rather than grammatical conjunctions.

Relation	Typical devices
Additive	and, also, moreover, in addition
Adversative / contrast	but, however, nevertheless, on the other hand
Causal / result	because, so, therefore, as a result
Temporal / sequence	first, next, then, finally, afterwards

Conjunction: step-by-step example

“Heavy rain flooded the road. Therefore, the bus was delayed.” (1) Identify the first event: flooding. (2) Identify the second event: delay. (3) Locate “therefore”. (4) Interpret the delay as a consequence of the flooding.

Applications: Ellipsis recovery supports dialogue interpretation; conjunction analysis supports argument understanding, summarization, and translation of discourse relations.

4. Cohesive Devices

Definition

Cohesive devices are the linguistic resources used to connect a text. The term includes lexical resources as well as grammatical resources; it is broader than “linking words”.

Characteristics

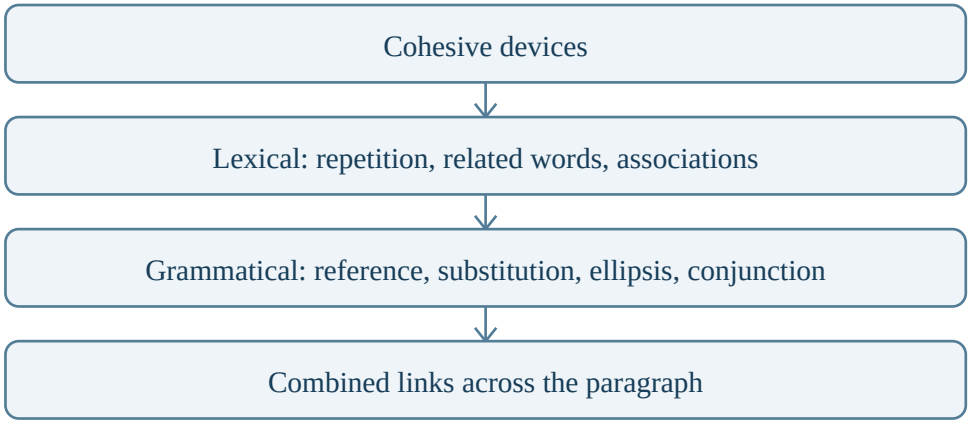
• Signal continuity, reference, contrast, cause, sequence, and other relations.

• Can be explicit words, lexical patterns, or recoverable omissions.

• Must match the intended relation and the surrounding context.

• Improve clarity when used accurately; excessive connectors can make writing mechanical.

Classification architecture



The middle two boxes represent complementary categories, not sequential processing stages.

Step-by-step paragraph improvement

Before: “Asha joined a course. Asha attended classes. Asha found the course difficult. Asha completed the course.”

• **Step 1:** Keep the first mention of Asha and the course.

• **Step 2:** Use “She” for later references to Asha.

• **Step 3:** Use “it” when the course is the clear referent.

• **Step 4:** Add “However” to express contrast between difficulty and completion.

After: “Asha joined a course. She attended classes and found it difficult. However, she completed it.”

Applications

Use cohesive devices in academic writing, instructions, reports, narratives, classroom explanations, and generated responses. Select them according to the actual relationship between ideas.

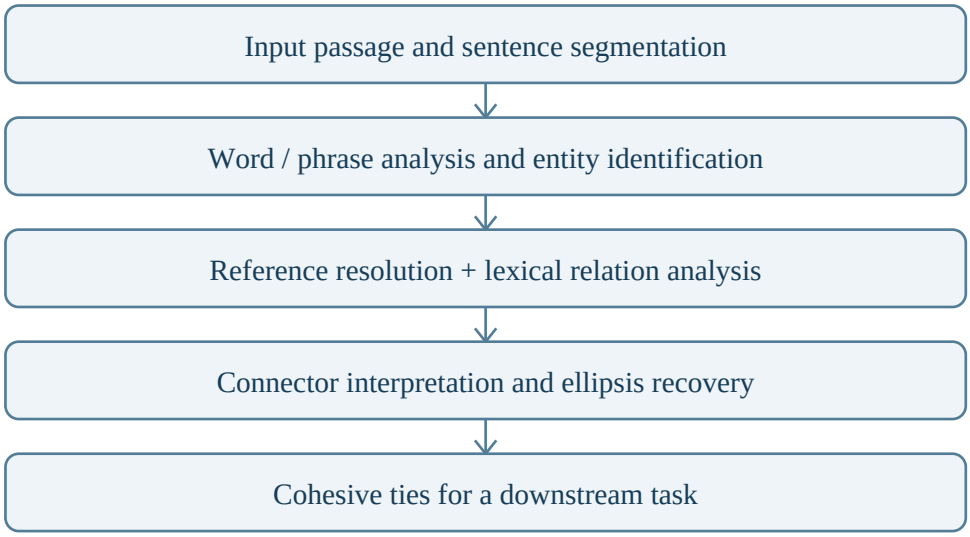
5. Complete Worked Analysis

Paragraph: “Anita visited a library. The library had many books. She selected a novel and borrowed it. Her friend wanted one too, but the librarian said that no other copies were available. Therefore, they decided to return the next day.”

Step-by-step analysis

Step	Identify and explain
1. Entities	Anita, a library, books, a novel, her friend, and the librarian.
2. Repetition	Library repeats in sentence 2, maintaining the setting.
3. Lexical relations	Novel is a type of book. Library, books, copies, borrowed, and librarian are associated with the borrowing setting.
4. Reference	She = Anita; it = the selected novel; Her = Anita; they = Anita and her friend. “The library” refers back to the introduced library.
5. Substitution	One substitutes for a novel. The friend wants a novel too; context may suggest the same title, but the substitute alone does not establish that.
6. Conjunction	And links actions; too adds a parallel desire; but signals contrast; therefore introduces the resulting decision.
7. Time and interpretation	The next day supplies a temporal relation. No clear ellipsis occurs in this paragraph. The librarian is understood through the library setting.

Illustrative NLP analysis architecture



Practical limitation: A real system may combine rules, lexical resources, and learned models. Pronoun ambiguity, implicit relations, and domain-specific vocabulary require contextual reasoning. Counting connectors alone is not a reliable quality measure.

6. Revision and Practice

Quick comparison

Device	Diagnostic question
Lexical cohesion	Are repeated or semantically associated content words linking the text?
Reference	What person, object, or comparison does this expression point to?
Substitution	Which earlier word, phrase, or proposition is replaced?
Ellipsis	Which recoverable wording is omitted?
Conjunction	What relation between ideas is signalled?

Practice questions

• 1. Define cohesion and explain why it does not guarantee coherence.

• 2. Identify the devices: “A dog entered the garden. The animal was hungry. It ate the food.”

• 3. Analyse: “I wanted a blue shirt, but I bought a white one.”

• 4. Recover the omission: “Meena chose physics; Ravi, chemistry.”

• 5. Write a short paragraph using lexical cohesion, reference, and a causal connector.

Answer hints

1. Cohesion provides linguistic ties; coherence requires meaningful overall relations. **2.** Dog-animal is a specific-to-general lexical relation; “The animal” and “It” refer to the dog. **3.** “But” expresses contrast; “one” substitutes for shirt; blue and white belong to the colour vocabulary field. **4.** Recover “chose”: Ravi chose chemistry. **5.** Example: “A student missed the bus. She waited for another vehicle. Therefore, she arrived late.”

Key takeaways

Cohesion = linguistic connections. Lexical cohesion = vocabulary links. Grammatical cohesion = reference, substitution, ellipsis, and conjunction. Cohesive devices = the complete set of these resources. Always analyse the device, its target or relation, and its contribution to the passage.

Foundational reading

Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English*. Longman.
Brown, G., and Yule, G. (1983). *Discourse Analysis*. Cambridge University Press.
These notes paraphrase standard concepts; all worked examples are illustrative teaching examples.