T. Y. Artificial Intelligence & Data Science
Semester VI

Prof. Mandar Diwakar
Asst. Professor
Department of AI & DS
Vishwakarma Institute of Information Technology

ADUA32203: Natural Language Processing

Unit III

Part of Speech Tagging and Parsing

Unit III:	Part of Speech Tagging and Parsing	7 Hrs

Word Classes and Part-of-Speech tagging, survey of POS tag sets for English and Indian Languages, Rule based PoS tagging, Transformation-Based Tagging, Evaluation and Error Analysis, Advanced Issues in Part-of-Speech Tagging,

Introduction to CFG, Parsing with Context-Free Grammars: Parsing as Search, Top-Down Parsing, Bttom-Up Parsing, Dynamic Programming Parsing Methods: CKY Parsing, The Earley Algorithm, Partial Parsing: Finite-State Rule-Based Chunking Machine Learning-Based Approaches to Chunking .

Natural Language Processing

WORD CLASSES AND PART OF SPEECH TAGGING

Part of Speech Tagging

- Words are traditionally grouped into equivalence classes called ***parts of speech*** (POS), ***word classes***, ***morphological classes***, or ***lexical tags***.

- The part of speech for a word gives a ***significant amount of information about the word and its neighbors.***

- A word's part-of-speech can tell us something about ***how the word is pronounced***.

- This can be useful in a language model for ***speech recognition***.

- ***Part-of-speech (POS) tagging*** is a popular Natural Language Processing process which refers to ***categorizing words in a text (corpus) in correspondence with a particular part of speech***, depending on the definition of the word and its context.

Part of Speech Tagging

- Parts of speech can also be used in ***stemming for informational retrieval (IR)***

- Parts of speech are very often used for '***partial parsing***' ***texts***, for example for quickly ***finding names or other phrases for the information extraction applications.***

- Three ***POS tagging*** algorithms:

1. Rule-based tagging.
2. Stochastic tagging.
3. Transformation-based tagging.

Part of Speech Tagging

- Parts of speech can be divided into two broad super categories, ***closed class types*** and ***open class types***.

- There are four major ***open classes*** that occur in the languages of the world: ***nouns***, ***verbs, adjectives***, and ***adverbs***.

- ***The closed classes***

✓ **prepositions**: on, under, over, near, by, at, from, to, with
✓ **determiners**: a, an, the
✓ **pronouns**: she, who, I, others
✓ **conjunctions**: and, but, or, as, if, when
✓ **auxiliary verbs**: can, may, should, are
✓ **particles**: up, down, on, off, in, out, at, by
✓ **numerals**: one, two, three, first, second, third

Part of Speech Tagging

• English also has many words of more or less unique function,

✓ **interjections** (oh, ah, hey, man, alas),

✓ **negatives** (no, not),

✓ **politeness markers** (please, thank you),

✓ **greetings** (hello, goodbye)

Part of Speech

1. NOUN

- **A noun is the name of a person, place, thing, or idea.**

man... Butte College... house... happiness

A noun is a word for a person, place, thing, or idea. Nouns are often used with an article (*the*, *a*, *an*), but not always. Proper nouns always start with a capital letter; common nouns do not. Nouns can be singular or plural, concrete or abstract. Nouns show possession by adding *'s*. Nouns can function in different roles within a sentence; for example, a noun can be a subject, direct object, indirect object, subject complement, or object of a preposition.

*The young **girl** brought me a very long **letter** from the **teacher**, and then she quickly disappeared. Oh my!*

Part of Speech

2. PRONOUN

- **A pronoun is a word used in place of a noun.**

She... we... they... it

A pronoun is a word used in place of a noun. A pronoun is usually substituted for a specific noun, which is called its antecedent. In the sentence above, the antecedent for the pronoun *she* is the girl. Pronouns are further defined by type: personal pronouns refer to specific persons or things; possessive pronouns indicate ownership; reflexive pronouns are used to emphasize another noun or pronoun; relative pronouns introduce a subordinate clause; and demonstrative pronouns identify, point to, or refer to nouns.

*The young girl brought **me** a very long letter from the teacher, and then **she** quickly disappeared. Oh my!*

Part of Speech

3. VERB

- **A verb expresses action or being.**

jump... is... write... become

The verb in a sentence expresses action or being. There is a main verb and sometimes one or more helping verbs. ("*She can sing.*" *Sing* is the main verb; *can* is the helping verb.) A verb must agree with its subject in number (both are singular or both are plural). Verbs also take different forms to express tense.

*The young girl **brought** me a very long letter from the teacher, and then she quickly **disappeared**. Oh my!*

Part of Speech

4. ADJECTIVE

- **An adjective modifies or describes a noun or pronoun.**

pretty... old... blue... smart

An adjective is a word used to modify or describe a noun or a pronoun. It usually answers the question of which one, what kind, or how many. (Articles [a, an, the] are usually classified as adjectives.)

*The **young** girl brought me a very **long** letter from the teacher, and then she quickly disappeared. Oh my!*

POS tag sets for English

5. ADVERB

- **An adverb modifies or describes a verb, an adjective, or another adverb.**

gently... extremely... carefully... well

An adverb describes or modifies a verb, an adjective, or another adverb, but never a noun. It usually answers the questions of when, where, how, why, under what conditions, or to what degree. Adverbs often end in -ly.

*The young girl brought me a **very** long letter from the teacher, and **then** she **quickly** disappeared. Oh my!*

Part of Speech

6. PREPOSITION

- **A preposition is a word placed before a noun or pronoun to form a phrase modifying another word in the sentence.**

by... with.... about... until

(by the tree, with our friends, about the book, until tomorrow)

A preposition is a word placed before a noun or pronoun to form a phrase modifying another word in the sentence. Therefore a preposition is always part of a prepositional phrase. The prepositional phrase almost always functions as an adjective or as an adverb. The following list includes the most common prepositions:

*The young girl brought me a very long letter **from the teacher**, and then she quickly disappeared. Oh my!*

Part of Speech

7. CONJUNCTION

- **A conjunction joins words, phrases, or clauses.**

and... but... or... while... because

A conjunction joins words, phrases, or clauses, and indicates the relationship between the elements joined. Coordinating conjunctions connect grammatically equal elements: and, but, or, nor, for, so, yet. Subordinating conjunctions connect clauses that are not equal: because, although, while, since, etc. There are other types of conjunctions as well.

The young girl brought me a very long letter from the teacher, ***and*** *then she quickly disappeared. Oh my!*

Part of Speech

8. INTERJECTION

- **An interjection is a word used to express emotion.**

Oh!... Wow!... Oops!

An interjection is a word used to express emotion. It is often followed by an exclamation point.

The young girl brought me a very long letter from the teacher, and then she quickly disappeared. ***Oh my****!*

Part of Speech Tagging

- ***Steps Involved in the POS tagging***

• **Collect a dataset of annotated text**: This dataset will be used to train and test the POS tagger. The text should be annotated with the correct POS tags for each word. • **Preprocess the text**: This may include tasks such as tokenization (splitting the text into individual words), lowercasing, and removing punctuation.
• **Divide the dataset into training and testing sets**: The training set will be used to train the POS tagger, and the testing set will be used to evaluate its performance. • **Train the POS tagger**: This may involve building a statistical model, such as a hidden Markov model (HMM), or defining a set of rules for a rule-based or transformation based tagger. The model or rules will be trained on the annotated text in the training set.
• **Test the POS tagger**: Use the trained model or rules to predict the POS tags of the words in the testing set. Compare the predicted tags to the true tags and calculate metrics such as precision and recall to evaluate the performance of the tagger.

Part of Speech Tagging

• ***Steps Involved in the POS tagging***

• **Fine-tune the POS tagger**: If the performance of the tagger is not satisfactory,

adjust the model or rules and repeat the training and testing process until the desired level of accuracy is achieved.

- **Use the POS tagger**: Once the tagger is trained and tested, it can be used to perform POS tagging on new, unseen text. This may involve preprocessing the text and inputting it into the trained model or applying the rules to the text. The output will be the predicted POS tags for each word in the text.

Use of Parts of Speech Tagging in NLP

- **To understand the grammatical structure of a sentence:** By labeling each word with its POS, we can better understand the syntax and structure of a sentence. This is useful for tasks such as machine translation and information extraction, where it is important to know how words relate to each other in the sentence.
- **To disambiguate words with multiple meanings:** Some words, such as "bank," can have multiple meanings depending on the context in which they are used. By labeling each word with its POS, we can disambiguate these words and better understand their intended meaning.
- **To improve the accuracy of NLP tasks:** POS tagging can help improve the performance of various NLP tasks, such as named entity recognition and text classification. By providing additional context and information about the words in a text, we can build more accurate and sophisticated algorithms.
- **To facilitate research in linguistics:** POS tagging can also be used to study the patterns and characteristics of language use and to gain insights into the structure and function of different parts of speech.

Part of Speech Tagging

Part of Speech Tagging

Rule based PoS tagging

➢ One of the oldest techniques of tagging is rule-based POS tagging.

➢ Rule-based taggers **use a dictionary** (i.e. it can store a number of words) or **lexicon for getting possible tags** for tagging each word.

➢ If the word has **more than one possible tag**, then rule-based taggers use **hand-written rules** to identify the correct tag.

➢ For example, suppose if the preceding word of a word is an article then the word must be a noun.

Rule based PoS tagging

➢ We can also understand Rule-based POS tagging by its two-stage architecture

1. The first stage

In the first stage, **create a dictionary** to assign each word a list of potential parts-of-speech.

2. The second stage

In the second stage, **create a large list of hand-written disambiguation rules** to sort down the list to a single part-of-speech for each word as per requirements.

Rule based PoS tagging

➢ **Rule-based POS taggers possess the following properties –**

1. These taggers are knowledge-driven taggers.
2. The rules in Rule-based POS tagging are built manually.
3. The information is coded in the form of rules.

4. We have a limited number of rules approximately around 1000. 5. Smoothing and language modeling is defined explicitly in rule-based taggers.

Rule based PoS tagging

Rule based PoS tagging

Stochastic POS Tagging

- The model that includes **frequency or probability (statistics)** can be called **stochastic**.
- Any number of different approaches to the problem of part-of-speech tagging can be referred to as **stochastic tagger**.

➢ **Word Frequency Approach**

1. In this approach, the stochastic taggers **disambiguate the words based on the probability that a word occurs with a particular tag**.
2. We can also say that the **tag encountered most frequently with the word in the training set is the one assigned to an ambiguous instance of that word**.

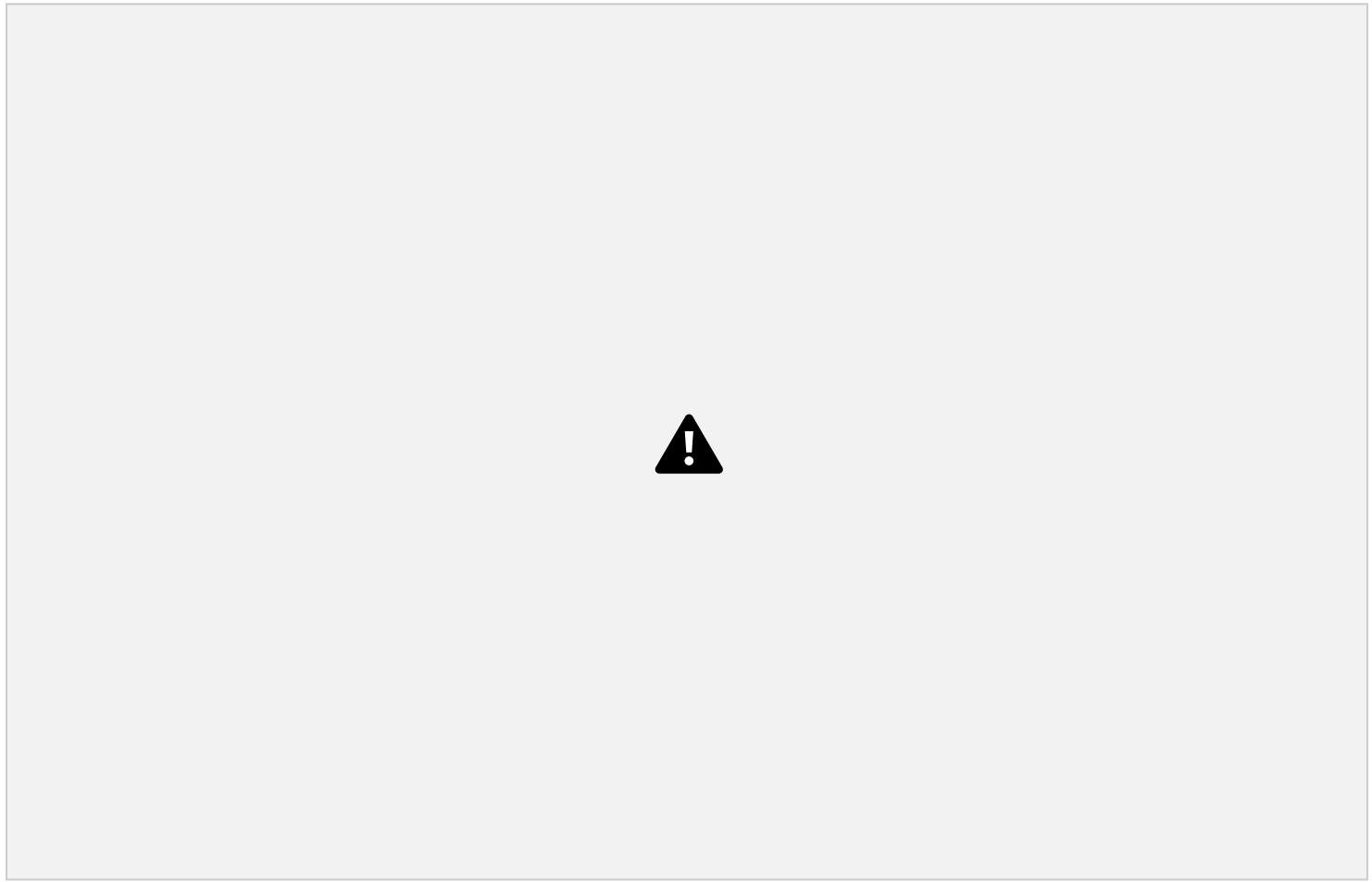
Stochastic POS Tagging

Stochastic POS Tagging

➢ **Tag Sequence Probabilities**

1. It is another approach of stochastic tagging, where the **tagger calculates the probability of a given sequence of tags occurring**.

2. It is also called **n-gram approach**. It is called so because the best tag for a given word is determined by the probability at which it occurs with the n previous



tags.

Stochastic POS Tagging

➢ **Properties of Stochastic POST Tagging**

1. This POS tagging is based on the **probability of tag occurring.**

2. It requires **training corpus.**

3. There would be no probability for the words that do not exist in the corpus. 4. It uses **different testing corpus** (other than training corpus). 5. It is the simplest POS tagging because **it chooses the most frequent tags associated with a word** in the training corpus.

Transformation based tagging

• Transformation based tagging is also called **Brill tagging**.

• It is an instance **of transformation-based learning (TBL),** which is a **rule based algorithm** for automatic tagging of POS to the given text. • TBL, allows us to have linguistic knowledge in a readable form, transforms one state to another state by using transformation rules.

• It draws **inspiration from both the previously explained taggers – rule based and stochastic**.

• If we see a similarity between rule-based and transformation tagger, then like rule based, it is also based on the rules that specify what tags need to be assigned to what words.

Transformation based tagging

- **Advantages of Transformation-based Learning (TBL)**

1. We learn a small set of simple rules and these rules are enough for tagging.
2. Development as well as debugging is very easy in TBL because the learned rules are easy to understand.
3. Complexity in tagging is reduced because in TBL there is the interlacing of machine-learned and human-generated rules.

Transformation based tagging

Top down Parsing

➢ Top down parsing starts with the **starting symbol** and **proceeds towards the goal**.

➢ We can say it is the process of construction the parse tree **starting at the root** and **proceeds towards the leaves.**

➢ It is a strategy of analyzing unknown data relationships by **hypothesizing general parse tree structures** and then considering whether the known fundamental structures are compatible with the **hypothesis**.

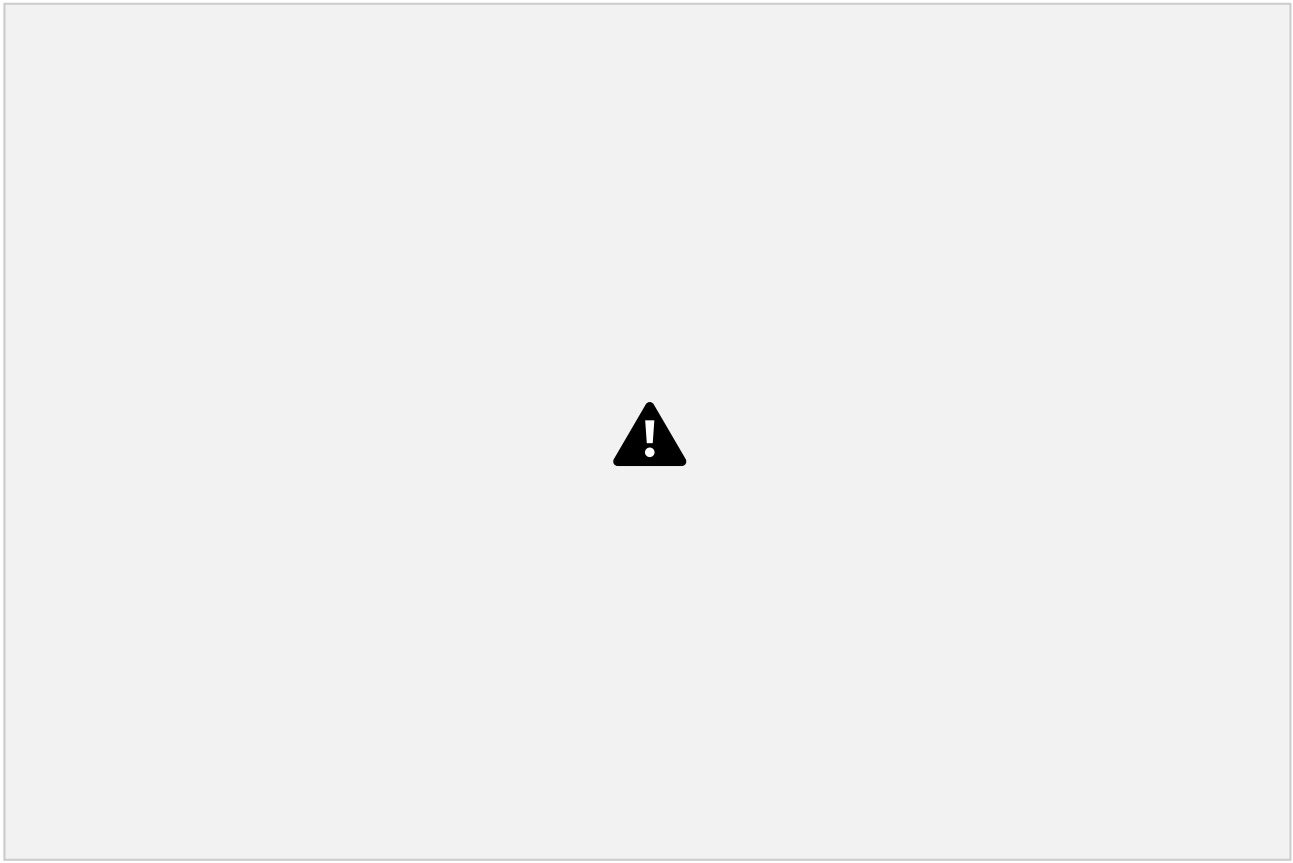
➢ In top down parsing words of the **sentence are replaced by their categories** like verb phrase (VP), Noun phrase (NP), Preposition phrase (PP), Pronoun (PRO) etc. Let us consider some examples to illustrate top down parsing.

Top down Parsing

- We will consider both the **symbolical representation** and the **graphical representation**.
- For parsing we will consider the previous symbols like PP, NP, VP, ART, N, V and so on. Examples of top down parsing are **LL (Left-to-right, left most derivation), recursive descent parser etc.**

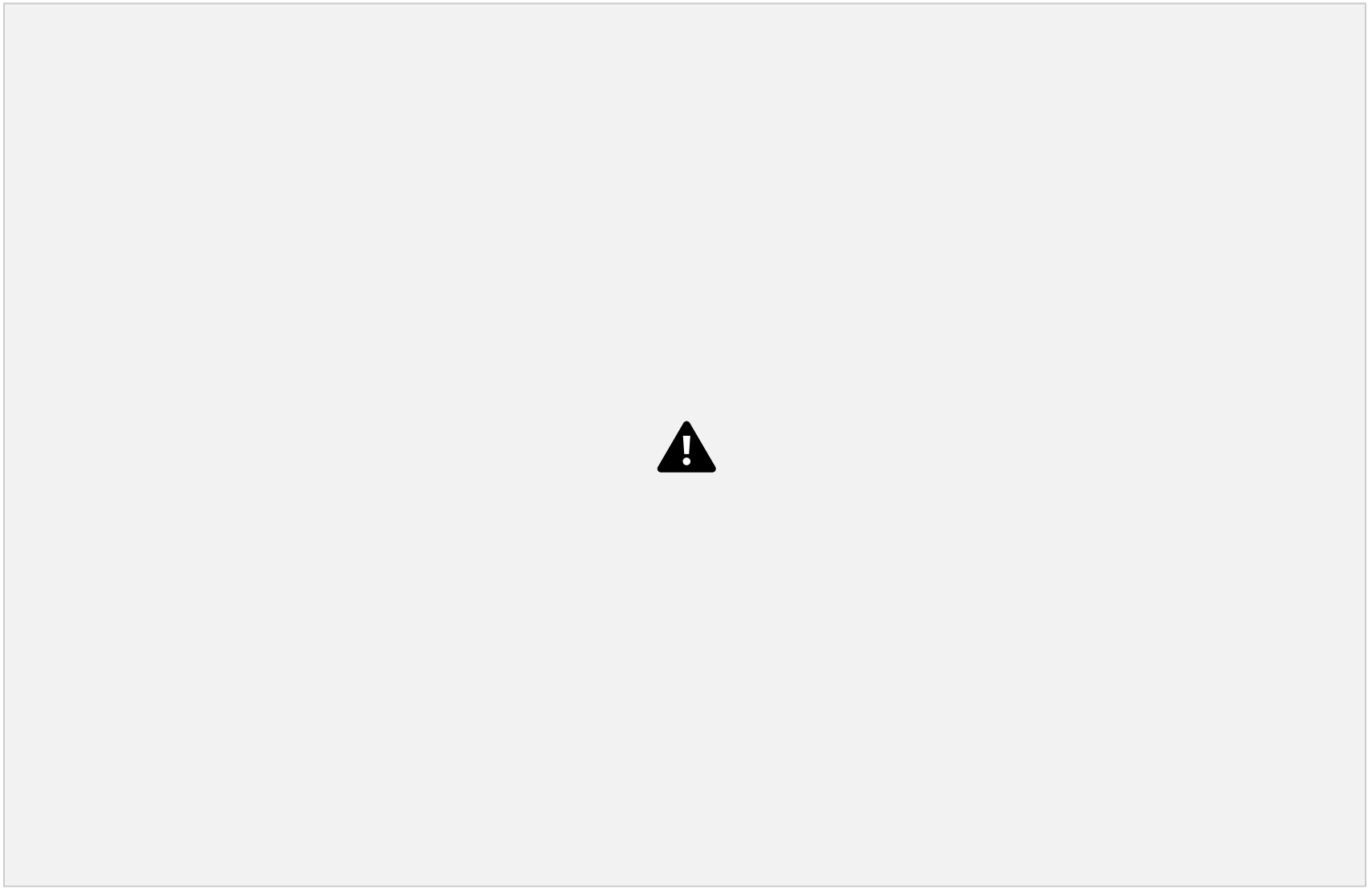
Top down Parsing

Rahul is eating an apple.



Top down Parsing

The small tree shades the new house by the stream.

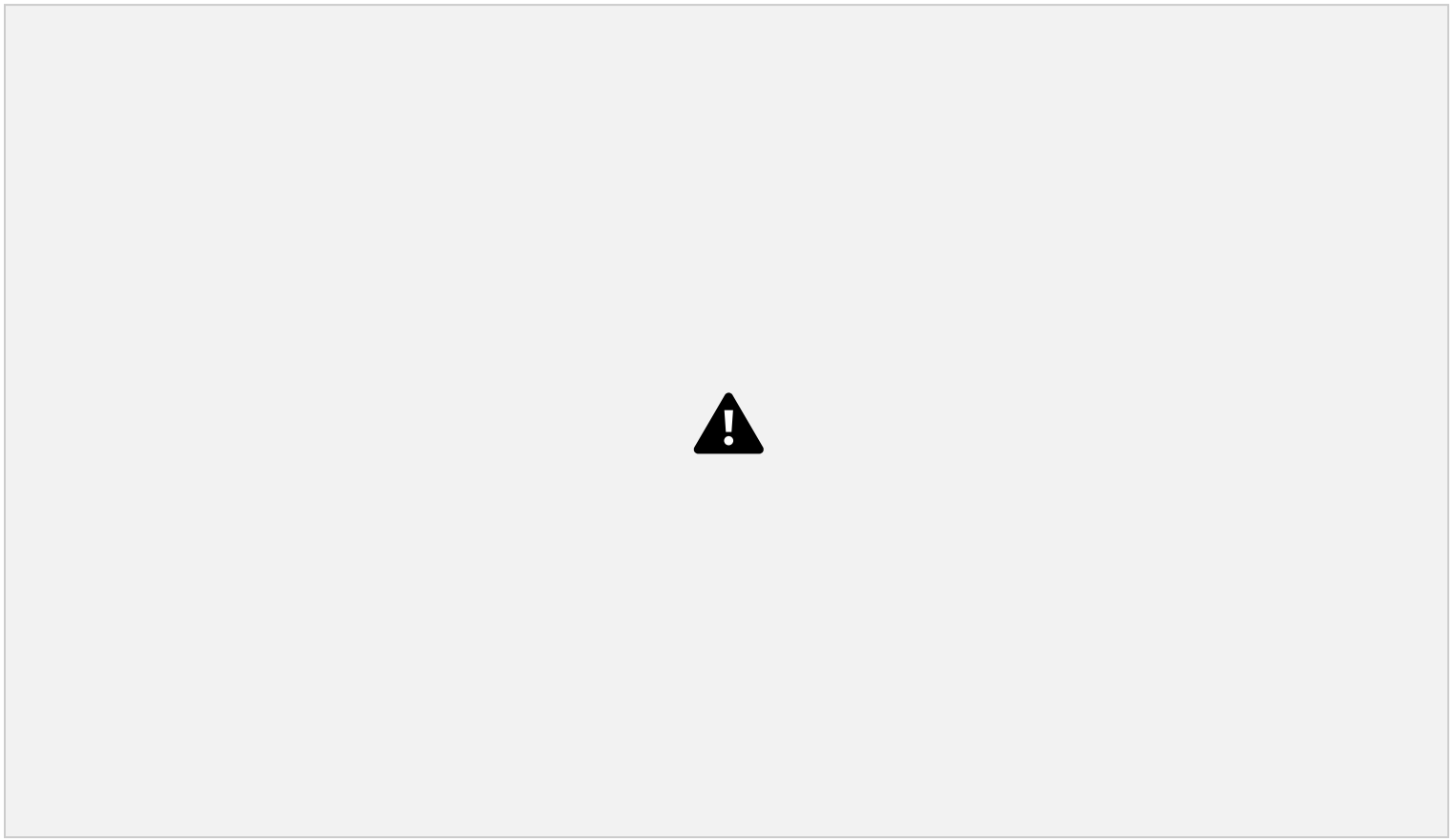


Bottom up Parsing

- In this parsing technique the process **begins with the sentence** and **the words of the sentence is replaced by their relevant symbols**.
- It is also called **shift reducing parsing**.
- In bottom up parsing the construction of parse tree **starts at the leaves** and **proceeds towards the root**.
- Bottom up parsing is a strategy for analyzing unknown data relationships that attempts to identify the most fundamental units first and then to infer higher order structures for them.
- This process occurs in the analysis of both **natural languages** and **computer languages**.

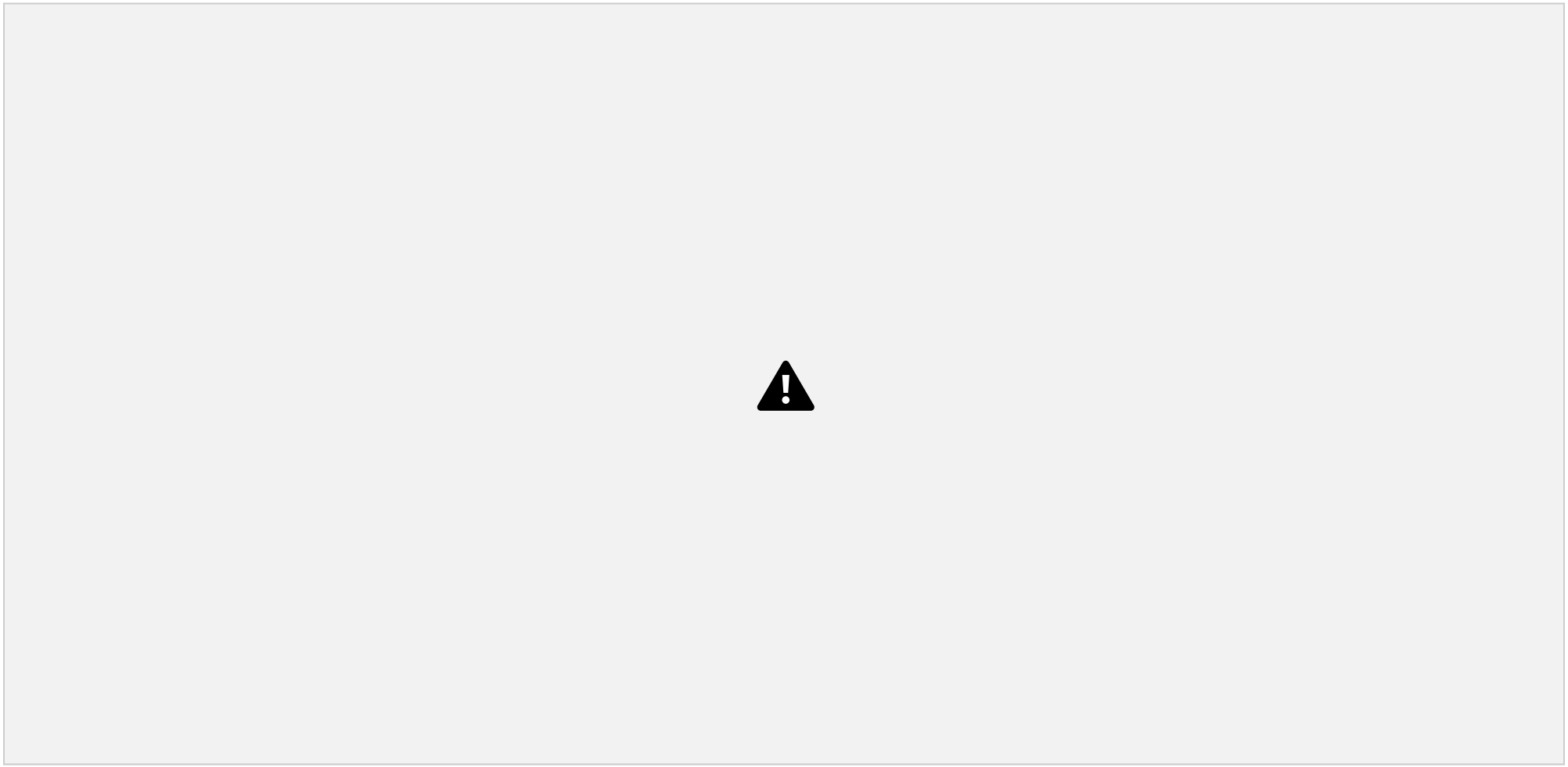
Bottom up Parsing

Rahul is eating an apple.



Bottom up Parsing

The small tree shades the new house by the stream



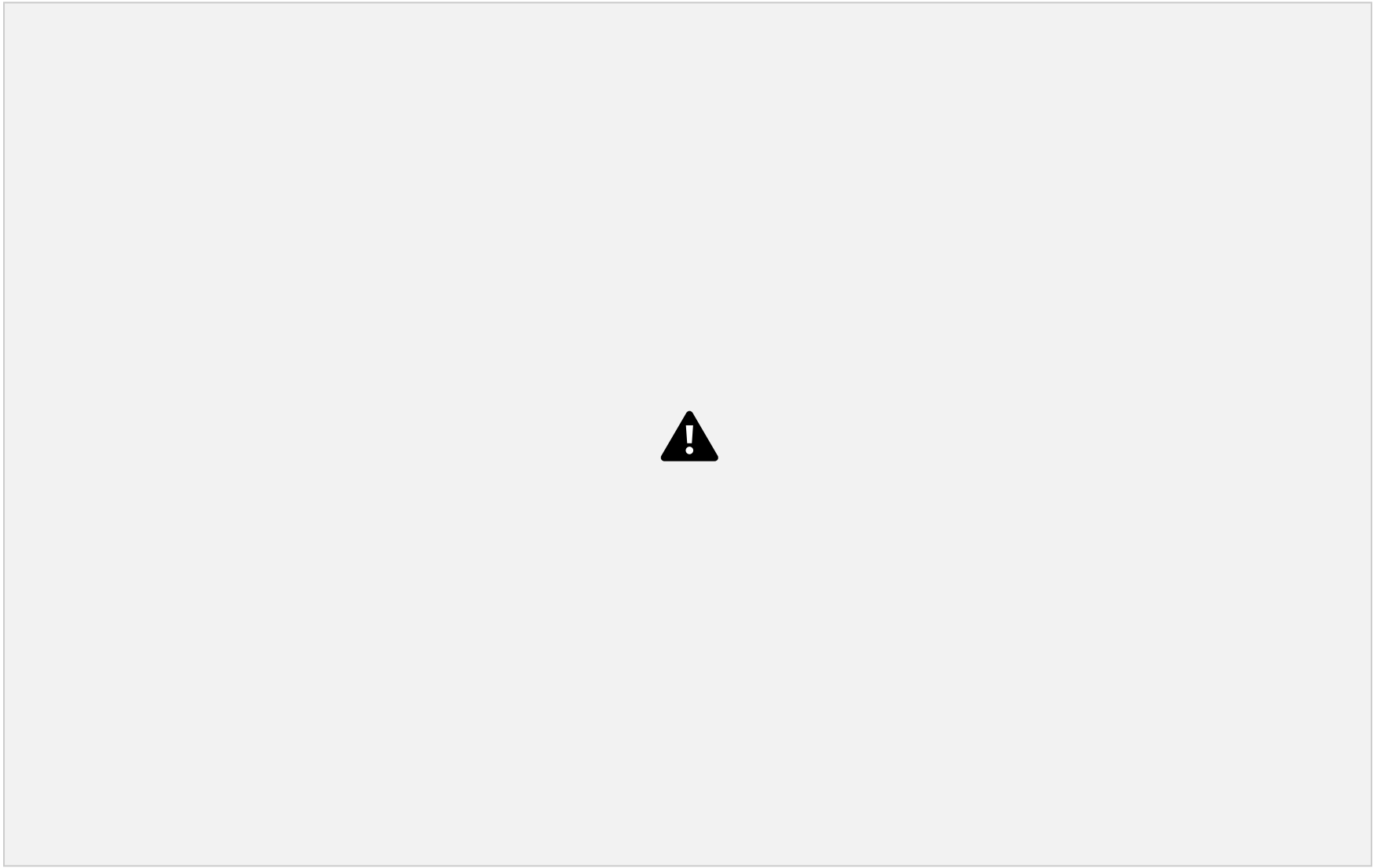
Dynamic Programming Parsing Methods

➢The idea is straightforward: when addressing a large complex problem that can be broken down into multiple sub problems, **save partial solutions as they are generated** so that **they can be reused later in the solution process** for the **full problem**.

➢This "**save and reuse of partial solutions**" is sometimes called **memorizing** the sub problem solutions for later reuse

1. CKY Parsing

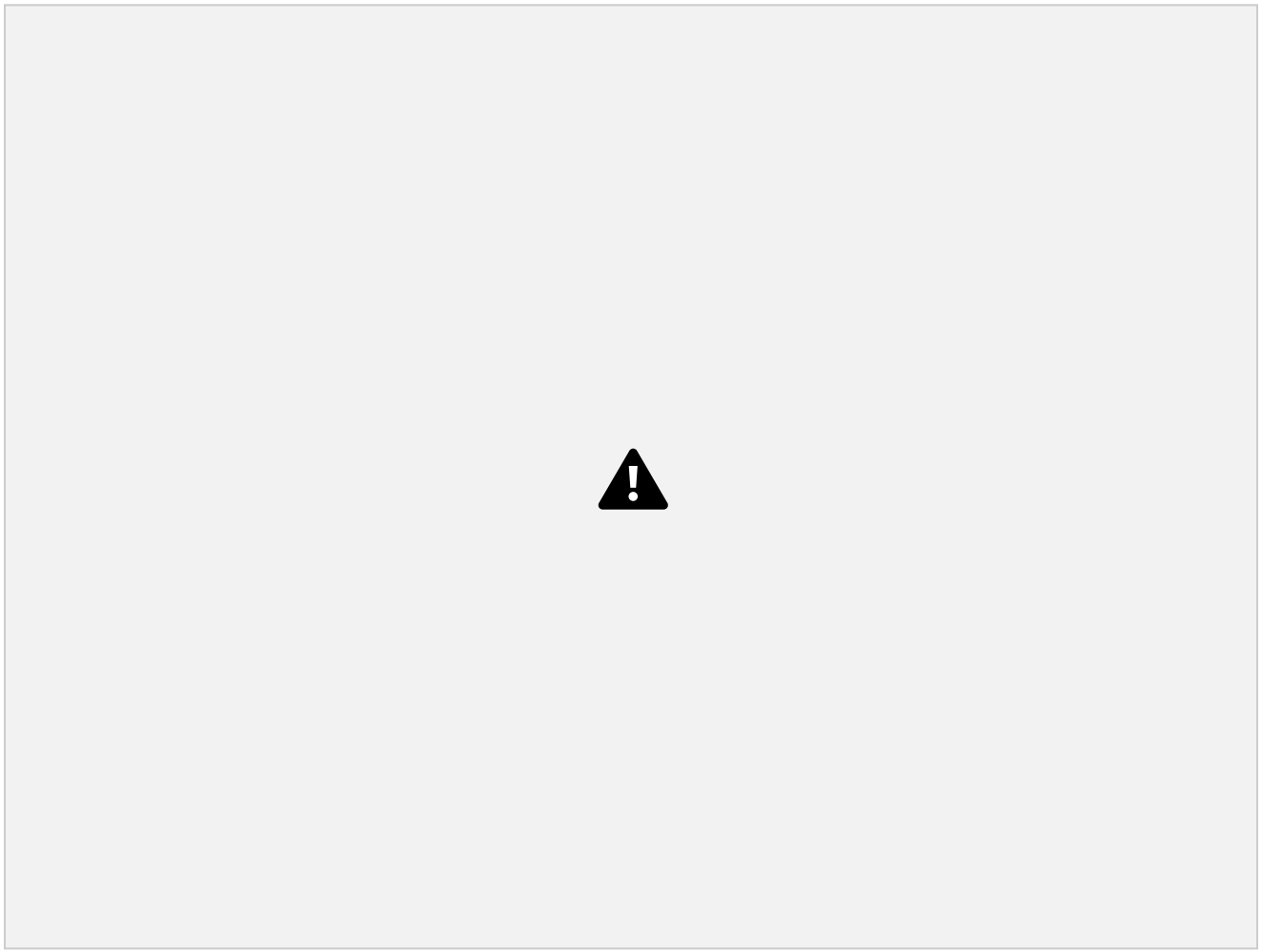
2. The Earley Algorithm

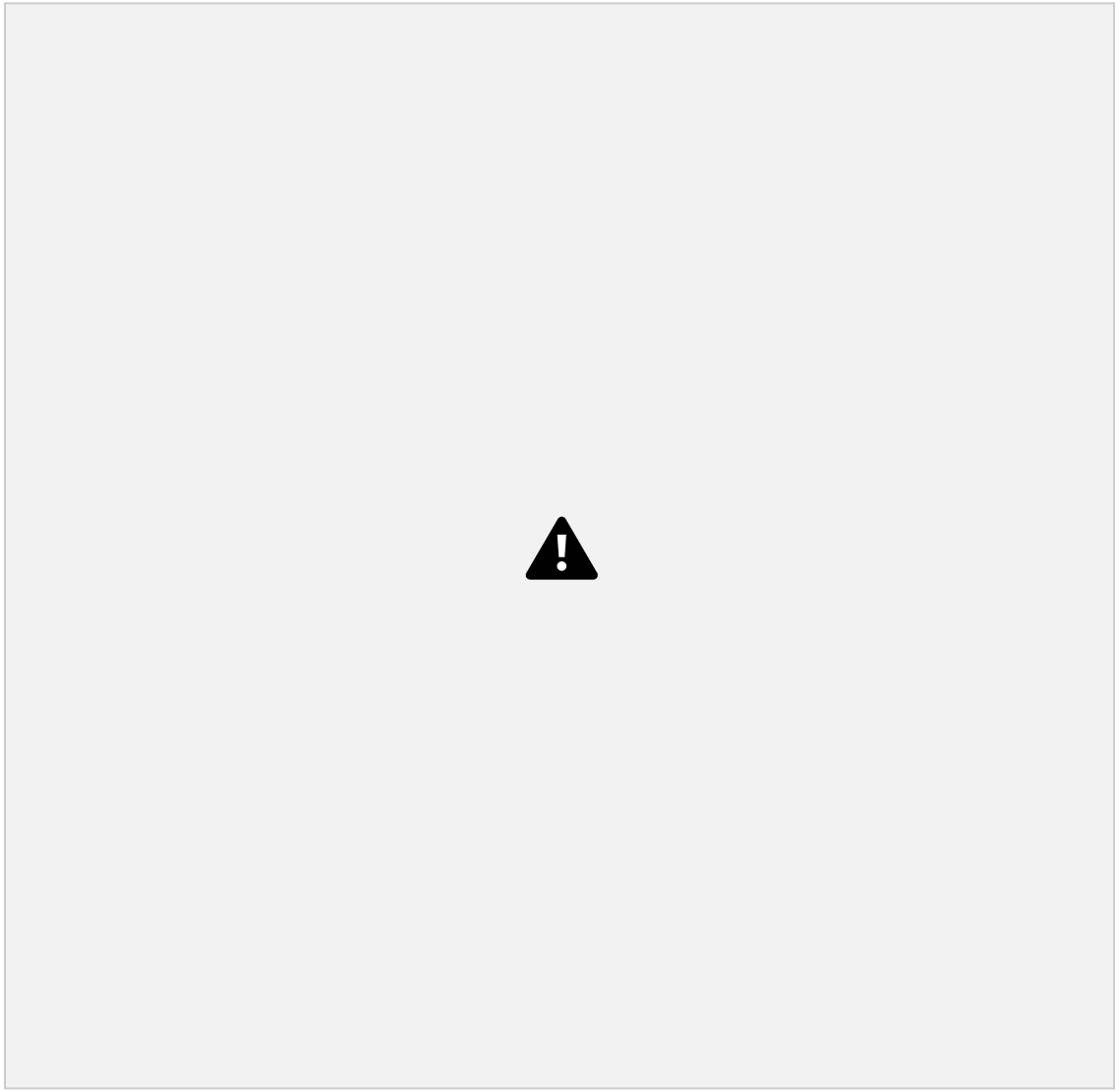


Context-Free Grammars in Chomsky Normal Form

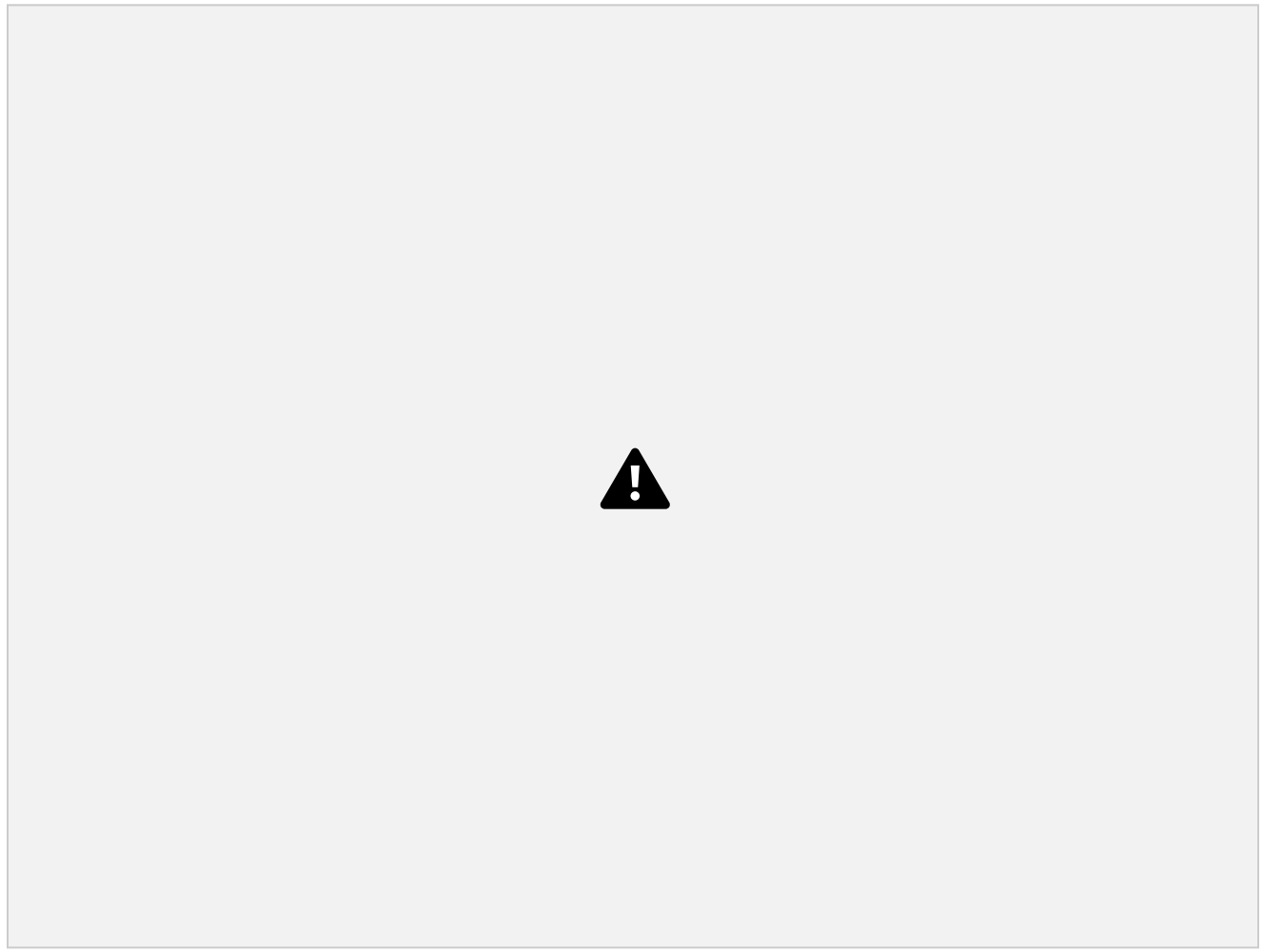
“Book this flight through

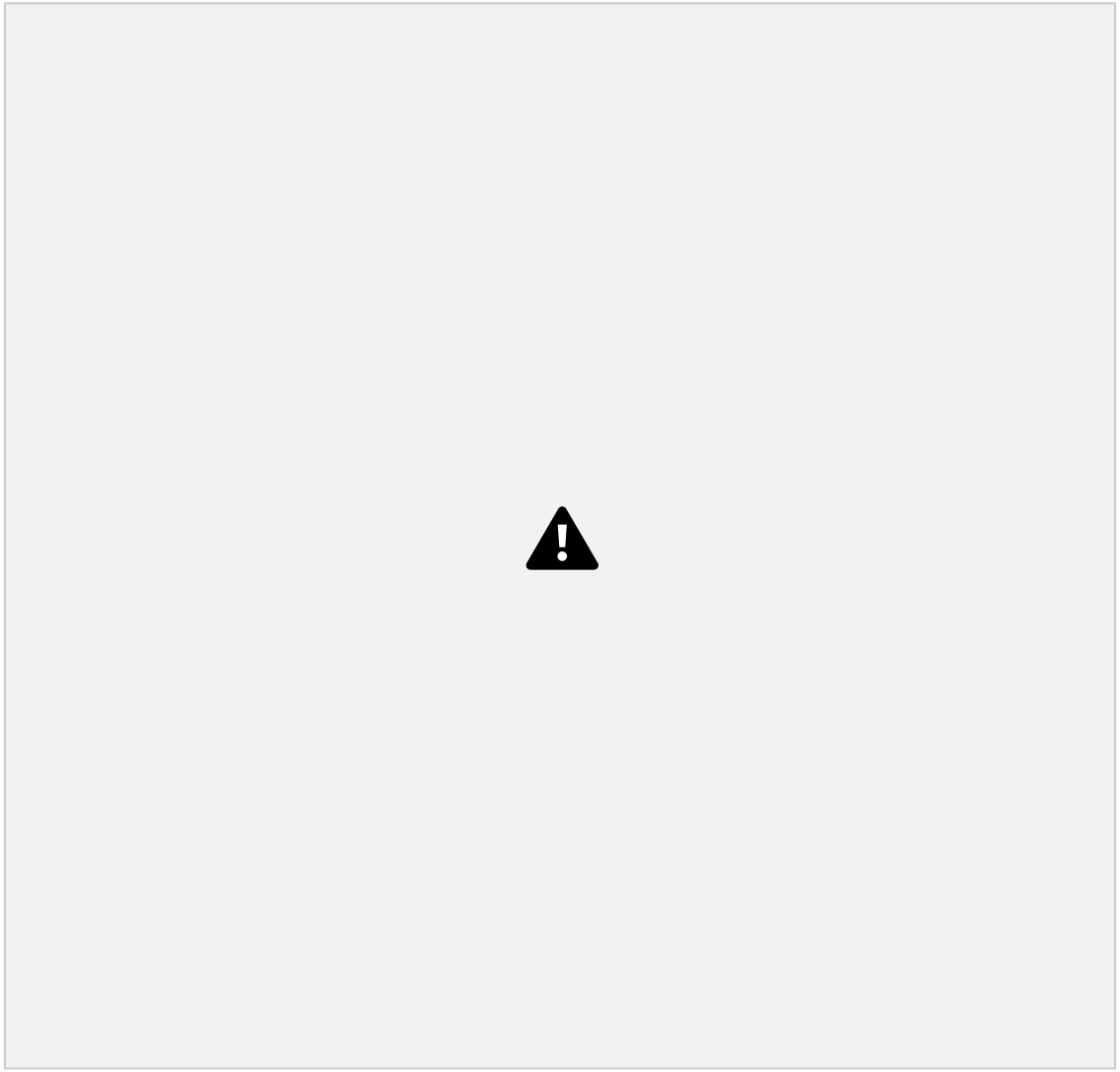
Houston”



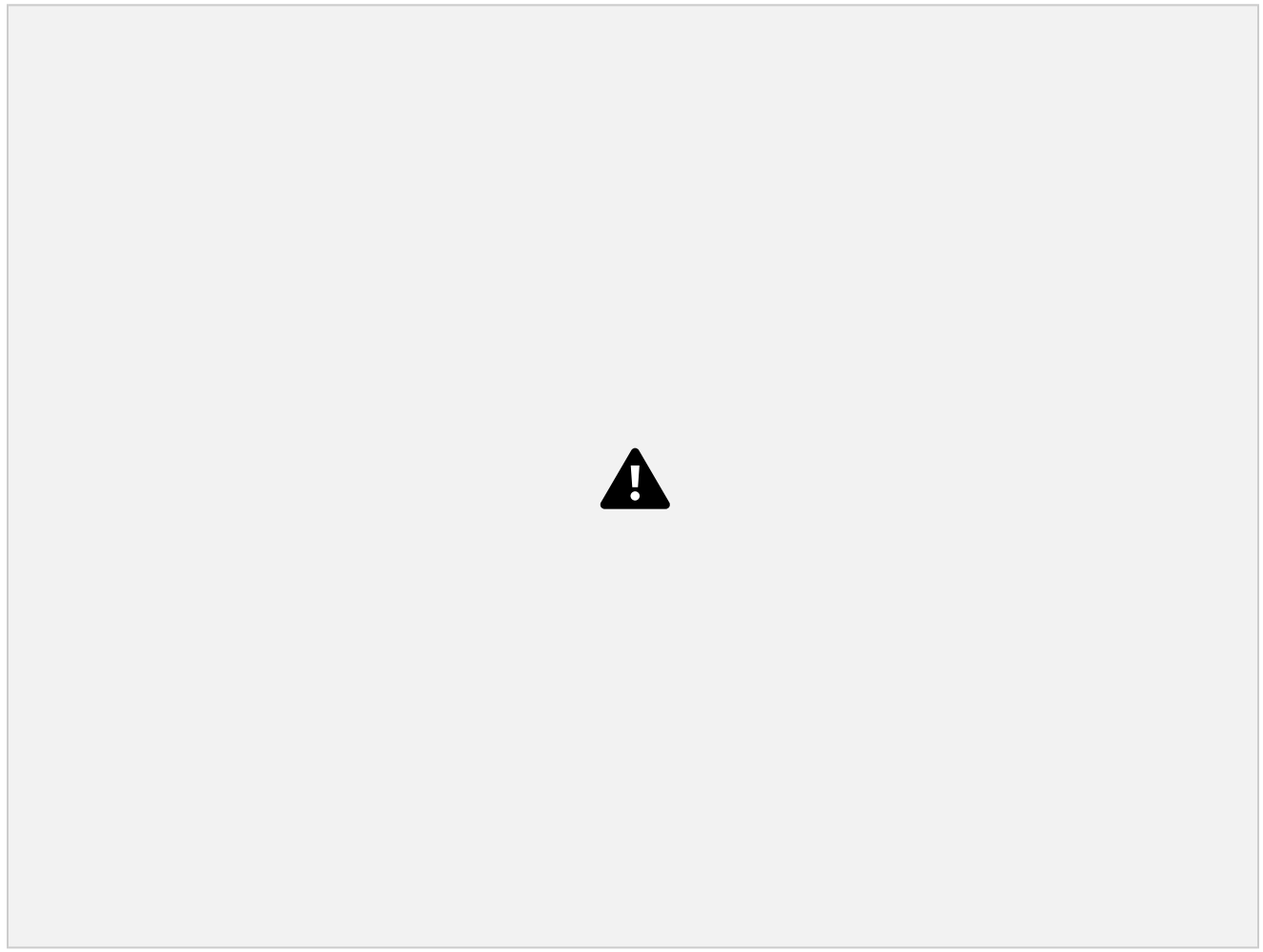


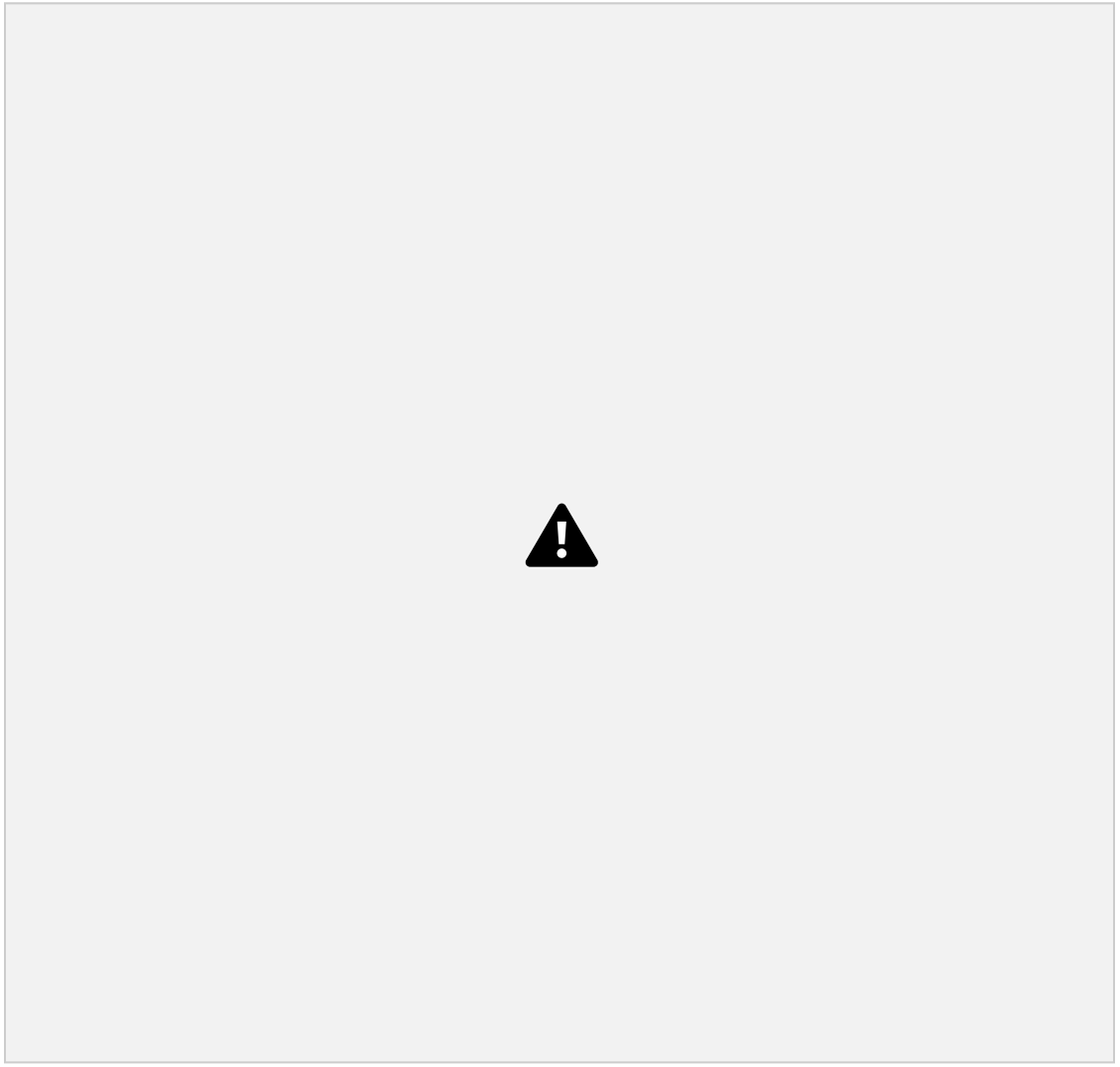
“Book this flight through Houston”



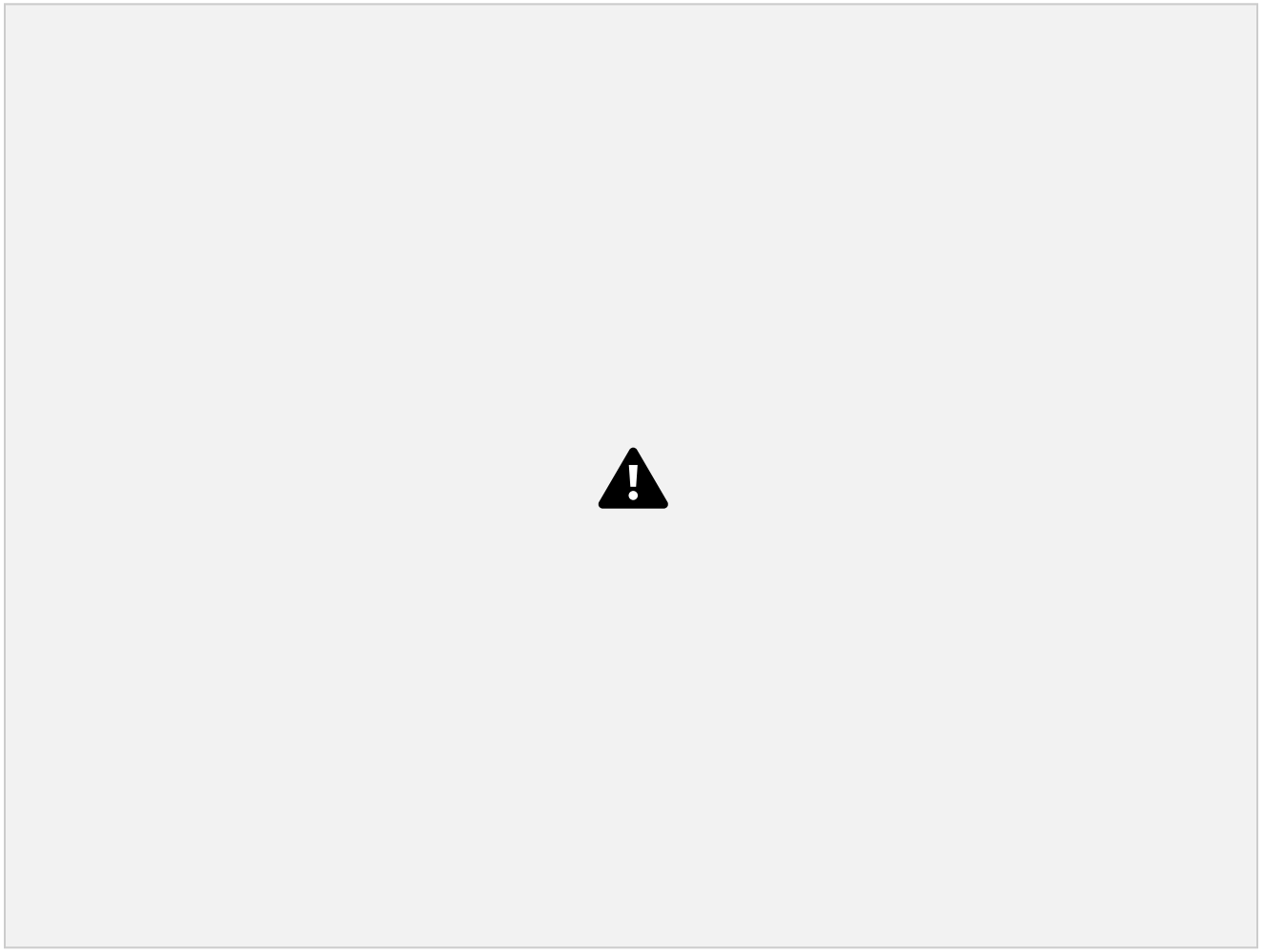


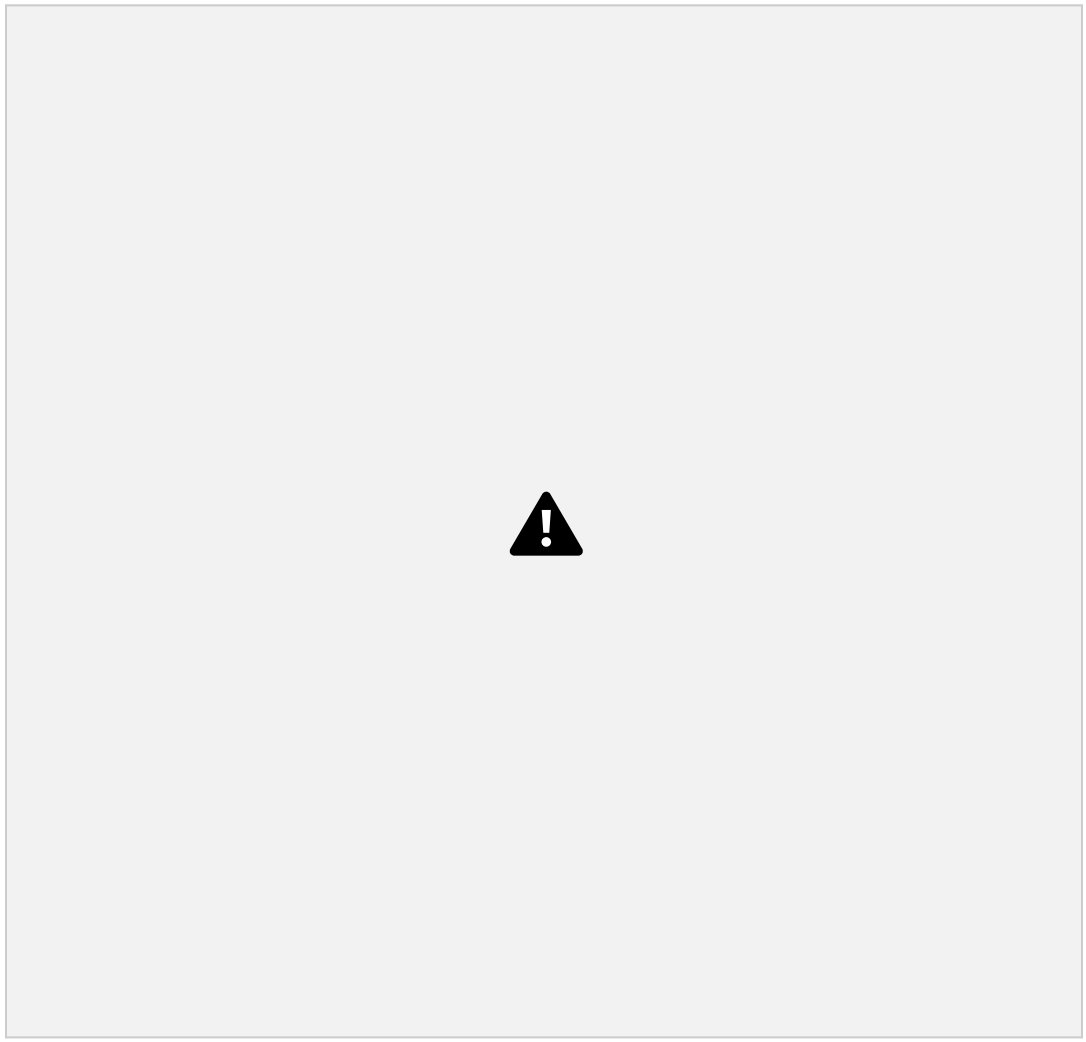
“Book this flight through Houston”



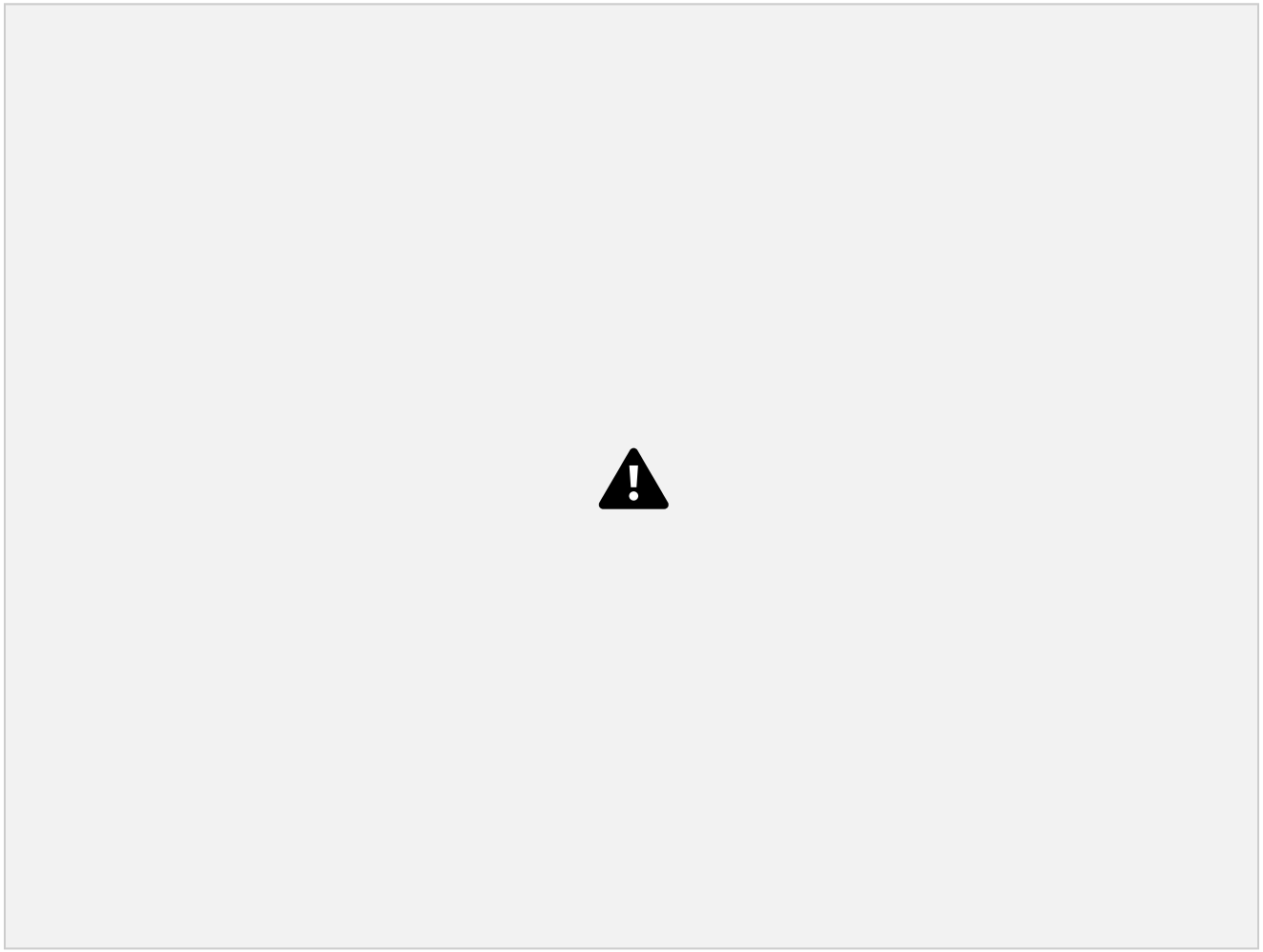


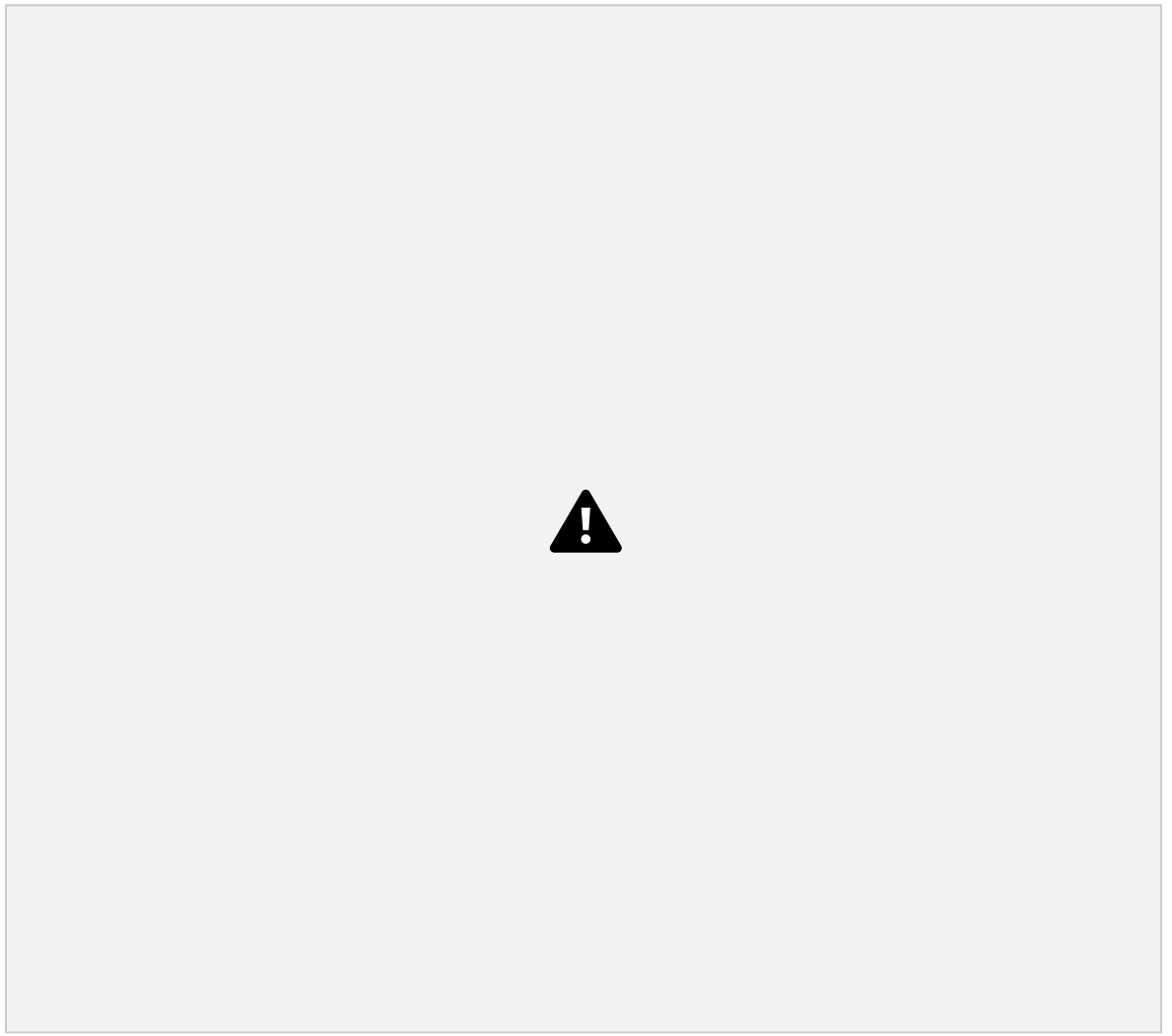
“Book this flight through Houston”



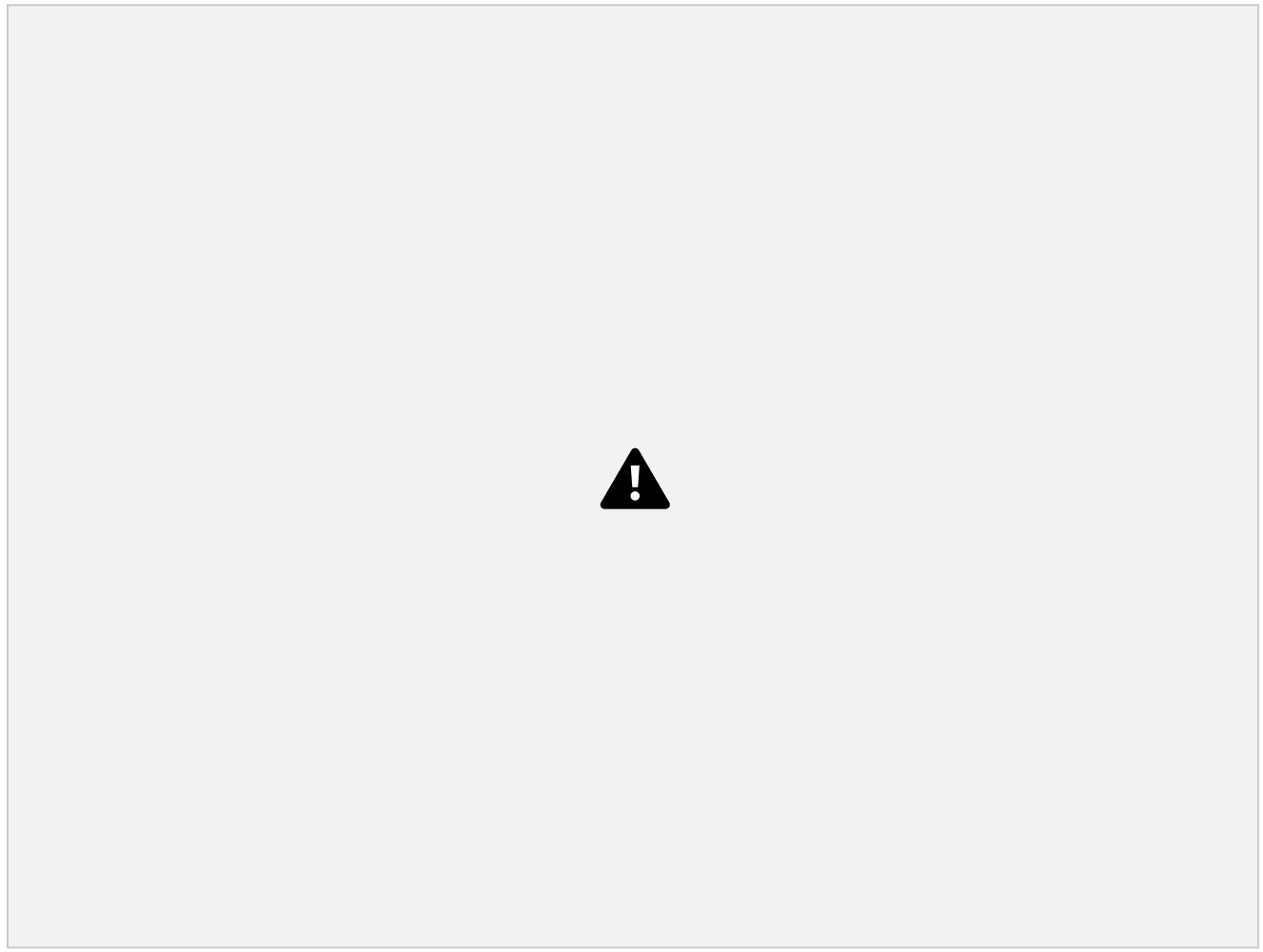


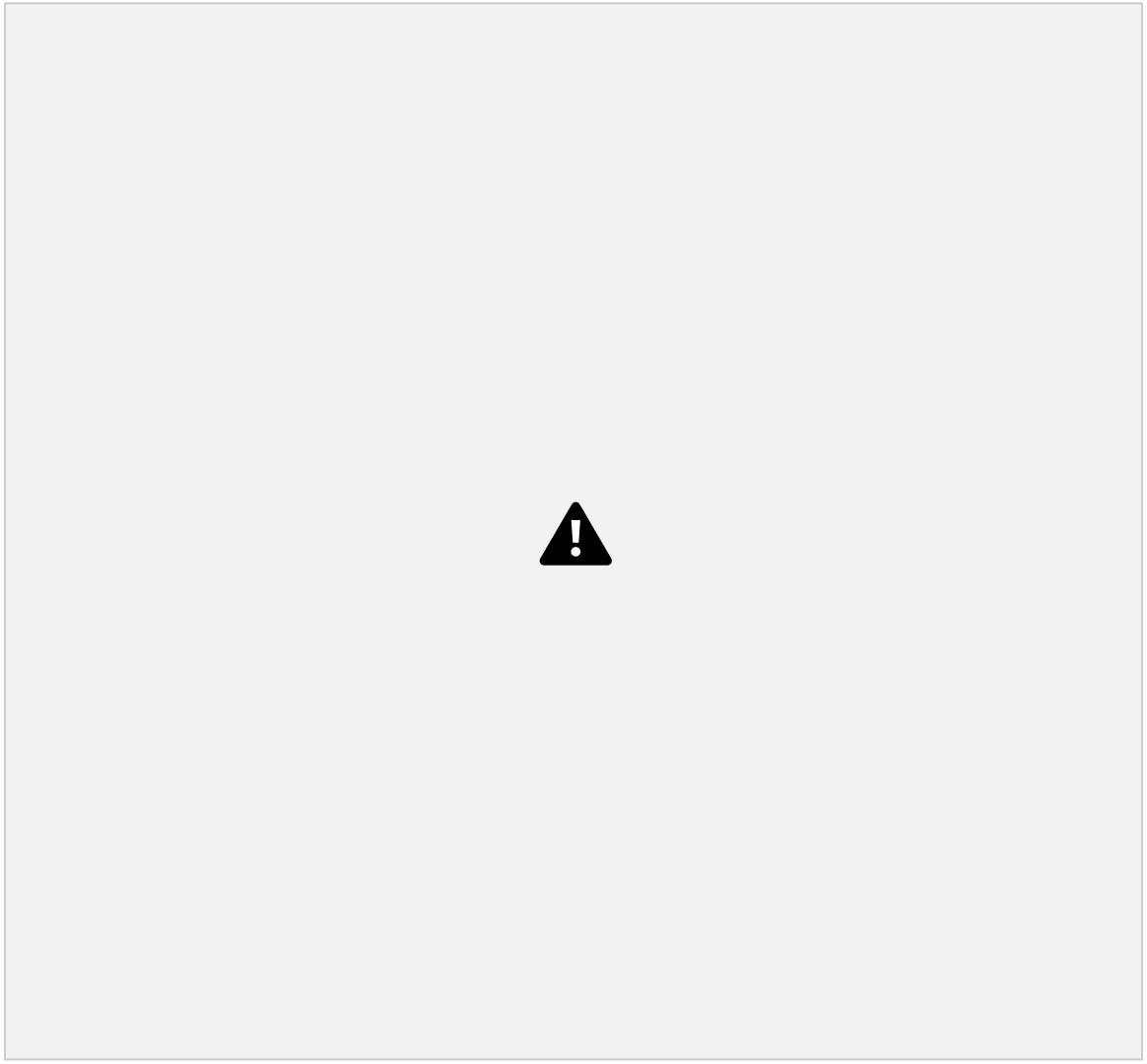
“Book this flight through Houston”



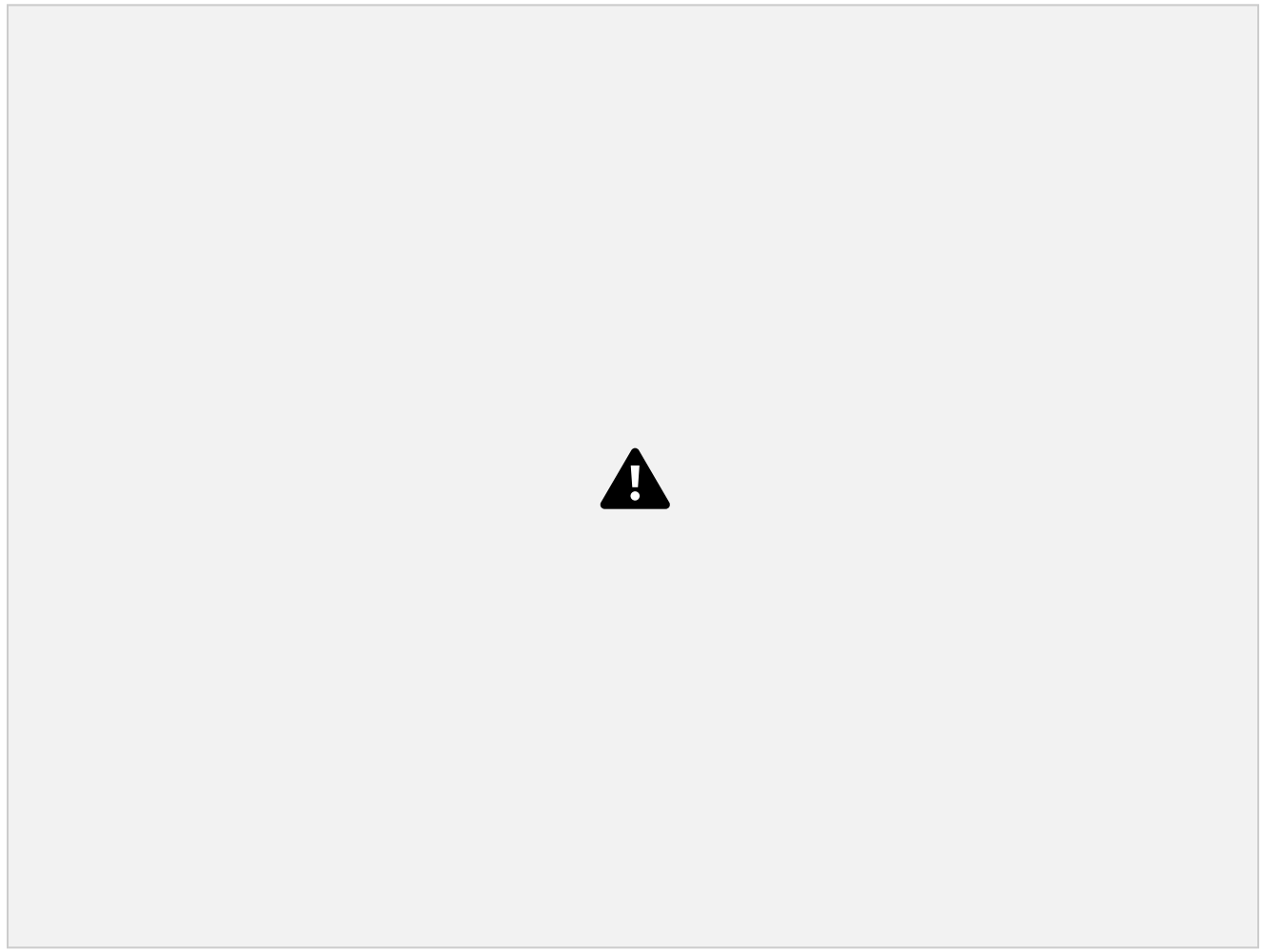


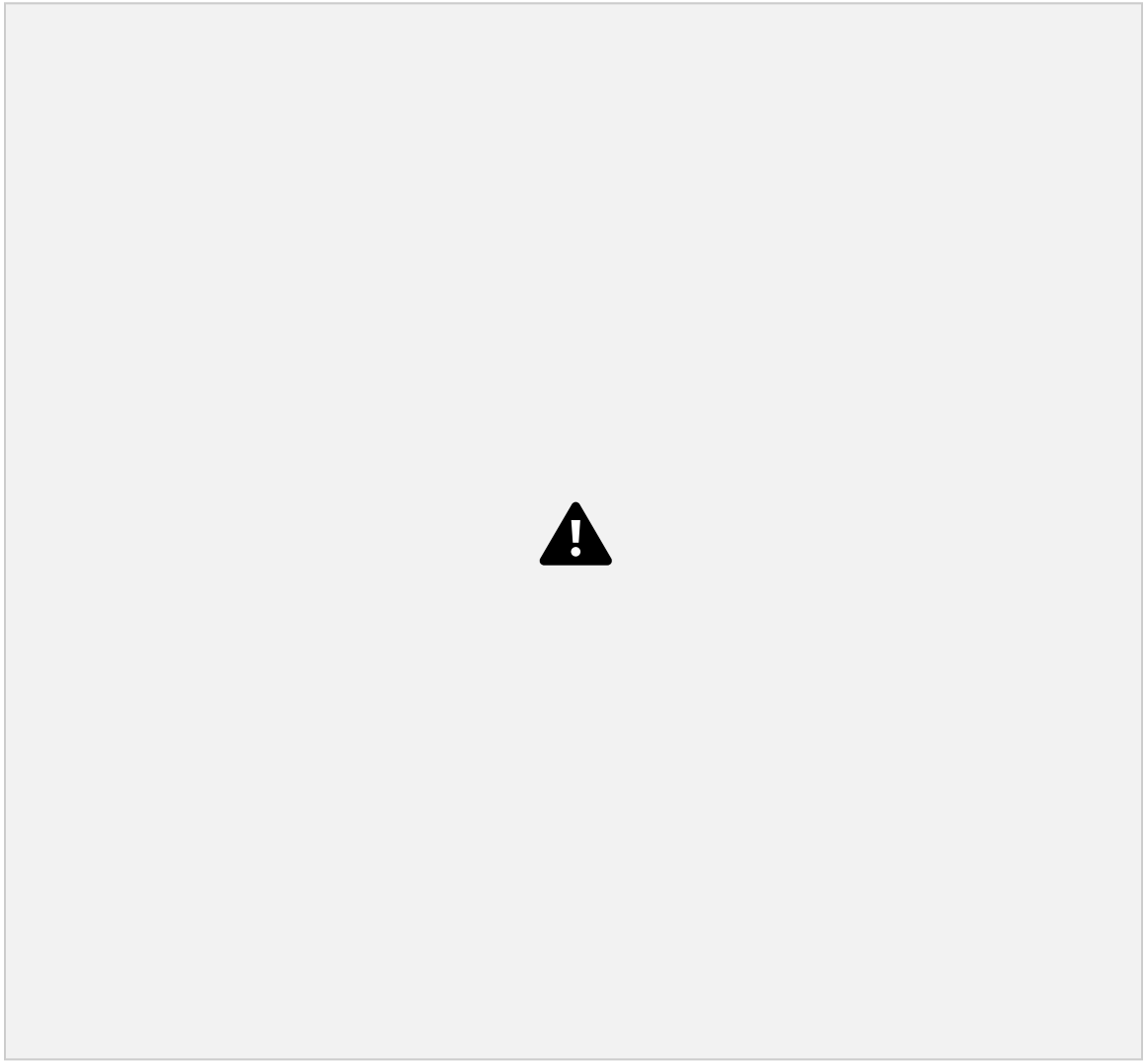
“Book this flight through Houston”





“Book this flight through Houston”





Earley's algorithm

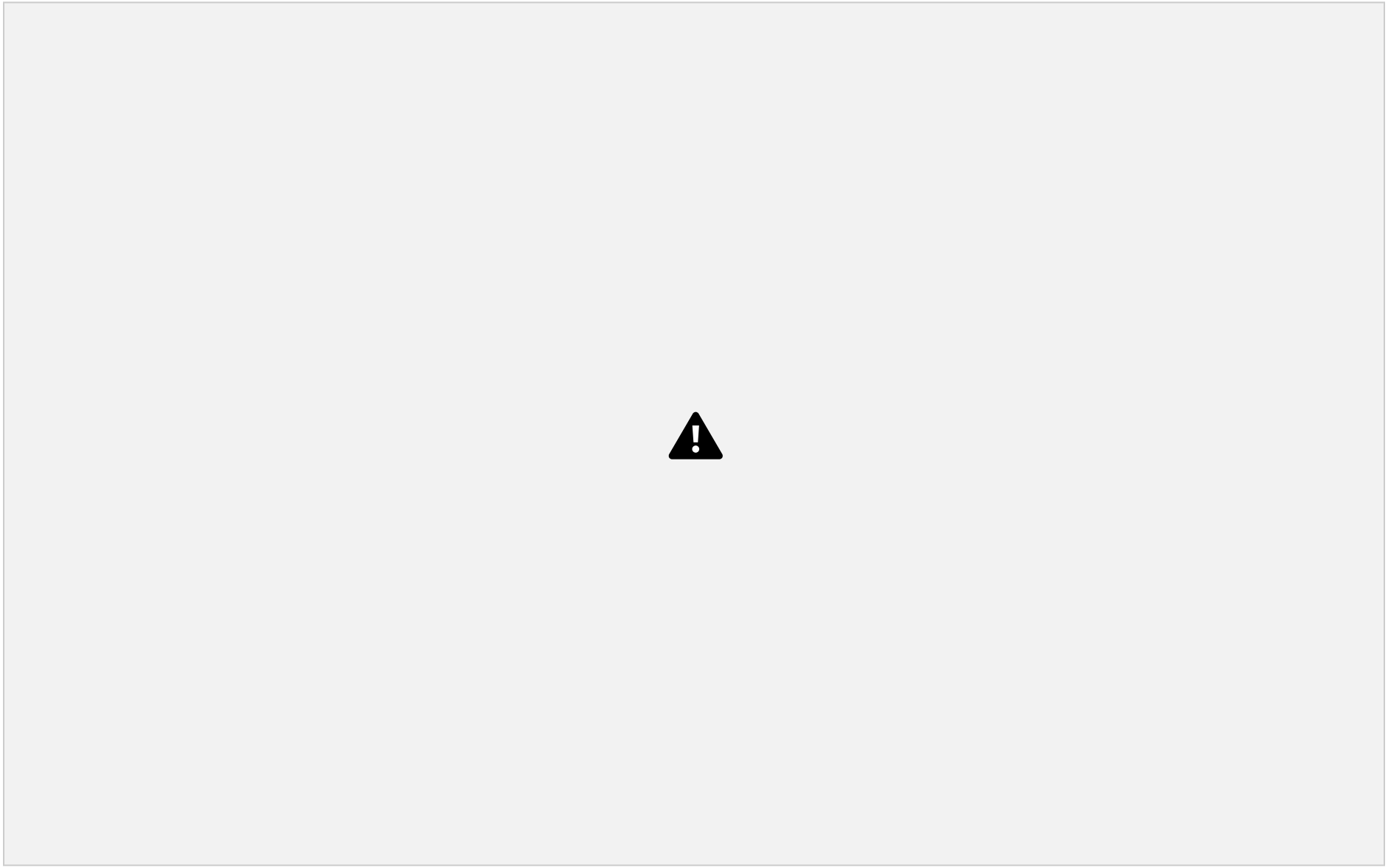
➢ In parsing with **Earley's algorithm** the memorization of partial solutions (partial parses) is done with a **data structure called a chart**.

➢ This is why the various alternative forms of the Earley approach to parsing are sometimes called **chart parsing**.

➢ The chart is generated through the use of **dotted grammar rules.**

➢ Every dotted rule falls into one of three categories, depending on whether the dot's position is at the **beginning**, somewhere in the **middle**, or at the end of the **right hand side**, RHS, of the grammar rule.

Earley's algorithm



Earley's algorithm