### 6.3 Hints

A better decision tree for articles:

![Decision Tree](image)

#### 6.3.1 When a noun can be used as a countable or an uncountable concept

The use of articles depends on the concept which is meant in the **current context**. For example, word *memory* can have at least three meanings:

1. The store of things learnt or the power or process of recalling (in our brains) → generally uncountable. "Memory can be divided into two classes: short-term memory and long-term memory. The short-term memory...” However, you can say: ”I have a good memory”.

2. The object of recall → countable. ”My earliest memories”

3. The capacity of a computer to store information → uncountable. In the cs context, you can suppose it as a known concept and use article the (always?). ”The data is loaded into the main memory”

*Time* is another word which can be used in different ways. It can mean a limited period or interval, an indefinite period or duration, or it can express an occasion of repeated actions. In addition, it occurs in several phrases. By default, time is uncountable (either no article or article "the").
1. Without any article:
   "Time will show..."
   "It is time to do sg."
   "It takes time..."
   "on time" (or "in time")

2. Article "the"
   "all the time"
   "at the same time"

3. Article "a":
   "It is a long time..."
   "one at a time" (i.e. one by one)

4. Plural:
   "many times"
   "modern times"

6.3.2  **Hint: could you use "any" or "some"?**

Hint: Try if you could use words any or some before the noun. If you can, it is indefinite. This means that you cannot use article "the".
"The grammar is not strict in (any) spoken language"
"The disk contains (some) space for back-up files"
"There is some reason for this behaviour" → "There is a reason for this behaviour".

6.3.3  **Hint: are you referring to sg particular?**

If you have need to say "This particular x", say "The x", where x is a noun. "This particular" hints that you have already talked about x and it is known (definite).
Don’t use pronouns, if you mean article "the"! "This x" can often be replaced by "the x" (where x is a noun).

6.3.4  **Hint: could you use ∃ or ∀?**

Imagine the concept C as a set (universum) of all its instances. E.g. concept "computer" is a set of all possible computers.
If you want to express ∃x ∈ C such that P(x) (there is some x in C for which holds property P), use article a/an. "A computer could solve this problem faster." (maybe not all of them, but some computers can)
If you want to express $\forall x \in CP(x)$ (for all $x$ in $C$ property $P$ holds; i.e. it holds for the whole set $C$), use article the. Now you refer to the whole class of $xs$ in $C$, which is definite. "The computer can solve only mechanical problems." (all computers can do this) Notice that this technique suits only for countable concepts!

6.3.5 Articles before variable names?

In cs, we often use the names of variables, data sets, models, etc.

- When you use the name without any modifying word → no article "$X$ is independent from $Y$", "$S$ contains no outliers"

- When you use a modifying word like "set", vector", ”model” etc. before the name → by default, no article

- Problem: in literature, article "the" is sometimes used in such cases??
6.3.6 Exercises

Task 1: Add the correct articles to the following sentences or mark the absence of articles by −!

1. ______true positive rate was higher in ______method X than ______method Y.

2. ______method X had ______higher true positive rate than ______method Y.

3. ______memory means ______power or ______process of recalling.

4. ______X is ______algorithm which solves ______Travelling Salesman problem. ______algorithm X is ______fastest among all ______known TSP algorithms.

5. ______data set X follows ______Normal distribution with ______parameters μ and σ². ______parameter μ is ______mean of ______set X and ______parameter σ² is ______variance of ______X.

6. ______problem X belongs to ______class P, if it has ______polynomial time algorithm Y. ______time complexity of ______algorithm Y is O(p(n)) where n is ______size of input and p is ______polynomial function.

7. In ______next section we introduce ______theory of ______Bloom filters.

8. To assess ______students’ program codes, we construct ______bug library. ______bug library contains all ______errors which have occurred in ______students’ programs.

9. ______infinite time Turing machines extend ______idea of ______traditional Turing machines.

10. In ______pattern extraction we produce ______set of ______new attributes from ______original ones. ______goal is to find such ______set of attributes which describes ______data ______best. ______goodness of representation depends on ______modelling purpose, and in ______practice we have to define ______appropriate goodness measure.

11. In ______clustering analysis we divide ______data points into ______clusters such that all ______data points in one cluster are similar to each other but different from ______data points in ______other clusters.
12. _____episode is _____set of _____events which occur together. If _____order of _____events is fixed, _____episode is called serial.

13. There is always _____danger that _____model overfits. _____danger that _____model overfits is unavoidable.

14. _____main parts of _____computer are _____central unit, _____hard disk, _____i/o devices. _____central unit is responsible for all _____computation.

**Task 2**: Are the following words countable or uncountable? Which articles can you use with them? Give example sentences!

- space
- requirement
- model
- program
- computation
- power
- capacity
- data
- information
- knowledge
- recognition
- software
- hardware
- code
- value
- property
- strength
- weakness
- use
- usability
7 Pronouns

Two important rules when you use pronouns:

1. When a pronoun refers to a noun in the preceding sentence, make sure that the referred is obvious!
2. Each pronoun should agree with the referant in number and gender.

7.1 Unclear references

- The simple pronouns – it, they, this, that, these, those – do often create ambiguities.
- Goal: the reader should not have to scan the previous sentence to understand what you mean.
- Recommendation: Avoid them, when possible! If you use them, always check twice that the meaning is not ambiguous!
- Never use ”those” – it is usually a sign that the sentence is foggy.
  ”There was no difference in the accuracy of models between those which belonged to group A and those which belonged to group B. → ”The models in groups A and B were equally accurate.”
- Do not use ”it” to begin a sentence, if it is not absolutely clear, what it refers! (Exception: expressions like ”It is difficult to estimate...” require ”it” as a formal subject.)
- Hint: often you can replace ”this/these” + noun by ”the” + noun!
  ”This experiment demonstrated...” → ”The experiment demonstrated...”

7.2 Pronouns which require singular verb form

{everybody, anybody, nobody, everyone, anyone, no one} → verb is singular

7.3 Every vs. all

<table>
<thead>
<tr>
<th>every</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ singular noun</td>
<td>+ singular or plural noun</td>
</tr>
<tr>
<td>when you talk generally</td>
<td>when you mean sg certain</td>
</tr>
</tbody>
</table>
7.4 Many vs. several
several < many
several ≈ some

7.5 Phrases
one – the other (singular)
some – the others (plural)

each other, e.g. "X and Y affect each other"
This kind of + singular noun, e.g. "This kind of system…"
If you want plural you have to say "Systems of this kind…"

on one’s own, e.g. "The students solved the task on their own”.

"All but one point belong to cluster 1"
"First of all, we have to initialize the parameters"
"On the one hand, the system is stable, on the other hand, it has poor accuracy”
"The initialization phase is time demanding. Otherwise the program is very efficient.”

7.6 Relative pronouns

Relative pronouns (who, which, that) are used in relative clauses. To understand their use we have to study also relative clauses.
→ Section Relative clauses.

8 Adjectives

These seem to be well mastered, just two notes:

1. Avoid vague adjectives!

2. How to derive and use comparative and superlative forms?

8.1 Vague adjectives

• Do not use vague adjectives. Especially the adjectives which describe amounts (large, small, huge) are very context-sensitive!
• E.g. for statisticians, a data set of 500 rows is quite large, while for a
data miner it is extremely small → numbers are more exact!

• The expressions become even vaguer, when you add modifiers “quite”,
“rather”, “very”, etc. Skip them always when possible!

8.2 Comparative and superlative

Basic rule: use -er/-est for short adjectives, and more/most for longer ones.

<table>
<thead>
<tr>
<th>Adjective type</th>
<th>Comparative</th>
<th>Superlative</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-syllable adjectives</td>
<td>-er</td>
<td>-est</td>
<td>strong, stronger, the strongest</td>
</tr>
<tr>
<td>2-syllables adjective with suffix</td>
<td>-er</td>
<td>est</td>
<td>narrow, narrower, the narrowest</td>
</tr>
<tr>
<td>2-syllables adjective with suffix</td>
<td>-er</td>
<td>est</td>
<td>noble, nobler, noblest</td>
</tr>
<tr>
<td>consonant + le</td>
<td>-er</td>
<td>est</td>
<td></td>
</tr>
<tr>
<td>all other adjectives</td>
<td>more + adj.</td>
<td>most + adj.</td>
<td>efficient, more efficient, the most efficient</td>
</tr>
<tr>
<td>participes verb+{-ed, -ing}</td>
<td>more</td>
<td>most</td>
<td>interesting, more interesting, the most interesting</td>
</tr>
<tr>
<td>when used as adjectives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>irregular adjectives</td>
<td></td>
<td></td>
<td>good, better, the best bad, worse, the worst</td>
</tr>
</tbody>
</table>

Notice the spelling:

• the consonant is doubled in a short stressed syllable: big, bigger, the biggest

• -y becomes -ie: easy, easier, easiest

8.3 When you compare things

When you use the comparative, make clear what you are referring!

"Problem X is easier to solve" (than what?)

Basic structure:

\[ X \text{ is as efficient as } Y \] (X and Y are equally efficient)

\[ X \text{ is more efficient than } Y \]
Exceptional expressions:

- $X$ is different from $Y$
- $X$ is similar to $Y$
- $X$ is the same as $Y$
- $X$ is inferior/superior to $Y$
- $X$ is equal to $Y$ (Notice: use “$X$ equals $Y$” only in math, for $X = Y$)

9 Other word groups

Verbs, nouns, pronouns, numerals, and adjectives compose the skeleton of sentences. The additional stuff consists of

- adverbs,
- prepositions, and
- conjunctions.

Adverbs modify verbs, adjectives, or other adverbs, while conjunctions join words, clauses or sentences together. Some words can be used either as adverbs or as conjunctions. Prepositions are always connected to other words (nouns, pronouns, or verbs in -ing form). Prepositional phrases (“in the beginning”, “through a gateway”) are used in the same way as adverbs.

10 Adverbs


- time (immediately, now, soon, later, next)
- place (here, there, everywhere)
- manner (easily, temporarily, well, poorly)
- degree (very, quite, ...) → Avoid in scientific texts!
- frequency (often, seldom, usually, sometimes)
- speaker’s attitude “Fortunately, the data set is small, and function $f$ can be computed in real time.” → use sparsely!
Notes:

- **Recommendation**: Use expressive verbs and nouns which express the most of message, and as few adverbs/prepositional phrases as possible!

- Use introductory adverbs like "fortunately, similarly, conversely, certainly" carefully, as a synonym to expressions "it is fortunate" or "in a similar manner". Drop them if they are not needed.

- Notice that "importantly" and "interestingly" are not proper adverbs. E.g.
  "More importantly, the accuracy can actually increase when the complexity is reduced"
  → "More important, the accuracy can actually increase when the complexity is reduced."

  "Interestingly, we found that..."
  → "An interesting finding was that..."

10.1 **The position of adverbs in a sentence**

The adverb can be

1. in the beginning, when you express time or attitude. E.g. "Evidently, the students’ learning outcomes depend on their effort", "Later, we realized that..."

2. in the end, when you express way, time or place. E.g. “This problem occurs frequently in sparse data.”

3. in the middle, when you express frequency or attitude. Notice that *already* behaves in the same way. E.g. "In knowledge discovery, we assume that the features have been already extracted"

An adverb should clearly refer to the word it modifies!

10.2 **Special cases**

**still** and **yet**

- Still (mostly in positive sentences): before the main verb, but after be-verb. "These enlargements are still unimplemented"
• Yet (mostly in negative or interrogative sentences): in the end. ”These enlargements have not been implemented yet.

• If still or yet is used in the beginning, it means ”however”.

so and such

• So: before adjectives or adverbs which are not succeeded by nouns. E.g. ”The time complexity is not so hard”

• Such: when an adjective is succeeded by a noun. E.g. ”Such time complexity is infeasible”

• Notice the article ”a/an”, if the noun is countable: ”such a system”, ”such an algorithm”

10.3 Extra: How to derive adverbs from adjectives?

10.3.1 Basic rule

<table>
<thead>
<tr>
<th>Adjective suffix</th>
<th>Adverb suffix</th>
<th>Adverb</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>-y</td>
<td>-ily</td>
<td>easy</td>
<td>easily</td>
</tr>
<tr>
<td>-e</td>
<td>-ly</td>
<td>whole</td>
<td>wholly, true</td>
</tr>
<tr>
<td>-ic</td>
<td>-ally</td>
<td>automatic</td>
<td>automatically, systematically</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exception: public</td>
<td>publicly</td>
</tr>
<tr>
<td>-able/-ible</td>
<td>-l disappears</td>
<td>sensible</td>
<td>sensibly</td>
</tr>
<tr>
<td>-ly</td>
<td>in a &lt;adj.&gt; way</td>
<td>in a friendly way</td>
<td></td>
</tr>
</tbody>
</table>

If you are not sure how to derive an adverb, check it from a dictionary!

10.3.3 Adverb = adjective

fast, hard, lat, straight, low, wrong, right, long

Notice the difference in meaning (both can be used as adverbs):
deep vs. deeply
hard vs. hardly
high vs. highly
most vs. mostly

Task: Draw a decision tree for deriving adverbs from adjectives!

10.4 Comparing adverbs

<table>
<thead>
<tr>
<th>Adverb type</th>
<th>Comparative, superlative</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ly suffix</td>
<td>more &lt;adv.&gt;, most &lt;adv&gt;</td>
<td>more carefully, most carefully</td>
</tr>
<tr>
<td>like adjective</td>
<td>-er, -est</td>
<td>faster, fastest</td>
</tr>
</tbody>
</table>

**Exceptions:** well, badly, much, little, far

"This is a less desirable solution“, ”The X algorithm peforms worse/better than Y algorithm”

Notice:

- far, farther, farthest, when you express distance, E.g.
  "Point x lies farthest from the centre.”

- far, further, furthest, when you express distance, time, or in an abstract context. E.g.
  "In Chapter X, we will analyze this problem further” or ”This problem is further analyzed in Chapter X”

11 Parallel structures

Conjunctions and some special phrases are used to combine words, word groups (phrases), clauses or sentences. Here we concentrate on combining parallel elements. A different structure is needed for combining a main clause and a subordinate clause. → Section Sentences.

Parallel structures are used to present parallel ideas.
Parallel structure = words, phrases, clauses or sentences combined by commas and/or conjunctions. Here we call the combined items as parallel items.

- Parallel items are combined by parallel conjunctions (and, or, but, ...).
- Notice that lists are also parallel structures!
- Often the parallel structure lists alternatives or makes some kind of comparison: the items belong to the same or similar classes or to two opposite classes.
- E.g. "Method X has several advantages: it is easy to implement, it works in polynomial time, and it can use both numeric and categorical data." contains two parallel structures: three advantages ("it is, it works, it can") in a list and "both numeric and categorical data"

11.1 Basic rules

The parallel structure should be consistent in two ways

- **Semantically**: the concepts referred by parallel items should be comparable, i.e. the comparison should make sense.

- **Syntactically**: the items should have similar grammatical structure. All of them should be either nouns, noun phrases, verb phrases, or clauses. In addition, they should be in the same form, e.g. you cannot combine "to" + verb and a verb without "to".

  "The problem is both hard to define and solve"

  → "The problem is both hard to define and to solve"

11.2 Parallel items combined by conjunctions and, or, but

The most common form of parallel structures!

"The method has low space but high time requirement”

→ "The method has low space requirement but high time requirement."
"The students were told to make themselves comfortable, to read the instructions, and that they should ask about anything they did not understand" → "The students were told to make themselves comfortable, to read the instructions, and to ask about anything they did not understand"

"The results show that $X$ did not affect the error rate and the model overfitted the data" → "The results show that $X$ did not affect the error rate and that the model overfitted the data"

11.3 Lists

Notice that elements in a list should be in a parallel form!

11.3.1 Example 1

"Boud [Bou89] has listed general characteristics which are typical for problem-based courses:

- Acknowledgement of learners’ experience.
- Emphasis on students taking responsibility of their own learning.
- Crossing of boundaries between disciplines.
- Focus on the processes of knowledge acquisition rather than the products of such processes.
- Change in staff role from instructor to facilitator.
- Students’ self- and peer assessment of learning.
- Focus on communication and interpersonal skills."

11.3.2 Example 2

"The clustering methods can be divided into three categories:

1. *Hierarchical methods* construct a hierarchy of (typically) nested clusters.
2. *Partitioning methods* try to find optimal partitioning into a specified number of clusters.
3. *Probabilistic model-based clustering* tries to find the underlying probabilistic model which has produced the data."
11.3.3 Example 3

"The whole procedure is following:

1. Determine the number of clusters $k$
2. Choose parametric models (density functions $f_j$) for each of the clusters.
3. Determine the component probabilities $\pi_k$ and parameters $\theta_k$ from data.
4. Assign each point to the most probable cluster."

11.4 Example 4

"According to O’Shea [OSh00], an intelligent tutoring system should be

- robust,
- helpfull
- simple,
- transparent
- flexible
- ...
- sensitive, and
- powerfull."

Notice! The previous kind of list should be avoided, because it can be written as normal sentences. A list was used above, because 13 items were listed (and they were analyzed later). If you list only a couple of items (e.g. less than 5), write them as a normal sentence!
11.5 Parallel items combined by conjunction pairs

Sometimes the parallel expression consists of two conjunctions like

- between...and,
- both...and,
- either...or,
- neither...nor, and
- not only...but.

The first conjunction should be immediately before the first part of the parallelism.

11.5.1 between – and

"between 20-22 years of age" → "between 20 and 22 years of age"

"We recorded the difference between the students who completed the first task and the second task"

→ "We recorded the difference between the students who completed the first task and the students who completed the second task."

11.5.2 both – and

"The task is both easy to solve and efficient." (Doesn’t make any sense!)

→ The task is both easy to solve and can be solved efficiently.

Or another structure:

"The task is easy and the solution is efficient."

11.5.3 either – or

"The students either gave the worst answer or the best answer."

→ "The students either gave the worst answer or gave the best answer." or

"The students gave either the worst answer or the best answer."
11.5.4 neither – nor

In negative clauses → less often needed in sciwri! (Say things in a positive way, when possible.)

"X solves the problems of traditional clustering algorithms. Neither outliers nor missing values affect the clustering quality."

(Grammatically correct, but better to say: "X solves the problems of traditional clustering algorithms. It is not sensitive to outliers or missing values.")

11.5.5 not only – but (also)

"The task is not only easy to solve but also efficient"
→ "The task is not only easy to solve but the solution is also efficient” or "The task is not only easy to solve but it can also be solved efficiently"

Once again: say in a positive way, when possible – clearer!

11.5.6 On the one hand – on the other hand

• A special expression: can combine either clauses or parallel sentences!

• An affective way to describe opposite points, like advantages and disadvantages!

"On the one hand, a complex model can describe the data well, but on the other hand, it overfits easily."

"There is always a wrestling between the descriptive power and the generalization ability. On the one hand, too complex a model describes the data well, but it does not generalize to any new data. On the other hand, too simple a model generalizes well, but it does not describe the essential features in the data."

11.6 The comparative – the comparative

The comparative forms of adjectives can used in a parallel way in the following structure:
"The more complex the model is, the better it describes the training data."

If $x$ and $y$ are missing, then no comma:
"The sooner the better."

Notice: Use sparsely!

11.7 Parallel sentences

Numerating properties or ideas is an efficient way to create logical structures into paragraphs. The sentences in the list begin by ordinal numbers "First, Second, Third". (Notice: you can say "Firstly", but there is no need for that!)

"$X$ model has three important properties: First, the model structure is easy to understand. This is a critical feature in adaptive learning environments, as we have noted before. Second, the model can be learnt efficiently from data. There are feasible algorithms for both numeric and categorial data. Third, the model tolerates noise and missing values."