

Annotation Guidelines: marking important units of text for summarisation
Latest Update: 15/07/02
Author: Laura Hasler

Length restriction: **30%** - mark **15%** of **essential/very important sentences**, and a further **15%** of **important sentences**.

Classification: very important/essential, important, unimportant.

Identify the main topic of the text.

Mark **sentences** which refer to this topic and give important information about it – mark those you see as **very important/essential**, keeping under 15% as far as possible.

Do not include sentences which contain the same information as others you have marked – if there are several sentences containing similar information, pick the most appropriate one (not necessarily the most descriptive or the longest etc., but that which most succinctly expresses the essential information).

e.g. “Inflow of export proceeds picking up: \$300 million likely by February 15” is preferable to “The inflow of stuck-up export proceeds has picked up pace and at least \$300 million are expected before the dead-line of February 15, say banking sources.”

Indicate sentences which should be included if other sentences are marked. Do not just mark them as very important, but also indicate that there is a relationship between them. For example, if you mark a sentence containing an anaphoric reference, you should also mark the sentence containing its antecedent.

e.g. If the second sentence here, containing “it” is marked as important, then the first sentence with the noun phrase “the Festival” also has to be marked. “For film lovers the Festival 's the place to be in September. It grows from strength to strength each year . . .”

Include sub-headings, as they usually summarise the following section, but only if important to the main topic of the text.

Do not mark sentences concerning sub-topics unless they directly influence the main topic (or present new, essential information on it) and do not repeat information – the main topic is essentially what the text is about.

Do not include direct speech unless it presents new and vital information concerning the main topic of the text not presented elsewhere. It is important to distinguish between speech and other text in quotation marks, which may be important.

e.g. “But its “killer app” is reinventing how the software industry works.”

Do not include examples (including constructions starting with e.g., for example, such as, like, for instance etc.)

Do not include tables or figures.

Once the very important/essential sentences have been marked, indicate those segments of these sentences which are not vital. Do not mark single words as unimportant. So, **within the sentences already marked as essential/very important:**

Mark irrelevant subordinate clauses as unimportant.

e.g. the following “which” clause could be viewed as unimportant: “Customer interest is high for the whole product line, which underlines the strong fundamentals of the new period of growth.”

Mark text in brackets and text occurring between dashes (- . . . -) as unimportant, unless vital to the main topic.

e.g. bracketed text would be considered important if it is an abbreviation which will replace a noun phrase later in the text, as in: “a Poverty Reduction and Growth Facility (PRGF) . . . It is intended that PRGF-supported programmes . . .”

Mark examples (see above) as unimportant.

Mark adjuncts (not just single words within sentences) containing dates, times, places etc. as unimportant unless vital to the main topic.

e.g. “during the last two months”

Mark constructions such as “a spokesman said”, “it was claimed”, “she told *The Guardian*”, “he explains” as unimportant.

Mark phrases which elaborate on information, such as “in addition to . . .”, “due to . . .”, “compared to . . .”, as unimportant unless vital to the main topic.

Once you have completed this, if the final amount of text marked is substantially below 15% of the full text, try and add more units which you consider very important/essential to bring the percentage up.

Then do the same for those sentences which you would classify as **important**, again keeping as close to 15% as possible.

Comment on the annotation – any problems, indecisions etc.