Byte Pair Encoding (BPE) is a significant algorithm in Natural Language Processing (NLP) with practical applications and benefits. There are 2 main types, which are known as original BPE and repurposing BPE.

Original BPE: Byte Pair Encoding (BPE) was originally developed for data compression, where the goal was to represent data using fewer bytes. The algorithm worked by iteratively replacing the most frequent pair of bytes in the data with a new byte not present in the data. This process continued until no more pairs could be merged, resulting in a compressed representation of the original data1.

Repurposing BPE: By taking original BPE and modifying or adapting it for a different use case. Here are some variant of BPE:

* WordPiece: This is a variant of BPE used in models like BERT.
* SentencePiece: This is another variant that treats the input as a raw input string, thus including space in the character set.
* Unigram Language Model: This is a subword regularization and BPE dropout. It is a variant of BPE that uses a unigram language model.

**Applications of BPE in Life:**

* **Machine Translation**: BPE helps in translating text between languages by managing large vocabularies efficiently.
* **Text Classification**: It is used to categorize text into predefined groups, which is useful for filtering spam or sorting articles.
* **Text Generation**: BPE aids in generating human-like text, which can be used in chatbots and virtual assistants.

**Advantages of BPE:**

* **Reduces Vocabulary Size**: BPE represents a large vocabulary with a smaller set of subword units, making it memory-efficient.
* **Dealing with rare words**: This refers to the challenge in Natural Language Processing (NLP) of handling words that don’t appear often in a text or dataset. These “rare” words can be difficult for a model to understand and use correctly.
* **Character level embeddings**: Character level embeddings are a way of representing words as a sequence of character-level information. This method is a part in handling rare words and it helps the machine understand the structure of words and the relationships between them at a very granular level. It’s especially good for understanding different forms of the same word and dealing with words that the computer hasn’t seen before. It’s also useful for dealing with languages that have complicated spelling rules.
* **Improves Model Performance**: By reducing the number of out-of-vocabulary words, BPE can enhance the performance of NLP models.
* **Multilingual Support**: BPE can handle multiple languages, especially those with complex morphology, making it valuable for multilingual applications.

**Disadvantages of BPE:**

* **Suboptimal for Certain Languages**: Some languages may benefit more from morphological segmentation than from BPE.
* **Greedy Algorithm**: BPE is a greedy algorithm, which means it may not always produce the globally optimal set of subword units.

**What Makes BPE Better Than Other Algorithms:**

* **Flexibility**: BPE’s ability to break words into subword units provides flexibility in representing words not seen during training.
* **Efficiency**: It is efficient in handling large vocabularies and can be used across various NLP tasks.
* **Longer sequences**: Due to basing on the character level embeddings, result in longer sequences compared to word-level embeddings, which can make the computational requirements higher.
* **Can’t capture the semantics of longer words:** Might not effectively capture the semantics of longer words compared with word-level or subword-level embeddings.
* **Byte-Level BPE**: This variant of BPE operates at the byte level, ensuring that every character, including emojis, is represented without being converted to an unknown token.

In summary, BPE brings the perfect balance between character- and word-level hybrid representations which makes it capable of managing large corpora. This behavior also enables the encoding of any rare words in the vocabulary with appropriate subword tokens without introducing any “unknown” tokens. This especially applies to foreign languages like German where the presence of many compound words can make it hard to learn a rich vocabulary otherwise.