and *"judge"* are an instance of *"same polarity substitution"*, while *"A federal magistrate ... ordered"* and *"Zuccarini was ordered by a federal judge..."* are an instance of *"diathesis alternation"*[2].

1a A federal **magistrate** in Fort Lauderdale ordered him held without bail.

1b Zuccarini was ordered held without bail Wednesday by a federal **judge** in Fort Lauderdale, Fla.

Second issue is that atomic paraphrases can appear in textual pairs that are not paraphrases. The two texts in 2a and 2b as a whole are not textual paraphrases, even if they have a high degree of lexical overlap and a similar syntactic structure. However, *"Microsoft"* and *"shares of Microsoft"* are an instance of *"same polarity substitution"* - both phrases have the same role and meaning in the context of the two sentences. This demonstrates the possibility of atomic paraphrases being present in textual non-paraphrases. [3]

2a Microsoft fell 5 percent before the open to $27.45 from Thursday's close of $28.91.

2b Shares in Microsoft slipped 4.7 percent in after-hours trade to $27.54 from a Nasdaq close of $28.91.

Third issue is that in certain cases, the semantic relation between the elements in an atomic paraphrase can only be interpreted within the context (as shown in the work of Shwartz and Dagan [2016]). The two texts in 3a and 3b are textual paraphrases. The out-of-context meaning of *"cargo"* and *"explosives"* differs significantly, however within the given context, they are an instance of *"same polarity substitution"*.

3a They had published an advertisement on the Internet on June 10, offering the cargo for sale, he added.

3b On June 10, the ship's owners had published an advertisement on the Internet, offering the explosives for sale.

in MRPC.

[2]These types and annotation are from Vila et al. [2015].

[3]In fact, it is possible to find atomic paraphrases within pairs of texts connected by various relations, such as entailment, simplification, summarization, contradiction, and question-answering, among others. This is illustrated by the significant overlap of atomic types in Paraphrase Typology research and typology research in Textual Entailment.

And finally, 4a and 4b illustrate an issue that is often overlooked in theoretical paraphrase research: the linguistic phenomena behind certain atomic paraphrases do not always preserve the meaning. The meanings of *"beat"* and *"battled"* are similar, and play the same syntactic and discourse role in the structure of the texts. Therefore, the substitution of *"beat"* for *"battled"* fulfills the formal requirements of a *"same polarity substitution"*. However, after this substitution, the resulting texts are not paraphrases as they differ substantially in meaning.

4a He beat testicular cancer that had spread to his lungs and brain.

4b Armstrong, 31, battled testicular cancer that spread to his brain.

5.3.3 Objectives of EPT and Research Questions.

We argue that the objectives behind a paraphrase typology are twofold: 1) to classify and describe the linguistic phenomena involved in paraphrasing (at the atomic level); and 2) to provide the means to study the function of atomic paraphrases within pairs of texts of arbitrary size and with various semantic relations (such as, textual paraphrases, textual entailment pairs, contradictions, and unrelated texts).

Traditionally, the authors of paraphrase typologies have focused on the first objective while the latter is mentioned only briefly or ignored altogether. In our work, we want to extend the existing work on paraphrase typology in the direction of Objective 2, as we argue that it is crucial for applications. We pose four research questions, that we aim to address with the creation of EPT and ETPC:

RQ1 what is the relation between atomic and textual paraphrases considering the distribution of atomic paraphrases in textual paraphrases?

RQ2 what is the relation between atomic paraphrases and textual non-paraphrases considering the distribution of atomic paraphrases in textual non-paraphrases?

RQ3 what is the role of the context in atomic paraphrases?

RQ4 in which cases do the linguistic phenomena behind an atomic paraphrase preserve the meaning and in which they do not?

5.3.4 The Extended Paraphrase Typology

The full Extended Paraphrase Typology is shown in Table 5.1. It is organized in seven meta categories: "Morphology", "Lexicon", "Lexico-syntax", "Syntax", "Discourse", "Other", and "Extremes". Sense Preserving (Sens Pres.) shows

Table 5.1 Extended Paraphrase Typology

ID	Type	Sense Pres.
	Morphology-based changes	
1	Inflectional changes	+ / -
2	Modal verb changes	+
3	Derivational changes	+
	Lexicon-based changes	
4	Spelling changes	+
5	Same polarity substitution (habitual)	+
6	Same polarity substitution (contextual)	+ / -
7	Same polarity sub. (named entity)	+ / -
8	Change of format	+
	Lexico-syntactic based changes	
9	Opposite polarity sub. (habitual)	+ / -
10	Opposite polarity sub. (contextual)	+ / -
11	Synthetic/analytic substitution	+
12	Converse substitution	+ / -
	Syntax-based changes	
13	Diathesis alternation	+ / -
14	Negation switching	+ / -
15	Ellipsis	+
16	Coordination changes	+
17	Subordination and nesting changes	+
	Discourse-based changes	
18	Punctuation changes	+
19	Direct/indirect style alternations	+ / -
20	Sentence modality changes	+
21	Syntax/discourse structure changes	+
	Other changes	
22	Addition/Deletion	+ / -
23	Change of order	+
24	Semantic (General Inferences)	+ / -
	Extremes	
25	Identity	+
26	Non-Paraphrase	-
27	Entailment	-