8.5 Analysis of the Results

Before this paper, the comparison between textual meaning relations was limited to measuring the overlap and correlation between the binary label of the pairs. Gold et al. [2019] present such an analysis. They find some expected results such as the high correlation and overlap between paraphrasing and (uni-directional) entailment and the negative correlation between paraphrasing and contradiction or entailment and contradiction. They also report some interesting and unexpected results. They point that in practical setting paraphrasing does not equal bi-directional entailment. With respect to specificity they find that it does not correlate with other textual meaning relations, and does not overlap with textual entailment.

In this section, we go further than the binary labels of the textual meaning relations and compare the distribution of types across all relations. A typological comparison can be much more informative about the interactions between the different relations.

This section is organized as follows. Section 8.5.1 analyzes and compares the frequency distribution of the different types in pairs with the following textual relations: Paraphrasing, Textual Entailment, and Contradiction. Section 8.5.2 discusses the Specificity relation and the types involved in it.

8.5.1 Type Frequency

To determine the similarities and differences between the textual meaning relations in terms of types, we measured the relative type frequencies for pairs that have the corresponding label. Table 8.4 shows the relative frequencies in pairs that have paraphrasing, entailment, or contradiction relations at textual level. For the entailment relation we consider only the pairs marked as "uni-directional entailment". That is, pairs that have entailment only in one of the directions. We discard the pairs that have bi-directional entailment to reduce the overlap with paraphrases (94 % of the bi-directional entailment pairs are also paraphrases).

For reference, we have also included the type frequencies for the paraphrase portion of the ETPC [Kovatchev et al., 2018a] corpus. ETPC is the largest corpus to date annotated with paraphrase types. The EPT typology used to annotate the ETPC also shares the majority of the linguistic types with SHARel. This allows us to put our results in perspective and to determine to what extent are they consistent with previous findings.

Table 8.4 Type Frequencies

ID	Type	Paraph.	Entailment	Contradiction	ETPC
Morphology-based changes					
1	Inflectional changes	4 %	4 %	1.9 %	2.78 %
2	Modal verb changes	0.25 %	1 %	0	0.83 %
3	Derivational changes	2 %	0	0.6 %	0.85 %
Lexicon-based changes					
4	Spelling changes	0.25 %	0.4 %	0	2.91 %
5	Same pol. sub. (habitual)	25.2 %	17 %	26 %	8.68 %
6	Same pol. sub. (contextual)	9.7 %	6.3 %	5.5 %	11.66 %
7	Same pol. sub. (named ent.)	0.7 %	0.4 %	1.2 %	5.08 %
8	Change of format	0.7 %	0.9 %	0	1.1 %
Lexico-syntactic based changes					
9	Opposite pol. sub. (habitual)	2.7 %	3.5 %	7.5%	0.07 %
10	Opposite pol. sub. (context.)	0.5 %	0.9 %	1.2 %	0.02 %
11	Synthetic/analytic sub.	6.7 %	6.8 %	3.7 %	3.80 %
12	Converse substitution	2.5 %	3.2 %	3.1 %	0.20 %
Syntax-based changes					
13	Diathesis alternation	1.5 %	2.2%	1.9 %	0.73 %
14	Negation switching	4 %	4 %	11.2 %	0.09 %
15	Ellipsis	0	0	0	0.30 %
16	Anaphora	1.7 %	2.7 %	0.6 %	0
17	Coordination changes	0	0	0	0.22 %
18	Subordination and nesting	0.25 %	0	0	2.14 %
Discourse-based changes					
18	Punctuation changes	0	0	0	3.77 %
20	Direct/indirect style altern.	0	0	0	0.30 %
21	Sentence modality changes	0	0	0	0
22	Syntax/discourse structure	0	0	0	1.39 %
Other changes					
23	Addition/Deletion	16.25 %	16.4 %	16.2 %	25.94 %
24	Change of order	0.5 %	0.9 %	0.6 %	3.89 %
Extremes					
25	Identity	12.5 %	14.5 %	11.8 %	17.5 %
26	Unrelated	0	0	0	3.81 %
Reasoning					
27	Cause and Effect	4.7 %	5.4 %	5 %	n/a
28	Conditions and Properties	2 %	6 %	0.6 %	n/a
29	Funct. and Mutual Exclus.	0	0.4 %	0	n/a
30	Named Ent. Reasoning	0	0	0	n/a
31	Numerical Reasoning	0	0	0	n/a
32	Temp. and Spatial Reasoning	0	0	0	n/a
33	Transitivity	0.25 %	0.9 %	0	n/a
34	Other (General Inference)	0.5 %	0.4 %	0.6 %	1.53 %

We can observe that the distribution of types is not balanced for any of the portions. Some types are over-represented, while others are under-represented or not represented at all. We focus our analysis on four different tendencies: 1) linguistic types that are frequent across all relations; 2) types whose frequency changes across the different relations; 3) the frequency of reason-based types; and 4) types that are infrequent or not represented at all.

Frequent linguistic types across all relations

The most frequent types across all relations are *same polarity substitution (habitual)* (#5), *same polarity substitution (contextual)* (#6), *same polarity substitution (named entity)* (#7), *addition/deletion* (#23), and *identity* (#25). These phenomena account for more than 50% of the types in the corpus. This finding is also consistent with the results reported in the ETPC. It is worth noting that in the ETPC, the distribution within the different *same polarity substitution* types (#5, #6, #7) differs from our annotation. The frequency of *same polarity substitution (habitual)* (#5) is lower, while *same polarity substitution (contextual)* (#6) and *same polarity substitution (named entity)* (#7) have a much higher frequency.

Other frequent types shared across all relations are *inflectional* (#1), *opposite polarity substitution (habitual)* (#9), *synthetic/analytic substitution* (#11), *converse substitution* (#12), *diathesis alternation* (#13), and *negation switching* (#14). For all of these types, the frequency that we obtain is substantially higher than in the ETPC corpus.

Differences in type frequencies across relations

We can observe that paraphrasing has the highest frequency of *Same Polarity Substitution*, both habitual (#5) and contextual (#6). This is a tendency that can also be observed in ETPC.

Entailment is the relation with the highest relative frequency of phenomena in the reason-based category. The reason-based phenomena (#27-#34) account for 13.1% of all phenomena within entailment, doubling the frequency of these phenomena in paraphrasing (5.65%) and contradiction (6.2%). Most of that difference comes from the "conditions/properties" (#28) type. The entailment relation also has the lowest frequency of same polarity substitutions (#5, #6, and #7).

Contradiction has the highest frequency of opposite polarity substitution (#9 and #10) and negation switching (#14), doubling the frequency of these phenomena in paraphrasing and entailment pairs. Interestingly, contradictions have a comparable frequency of same polarity substitution (#5, #6, and #7) and identity (#25)