to paraphrases. This suggests that contradictions are more similar to paraphrases than to entailment, at least in terms of the phenomena involved.

Frequency of reason-based types

We can observe that reason-based types (#27-#34) are much less frequent than linguistic types. Reasoning accounts for less than 14% of the examples across all relations. That means that in the majority of the cases, the textual relation can be determined via linguistic means and does not require reasoning or world knowledge. The most frequent reasoning type across all relations is *cause/effect*.

It is important to note that the frequency of reasoning phenomena in our annotation is much higher than the 1.5% reported in ETPC. In ETPC all reason based phenomena were annotated with a single label - *Other (General Inferences)* (#34) so the frequency of this type corresponds to the sum of all types from #27 to #34 in our annotation. These findings indicate that the methodology of Gold et al. [2019] successfully addresses one of the problems in the ETPC corpus, already emphasized by other researchers - the lack of reason-based types.

Low frequency types and missing types

In our annotation, there are several linguistic and reason-based types that are not represented at all. Regarding the linguistic types, there are no discourse based types, no *ellipsis* (#15), no *coordination changes* (#17), and almost no *subordination and nesting changes* (#18). Regarding the reason-based types, there are no *Named Entity Reasoning* (#30), *Numerical Reasoning* (#31), and no *Temporal and Spatial Reasoning* (#32).

We argue that the absence of these types in our annotation is due to the way in which the Gold et al. [2019] corpus was created. The authors of that corpus aimed at obtaining simple, one-verb sentences. The average length of a sentence is 10.5 tokens, which is much lower than the length of sentences in other corpora (ex.: 22 average length for ETPC). The corpus contains almost no Named Entities (proper names, locations, or quantities). These characteristics of the corpus do not facilitate transformations at the syntactic and discourse levels or Named Entity Reasoning.

Our intuition that the lack of these types is due to the corpus creation is further reinforced by the fact that these types are missing across all meaning relations. However, these missing types can be observed in other paraphrasing and entailment corpora, such as Sammons et al. [2010], Cabrio and Magnini [2014], and Kovatchev et al. [2018a]. For these reasons we decided to keep them as part of the

ShaRel typology. It would, nevertheless, require a further research and richer corpora to empirically determine the importance of these phenomena for the different meaning relations.

Summary The similarities and common tendencies between paraphrases, entailment, and contradiction clearly indicate that these relations belong within the same conceptual framework and should be studied and compared together. The results also suggest the possibility of the transfer of knowledge and technologies between these relations.

The differences between the textual meaning relations in terms of the involved types can help us to understand each of the individual relations better. This information can also be useful in the automatic classification of the different relations in a practical task.

8.5.2 Decomposing Specificity

We define specificity as the opposite of generality or fuzziness. Yager [1992] defines specificity as the degree to which a fuzzy subset points to one element as its member. This meaning relation has not been studied extensively. It has also not been decomposed. To the best of our knowledge this is the first work to do so. Gold et al. [2019] show that there is no direct correlation between specificity and the other textual meaning relations, including textual entailment. For that reason, we took a different approach to the decomposition of specificity and treat it separately from the other relations. We added one extra step in the annotation process, focused on the specificity relation.

The corpus of Gold et al. [2019] is annotated for specificity at the textual level. That is, the crowd workers identified which of the two given sentences is more specific. In 9, the annotators would indicate that **b** is more specific than **a**.

9 **a** All children receive the same education.
b The same education is received by all girls.

In our annotation, we performed an additional annotation of the specificity and we identified the particular elements (words, phrases, clauses) in one sentence that were more specific than their counterpart. In example 9, we can identify that "girls" is more specific than "children". The difference in the specificity of "girls" and "children" is the reason why **b** is annotated as more specific than **a**. We called that "scope of specificity".

In 80% of the pairs with specificity at textual level, our annotators were able to point at one or more particular elements that are responsible for the difference in specificity. In 20% of the pairs, the specificity was not decomposable. This

finding also confirms Ko et al. [2019]'s findings, who showed that frequency-based features are well-suited for automatic specificity detection.

In our analysis on the nature of the specificity relation, we combined the annotation of "scope of specificity" and the traditional annotation of linguistic and reason-based types discussed in the previous sections. In particular, we looked for overlap between the "scope of specificity" and the scope of linguistic and reason-based types. Example 10 shows the two separate annotations side by side. In **a** and **b**, we show the annotation of the linguistic and reason-based types: *"same polarity substitution (habitual)"* of "children" and "girls", and *"diathesis alternation"* of "receive" and "is received by". In **c** and **d** we show the annotation of the specificity: "children" - "girls". When we compare the two annotations we can observe that the "scope of specificity" overlaps with the scope of *"same polarity substitution (habitual)"*.

10 **a** All children *receive* the same education.
b The same education *is received* by all girls.
c All **children** receive the same education.
d The same education is received by all **girls**.

We argue that when there is an overlap between the "scope of specificity" and a linguistic or a reason-based type, it is the linguistic or reason-based phenomenon that is responsible for the difference in specificity. In example 10 we can say that the substitution of "children" and "girls" is responsible for the difference of specificity.

Table 8.5 shows the overlap between "scope of specificity" and "atomic types". In 97 % of the cases where specificity was decomposable the more/less specific elements overlapped with an atomic type. In 50 % of the cases the specificity was due to additional information (#23). The other frequent cases include *same polarity substitution* (#5, #6, and #7), *synthetic/analytic substitution* (#11), and *cause and effect* (#27) reasoning. While the overall tendencies are similar to the other meaning relations, specificity also has its unique characteristics. We found almost no specificity at morphological level and the frequency of *Same polarity substitution* (#5, #6, and #7), while still high, was lower than that of paraphrasing and contradiction pairs. The relative frequency of *Synthetic/analytic substitution* (#11) was the highest of all relations and the reasoning types were almost as frequent as in entailment pairs, although the type distribution is different. We found no syntactic or discourse driven specificity changes.