Towards a Taxonomy of Textual Entailments*

Adam Przepiórkowski, Jakub Kozakoszczak, Jan Winkowski, Daniel Ziembicki, and Tadeusz Teleżyński

[1] Institute of Computer Science, Polish Academy of Sciences, Warszawa, Poland
[2] University of Warsaw, Warszawa, Poland
adamp@ipipan.waw.pl, jkozakoszczak@gmail.com, jaswinko@gmail.com, seattlegrunge@wp.pl, tadekte@gmail.com

Abstract

This paper reports on work in progress which aims at the creation of a comprehensive taxonomy of textual entailment rules and – as a proof of concept – its application to an RTE corpus. In particular, a new formalism for encoding entailment rules is proposed, with some emphasis on encoding the properties of such rules and their converses in different polarity contexts.

1 Introduction

Ever since Aristotle, entailment has been in the centre of interest in logic and semantics, and since mid-2000s the more loosely defined textual entailment has become increasingly important in Natural Language Processing (NLP), where semantic systems are evaluated by their performance on the Recognising Textual Entailment (RTE) task (Dagan *et al.* 2006, 2013, Sammons 2015; see also Bos 2008). To this end, RTE corpora have been created[1] consisting of ⟨T, H⟩ pairs tagged by humans with information whether the hypothesis H is textually entailed by the attested text T, or not; in case of some corpora, a third tag, signalling contradiction between T and H, is also employed. The quality of a system is measured by the degree of agreement between human judgements and tags assigned by the system. The additional difficulty of the task lies in the fact that textual entailment is understood very generally and a little vaguely; the following definition is typical: "[The] applied notion of *textual entailment* is defined as a directional relationship between pairs of text expressions, denoted by T – the entailing 'Text', and H – the entailed 'Hypothesis'. We say that T entails H if, typically, a human reading T would infer that H is most likely true" (Dagan *et al.* 2006: 178).

Evaluation against such RTE corpora is purely quantitative: the only information it provides is the percentage of correctly tagged pairs, and not the kinds of phenomena the system does not handle correctly or the kinds of knowledge it lacks. For this reason, since around 2010, the need has been voiced for corpora containing pairs tagged with the types of phenomena or knowledge involved in the process of reasoning (Sammons *et al.* 2010). Different types of textual inference had, of course, been discussed earlier (e.g. Cooper *et al.* 1996, Zaenen *et al.* 2005, Garoufi 2007, Clark *et al.* 2007, Bos 2008), but such discussions resulted in identifying some types of phenomena or kinds of knowledge, without an attempt at creating an exhaustive taxonomy of types of textual entailment. Other work discusses selected types or aspects of reasoning, e.g. monotonicity effects (MacCartney and Manning 2007), common-sense knowledge (LoBue and Yates 2011), linguistic modification (Toledo *et al.* 2012), etc.

*Work reported here has been partially financed by the Polish Ministry of Science and Higher Education within the CLARIN ERIC programme 2015–2016 (http://clarin.eu/).

[1] Some are listed at http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Resource_Pool#RTE_data_sets.

To the best of our knowledge, no comprehensive taxonomy of types of textual entailment has been proposed so far, and the most advanced attempt is still that of Bentivogli *et al.* 2010. There, 90 pairs extracted from the RTE-5 (Bentivogli *et al.* 2009) corpus were annotated with kinds of phenomena involved in the entailment. The 36 distinguished phenomena were grouped into 5 categories: lexical (e.g. *hypernymy* and *geographical knowledge*), lexical-syntactic (e.g. *causative* and *paraphrase*), syntactic (e.g. *negation*, *list* and *apposition*), discourse (e.g. *coreference* and *ellipsis*) and reasoning (e.g. *apposition* again, *meronymy*, *relative clause* and the most frequent "phenomenon" of *common background / general inferences*). As the authors note themselves, "the list is not exhaustive and reflects the phenomena [they] detected in the sample of RTE-5 pairs [they] analyzed". Unfortunately, the scope of particular phenomena and reasons for classifying, say, *geographical knowledge* as a lexical phenomenon and *meronymy* as reasoning, are not elucidated, as that short paper concentrates on general methodological issues.

The aim of the current paper, which reports on work in progress, is to start filling this gap, i.e. to report on early research leading to 1) the creation of a comprehensive taxonomy of textual entailment steps and – as a proof of concept – 2) its application to an RTE corpus. We assume that each ⟨T, H⟩ pair tagged as involving entailment may be associated with a sequence of atomic entailment steps, cf. §2, that each such a step may be formalised as a rule transforming one text into another text (or, more declaratively, specifying a relation between texts), cf. §3, and that such rules may be organised into a taxonomy of textual entailment types, cf. §4. In specifying the format of the rules and the resulting taxonomy, we combine and extend ideas from the literature which have not been put together before.

While all the examples provided in this paper involve English, the actual entailment corpus is a Polish translation of the development part of the RTE-3 corpus (Giampiccolo *et al.* 2007) consisting of 800 ⟨T, H⟩ pairs, including 411 pairs tagged as involving true entailment. At the point of submitting this paper to the proceedings, over 120 ⟨T, H⟩ pairs are fully tagged with almost 800 atomic entailment steps (over 6 steps per pair, on average), and some 50 of the rules used in these entailments – jointly covering about 120 atomic entailment steps – are fully formalised in the sense of §3. While, as discussed there, some of the rules are language-specific and others are language-independent, they are organised into a taxonomy outlined in §4, which is designed to be largely language-independent. The possible uses of such a corpus and taxonomy are mentioned in the concluding remarks in §5.

2 Entailment corpus

One novel aspect of the research presented here consists in the construction of a textual entailment corpus, where particular entailment pairs are not only tagged with kinds of reasoning involved, as already postulated (but not fully realised) in the literature, but where all transformation steps leading from the initial text to the final hypothesis are explicitly shown, as illustrated in Figure 1.[2] The five steps shown there conflate eight actual steps in the corpus. In the first step of the figure, appropriate rules for dropping a modifier are applied three times: for the two temporal modifiers *In 1927* and *until his death in 1962*, and for the subordinate *following on its success*. Similarly, the final step in the figure involves two applications of a rule substituting a term with its synonym: *penned* is replaced by *authored* and *numerous* – by many.

Obviously, this is only one of possible orderings of rule applications, but in many cases, including the example above, the ordering does not matter. An example of where the order of

[2] In the actual corpus, particular steps are labelled with the identifiers of specific rules. Moreover, the rule alluded to as *distributing shared dependent* is not strictly necessary here, but we leave it here as its formalisation will be discussed in the following section.

T:	In 1927 Harnold Lamb wrote a biography of Genghis Khan, and following on its success turned more and more to the writing of non-fiction, penning numerous biographies and history books until his death in 1962.	
	→ Harnold Lamb wrote a biography of Genghis Khan, and turned more and more to the writing of non-fiction, penning numerous biographies and history books.	(dropping modifiers)
	→ Harnold Lamb turned more and more to the writing of non-fiction, penning numerous biographies and history books.	(dropping conjunct)
	→ Harnold Lamb penned numerous biographies and history books.	(extracting *-ing* modifier)
	→ Harnold Lamb penned numerous biographies and numerous history books.	(distributing shared dependent)
	→ Harnold Lamb penned numerous biographies.	(dropping conjunct)
H:	→ Harnold Lamb authored many biographies.	(substituting synonyms)

Figure 1: Entailment steps from text T to hypothesis H for the RTE-3 pair id=65

rule applications does matter is given in Figure 2, where the first sentence of the text T may be dropped only once the coreference rule fires, as otherwise *the team* could not be replaced by the (already dropped) *The Kinston Indians*.

T:	The Kinston Indians are a minor league baseball team in Kinston, North Carolina. The team, a Class A affiliate of the Cleveland Indians, plays in the Carolina League.	
	→ The Kinston Indians are a minor league baseball team in Kinston, North Carolina. The Kinston Indians, a Class A affiliate of the Cleveland Indians, plays in the Carolina League.	(substituting coreferring term)
	→ The Kinston Indians, a Class A affiliate of the Cleveland Indians, plays in the Carolina League.	(dropping sentence)
	→ The Kinston Indians plays in the Carolina League.	(dropping modifier)
	→ Kinston Indians play in the Carolina League.	(substituting synonym)
H:	→ Kinston Indians participate in the Carolina League.	(lexical implication)

Figure 2: Entailment steps from text T to hypothesis H for the RTE-3 pair id=92

Occasionally, it is possible that the same hypothesis may be reached from the text via an application of different sets of rules. Nevertheless, as the task of devising a single line of reasoning is already rather difficult, no attempt is made to provide all possible minimal sequences of rule applications, even once a set of rules is fixed. It should be noted that this procedure goes further than that of Bentivogli *et al.* 2010, where particular reasoning steps were only applied to the initial Text, without an attempt at constructing a complete transformation path from Text to Hypothesis.

3 Textual entailment rules

We propose an extension and further formalisation of the usual textual entailment rules (Szpektor *et al.* 2007) mapping one text into another (uni- or bidirectionally). We assume that such rules operate on dependency trees, i.e. their Left- and Right-Hand Sides are templates matching a dependency (sub)tree.[3] The rules constructed so far operate at the level of syntactic dependency trees, where word nodes are marked with lemmata and with morphosyntactic information (part of speech, case, gender, etc.), and with dependency labels such as *subject* and *object*,[4] but we envisage the possibility of some rules operating at the level of semantic dependencies, with labels such as *agent* or *instrument*.[5]

Formally, a rule is a quintuple whose first two elements are dependency trees (to match the input and the output of the rule), followed by a polarity signature,[6] strength specification and applicability conditions: ⟨LHS, RHS, P, S, C⟩. A preliminary version of a simple rule for dropping a premodifier expressed by a nominal or prepositional phrase is shown in (1):[7]

(1) ⟨ nmod: X.G Y.L, Y.L, ⟨+, ∘, ∘, +⟩, S, {} ⟩

The LHS and RHS elements are particularly simple here: LHS matches any dependency of the type `nmod` (nominal modifier) where the dependent (indicated by the variable X) is on the left of the head (indicated by Y), and the RHS is the result of dropping the dependent (i.e. leaving Y). An example of the operation of this rule is given in Figure 3. This example also illustrates

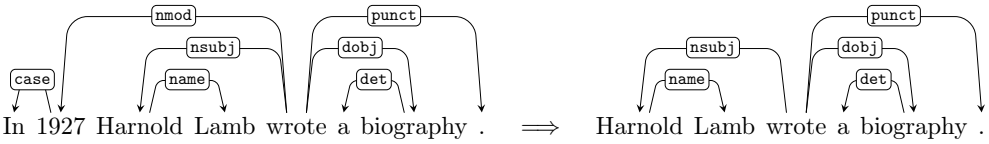


Figure 3: An example of the operation of rule (1)

the need for the markers L and G adorning the nodes in LHS and RHS: the former stands for *lazy match*, the latter – for *greedy match*. Lazy match occurs when a node is matched but those of its dependents which are not explicitly specified in the rule are not "consumed" during the match – such dependents will be rewritten to the output of the rule. On the other hand, greedy match indicates the need to "consume" – i.e. get rid of – any unspecified dependents. Hence, when Y.L matches *wrote*, all its dependents which are not explicitly mentioned by the rule – i.e. the subject *Harnold Lamb* and the direct object *a biography* – will be rewritten together with Y in the output. On the other hand, when X.G matches *1927*, its dependent[8] *In* will also be consumed rather than being adjoined to some higher node in the output.

[3] See Tesnière 1959, 2015 for the original work on dependency grammar and de Marneffe *et al.* 2014 for the approach assumed here.

[4] In all examples, part of speech tags and dependency labels follow the Universal Dependencies (`http://universaldependencies.github.io/docs/`).

[5] Two of the linguistic theories that posit multiple parallel dependency levels are Meaning-Text Theory (Mel'čuk 1981, 1988) and Functional Generative Description (Sgall *et al.* 1986).

[6] Generalisation to monotonicity is planned for future work.

[7] Analogous rules take care of nominal postmodifiers and clausal premodifiers, as also needed in the first step of Figure 1.

[8] Admittedly, treating nouns as heads and prepositions as their dependents in prepositional phrases is a controversial aspect of Universal Dependencies.

The third element of the rule quintuple is the polarity signature, which indicates the behaviour of the rule (the first two characters) and its converse (the next two characters) depending on the polarity of the context (to be discussed below). It is best to visualise this quadruple signature as a 2×2 table, where the first column indicates the behaviour of the rule in positive polarity contexts and the second column – in negative polarity contexts, and the first row describes the rule from LHS to RHS, while the second row – its converse, i.e. the rule from RHS to LHS:

(2) $\langle 1, 2, 3, 4 \rangle \equiv$

	Pos	Neg
LHS $\rightarrow$ RHS	1	2
LHS $\leftarrow$ RHS	3	4

Following the discussion in Nairn *et al.* 2006, we assume that one of three possible values may occur in each of the cells:

(3)

$+$	the rule is valid in the specified direction and polarity context without any further conditions,
$-$	the rule is valid in the specified direction and polarity context, but the polarity of the consequent must be reversed,
$\circ$	neither of the above, i.e. the rule is not applicable in the specified direction and polarity context.

Hence, the polarity specification $\langle +, \circ, \circ, + \rangle$ in (1) means that the rule may be applied in positive contexts (as in Figure 3), but not in negative contexts, and its converse may be applied in negative – but not in positive – contexts (e.g. to transform *Harnold Lamb didn't write a biography.* into *In 1927 Harnold Lamb didn't write a biography.*).

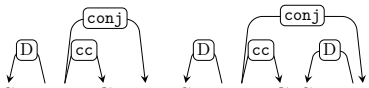
Following the suggestion of Zaenen *et al.* 2005, we also further classify rules as *strong* (*S*; including semantic entailments and presuppositions) and *defeasible* (*DF*; including conversational implicatures). Since an attempt is made to annotate with entailment steps all ⟨T, H⟩ pairs marked in the development section of RTE-3 as involving true entailment, even those where entailment steps are so weak or counterintuitive that they should not be even marked as *DF*, some rules are tagged as *wishful thinking fallacies* (*WTF*).

The final element is a set of additional constraints on the applicability of the rule. In case of (1), this set is empty, but it may involve references to word classes or relations between words, among other constraints, as in the following rule, which relates for example *a team of European astronomers* to *a team of astronomers from Europe* (Bentivogli *et al.* 2010: 3544). This rule contains constraints on parts of speech of all nodes and on the lexical relation of demonymy between nodes X and Z:

(4) ⟨X Y.L, Y.L from Z, ⟨+, +, +, +⟩, *S*, {ADJ(X), NOUN(Y), PROPN(Z), DEMONYM(X,Z)}⟩

This rule defines a full equivalence, valid in both types of context and in either direction, hence the four pluses in the signature. Note also that some nodes in the LHS and RHS are not adorned with either L or G – they are assumed to have no further dependents (this effectively becomes another condition on the applicability of this rule).

We do not propose a subformalism for expressing constraints, but acknowledge that much of the expressiveness of the proposed formalism for textual rules lies in the expressiveness of the constraint subformalism. Consider the antepenultimate entailment step of Figure 1, where the relevant rule is alluded to as *distributing shared dependent*. The rule may be formalised as in (5), where D stands for any dependency relation, and an example of its use is given in Figure 4.



(5) $\langle$ S.L X.L C Y.L, S.L X.L C S.L Y.L, $\langle +, +, +, + \rangle$, DF, {} $\rangle$

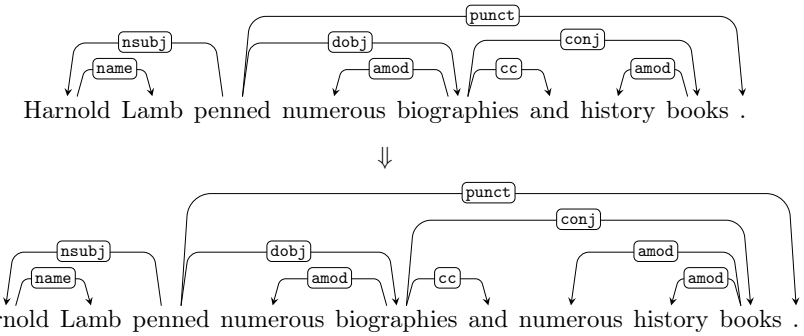
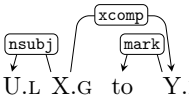


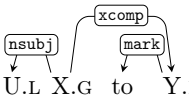
Figure 4: An example of the operation of rule (5)

This rule is marked as defeasible, as constructions such as *numerous biographies and history books* are infamously ambiguous (here *numerous* may refer to both conjuncts, *biographies* and *history books*, or just to the first conjunct, *biographies*), and the rule is justified only in one (perhaps the more natural one) of the two readings. However, if the texts are parsed and disambiguated semantically, a constraint could be added that refers to the disambiguated representation of LHS and makes sure that S indeed scopes over both X and Y – in such a case, the rule could be marked as strong.

So far we have only seen rules containing '+' and '$\circ$' in their polarity signatures. Two examples of rules containing '$-$' are (6)–(7).[9]



(6) $\langle$ U.L X.G to Y.L, U.L Y.L, $\langle -, -, \circ, \circ \rangle$, S, {LEMMA(X) $\in$ {FORGET, FAIL, ...}, ...} $\rangle$



(7) $\langle$ U.L X.G to Y.L, U.L Y.L, $\langle \circ, -, \circ, \circ \rangle$, S, {LEMMA(X) $\in$ {HESITATE, ...}, ...} $\rangle$

For simplicity, we assume that such rules are unidirectional, i.e., that the converse is not applicable (hence the final two '$\circ$' characters in polarity signatures). The former of these rules targets constructions headed by negative two-way implicatives such as FORGET or FAIL: saying that John forgot or failed to come implies that John didn't come, and saying that Mary didn't forget or fail to come implies the Mary indeed came. On the other hand, the latter rule is applicable only in negative polarity contexts: saying John didn't hesitate to come implies that John came, but saying that Mary hesitated to come doesn't imply whether Mary came or not.

Such rules illustrate a certain subtlety when it comes to the definition of context for the purpose of judging its polarity and to reversing the polarity of the output. Consider the following sentences as the intended input and output of rule (7):

[9]These rules are much simplified. Ellipsis signals the omission of the lemmata of other relevant implicatives and of other constraints, including those taking care of appropriate forms of verbs X and Y.

(8) Nobody hesitated to come.
(9) Everybody came.

The polarity context of the input is clearly negative, but this negative force does not come from the context in which the rule is applied, as it would in case of *It's not the case that John hesitated to come*, where negation is expressed outside the match of LHS, and also in case of *John didn't hesitate to come*, where negation is expressed in the unspecified part of the subtree headed by X. Instead, in (8), negation is introduced by the negative quantifier in the subject position, i.e. matched by the variable U of the LHS of this rule. Hence, the polarity of the context should be understood as the relative polarity of the head of the match, and not as the relative polarity of the whole match: an expression reversing the polarity may be part of the match, as in (9).

What the input–output pair (8)–(9) also illustrates is the convenience of understanding the polarity reversing effect of such rules rather broadly. The narrowest interpretation of this effect would simply add negation to the output, as in *Nobody didn't come*, and perhaps another rule would transform this sentence into (9). Similarly, perhaps the intermediate *I didn't not get high* could be used to get from (10) to (11). But for the purpose of this paper we assume that polarity reversal expressed by '−' is powerful enough to get in one step from the input to the output in both (8)–(9) (via rule (7)) and (10)–(11) (via rule (6)).

(10) I forgot not to get high.
(11) I got high.

Let us finally note that the simple picture where all rules are rewriting rules, replacing one subtree with another (perhaps empty, effectively pruning the tree), cannot be maintained. As an example, take the following pair (id=44) from the RTE-3: dataset:

(12) T: A former employee of the company, David Vance of South Portland, said Hooper spent a lot of time on the road, often meeting with customers between Portland and Kittery.
(13) H: David Vance lives in South Portland.

Here, the noun phrase (NP) *David Vance of South Portland* must be transformed into the sentence *David Vance lives in South Portland*. But of course the resulting sentence cannot simply replace the original NP – this would result in ungrammaticality. Neither can it simply replace the minimal sentence containing this NP: *It's not the case that David Vance of South Portland arrived* should not be replaced by *It's not the case that David Vance lives in South Portland*. Finally, the output of the rule should not replace the whole matrix sentence, as it may contain other similar bits of information to be used in the reasoning (e.g. *David Vance of South Portland met Hooper of North Dakota*). Instead, a second type of rules is needed which transform the input text by adding a new sentence. We currently assume that such rules are strictly unidirectional and are only used to extract conventional implicatures, which are independent of the polarity of context, hence, they may be formalised as quadruples ⟨LHS, RHS, S, C⟩, where RHS is interpreted as being added to the text rather than replacing LHS.

A simplified example of such a rule for the above pair is given in (14), and the complete reasoning – in Figure 5.

(14) ⟨ X.L of Y.L, X.L lives in Y.L . , *S*, {NAME(X), PLACE(Y)} ⟩

T:	A former employee of the company, David Vance of South Portland, said Hooper spent a lot of time on the road, often meeting with customers between Portland and Kittery.	
	→ A former employee of the company, David Vance of South Portland, said Hooper spent a lot of time on the road, often meeting with customers between Portland and Kittery. David Vance lives in South Portland.	(rule (14))
H:	→ David Vance lives in South Portland.	(dropping sentence)

Figure 5: Entailment steps from text T to hypothesis H for the RTE-3 pair id=44

4 Towards a taxonomy

Textual entailment rules are very diverse and involve various kinds of knowledge: lexical, syntactic, semantic, pragmatic and world knowledge, including the awareness of common scripts or frames (in the sense of Artificial Intelligence), so there is probably no single most natural way to organise them in a taxonomical hierarchy. Any attempt at such a taxonomy should be made with the view of its usefulness for the evaluation of semantic systems. For example, it makes sense to group rules reflecting phenomena such as coreference, zero anaphora and ellipsis, as a system which deals poorly on entailment pairs involving such rules probably lacks a reasonable discourse component. Similarly, grouping rules referring to lexical synonymy, hyperonymy, meronymy, etc., helps evaluate the use of wordnet-like knowledge sources by the system.

At the highest level, the proposed taxonomy distinguishes:

1. rules which do not refer to co(n)text, e.g. those discussed in §3 and almost all of the rules alluded to in Figures 1–2,
2. rules referring to the *cotext* of a given constructions, i.e. to bits of text not structurally related to this construction; e.g. to an earlier noun phrase (NP) coreferent with a given NP, as in the first rule of Figure 2, or to a bit of text corresponding to the missing part of an elliptical construction,
3. rules referring to the non-textual *context*, e.g. to the date of publication of the text, as in the textual entailment from *John Coltrane died in 1967* to *John Coltrane died almost 50 years ago* in case of texts published in 2015.

Within each class, rules are grouped according to how much lexical knowledge they require:

a. rules reflecting lexical equivalence relations: not only the usual synonymy (as in wordnets or thesauri), but also (near-)synonymous diathesis,
b. rules reflecting various kinds of lexical implication: hyperonymy, meronymy, etc.,
c. rules reflecting the implication of predicate dependents of factive and implicational verbs (as in Nairn *et al.* 2006),
d. constructional rules involving some lexical knowledge, e.g. rules allowing for the dropping of intersective or subsective modifiers (requires the knowledge of which lexemes may act as such modifiers),
e. constructional equivalence correspondences: variations in word order, distribution of dependents of a coordination over conjuncts (e.g. from *nice boy and girl* to *nice boy and nice girl* in languages in which the adjective on the left-hand side is in the plural and the adjectives on the right-hand side – in the singular), etc.; also rules which add new sentences on the basis of existing ones, including rules which extract relative clauses into standalone sentences (e.g. from *... Coltrane, who died in 1967,...* to *Coltrane died in 1967.*),

f. constructional implication relations: pruning of whole sentences, removing parentheticals, removing various kinds of modifiers, etc.

Additionally, a few groups of rules correspond to such subsystems of reasoning as temporal reasoning (e.g. from *Coltrane died on 17 July 1967* to *Coltrane died in July 1967*), numerical reasoning (further to *Coltrane died almost 50 years ago*; $2015 - 1967 = 48 = 50 - \epsilon$), etc.

At the time of submitting this paper to the proceedings, the complete taxonomy is not yet fully stable, so we do not provide it here. It should be emphasised, however, that while particular rules are often language-dependent, the taxonomy is constructed in a way that maximises its language-independence. This should facilitate the qualitative comparison of RTE systems for different languages.

5 Conclusion

The research reported here builds incrementally on a wide range of previous work in NLP (the RTE task) and linguistics, with the following novel features: 1) an extended and further formalised format of textual entailment rules, with multiple specific rules encoded, 2) a principled taxonomy of kinds of textual entailment (in place of previous more or less unordered collections of selected entailment types), 3) an entailment corpus, still at the initial stage of development, but already containing over 120 pairs fully tagged with almost 800 atomic entailment steps. We expect the corpus to be the most valuable result in practice, as it will not only lead to a qualitative evaluation of semantic NLP systems, but will also make it possible to construct training corpora – consisting of particular kinds of entailment steps – specialised for particular types of reasoning, as postulated already in Bentivogli *et al.* 2010.

References

Bentivogli, L., Dagan, I., Dang, H. T., Giampiccolo, D., and Magnini, B. (2009). The fifth PASCAL Recognizing Textual Entailment challenge. In *Proceedings of the Second Text Analysis Conference (TAC 2009)*.

Bentivogli, L., Cabrio, E., Dagan, I., Giampiccolo, D., Leggio, M. L., and Magnini, B. (2010). Building textual entailment specialized data sets: a methodology for isolating linguistic phenomena relevant to inference. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC 2010*, pp. 3542–3549, Valletta, Malta. ELRA.

Bos, J. (2008). Let's not argue about semantics. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC 2008*, pp. 2835–2840, Marrakech. ELRA.

Clark, P., Harrison, P., Thompson, J., Murray, W., Hobbs, J., and Fellbaum, C. (2007). On the role of lexical and world knowledge in RTE3. In *Proceedings of the ACL–PASCAL Workshop on Textual Entailment and Paraphrasing at ACL 2007*, pp. 54–59, Prague.

Cooper, R., Crouch, D., van Eijck, J., Fox, C., van Genabith, J., Jaspars, J., Kamp, H., Milward, D., Pinkal, M., Poesio, M., and Pulman, S. (1996). FraCaS: Using the framework. Deliverable D16, University of Edinburgh.

Dagan, I., Glickman, O., and Magnini, B. (2006). The PASCAL Recognising Textual Entailment challenge. In J. Quiñonero-Candela, I. Dagan, B. Magnini, and F. d'Alché Buc, eds., *Machine Learning Challenges: Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment. Proceedings of the First PASCAL Machine Learning Challenges Workshop (MLCW 2005)*, pp. 177–190, Berlin. Springer-Verlag.

Dagan, I., Roth, D., Sammons, M., and Zanzotto, F. M. (2013). *Recognizing Textual Entailment: Models and Applications*. Morgan & Claypool.

de Marneffe, M.-C., Dozat, T., adn Katri Haverinen, N. S., Ginter, F., Nivre, J., and Manning, C. (2014). Universal Stanford Dependencies: A cross-linguistic typology. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, eds., *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, pp. 4585–4592, Reykjavík, Iceland. ELRA.

Garoufi, K. (2007). *Towards a Better Understanding of Applied Textual Entailment: Annotation and Evaluation of the RTE-2 Dataset*. Master's thesis, Universität des Saarlandes.

Giampiccolo, D., Magnini, B., Dagan, I., and Dolan, B. (2007). The third PASCAL Recognizing Textual Entailment challenge. In *Proceedings of the ACL–PASCAL Workshop on Textual Entailment and Paraphrasing at ACL 2007*, pp. 1–9, Prague.

LoBue, P. and Yates, A. (2011). Types of common-sense knowledge needed for recognizing textual entailment. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 329–334, Portland, OR.

MacCartney, B. and Manning, C. D. (2007). Natural logic for textual inference. In *Proceedings of the ACL–PASCAL Workshop on Textual Entailment and Paraphrasing at ACL 2007*, pp. 193–200, Prague.

Mel'čuk, I. (1981). Meaning-Text Models: A recent trend in Soviet linguistics. *Annual Review of Anthropology*, **10**, 27–62.

Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. The SUNY Press, Albany, NY.

Nairn, R., Condoravdi, C., and Karttunen, L. (2006). Computing relative polarity for textual inference. In *Proceedings of the Fifth International Workshop on Inference in Computational Semantics (ICoS-5)*, Sydney, Australia.

Przepiórkowski, A. (2015). Towards a linguistically-oriented textual entailment test-suite for Polish based on the semantic syntax approach. *Cognitive Studies / Études Cognitives*, **15**. To appear.

Sammons, M. (2015). Recognizing textual entailment. In S. Lappin and C. Fox, eds., *The Handbook of Contemporary Semantic Theory*, pp. 523–558. Wiley-Blackwell, 2nd edition.

Sammons, M., Vydiswaran, V., and Roth, D. (2010). "Ask not what textual entailment can do for you...". In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1199–1208, Uppsala, Sweden.

Sgall, P., Hajičová, E., and Panevová, J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.

Szpektor, I., Shnarch, E., and Dagan, I. (2007). Instance-based evaluation of entailment rule acquisition. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 456–463, Prague.

Tesnière, L. (1959). *Éléments de Syntaxe Structurale*. Klincksieck, Paris.

Tesnière, L. (2015). *Elements of Structural Syntax*. John Benjamins, Amsterdam.

Toledo, A., Katrenko, S., Alexandropoulou, S., Klockmann, H., Stern, A., Dagan, I., and Winter, Y. (2012). Semantic annotation for textual entailment recognition. In I. Batyrshin and M. G. Mendoza, eds., *Advances in Computational Intelligence: 11th Mexican International Conference on Artificial Intelligence, MICAI 2012San Luis Potosí, Mexico, October 27 – November 4, 2012*, pp. 12–25. Springer-Verlag.

Zaenen, A., Karttunen, L., and Crouch, R. (2005). Local textual inference: Can it be defined or circumscribed? In *Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment*, pp. 31–36, Ann Arbor, Michigan.