ORIGINAL PAPER

An annotated corpus for the analysis of VP ellipsis

Johan Bos · Jennifer Spenader

Abstract   Verb Phrase Ellipsis (VPE) has been studied in great depth in theoretical linguistics, but empirical studies of VPE are rare. We extend the few previous corpus studies with an annotated corpus of VPE in all 25 sections of the Wall Street Journal corpus (WSJ) distributed with the Penn Treebank. We annotated the raw files using a stand-off annotation scheme that codes the auxiliary verb triggering the elided verb phrase, the start and end of the antecedent, the syntactic type of antecedent (VP, TV, NP, PP or AP), and the type of syntactic pattern between the source and target clauses of the VPE and its antecedent. We found 487 instances of VPE (including predicative ellipsis, antecedent-contained deletion, comparative constructions, and pseudo-gapping) plus 67 cases of related phenomena such as *do so* anaphora. Inter-annotator agreement was high, with a 0.97 average F-score for three annotators for one section of the WSJ. Our annotation is theory neutral, and has better coverage than earlier efforts that relied on automatic methods, e.g. simply searching the parsed version of the Penn Treebank for empty VP's achieves a high precision (0.95) but low recall (0.58) when compared with our manual annotation. The distribution of VPE source–target patterns deviates highly from the standard examples found in the theoretical linguistics literature on VPE, once more underlining the value of corpus studies. The resulting corpus will be useful for studying VPE phenomena as well as for evaluating natural language processing systems equipped with ellipsis resolution algorithms, and we propose evaluation measures for VPE detection and VPE antecedent selection. The stand-off annotation is freely available for research purposes.

Keywords   Ellipsis · Annotation · Evaluation · VP ellipsis

J. Bos · J. Spenader (✉)
University of Groningen, Groningen, The Netherlands
e-mail: j.spenader@ai.rug.nl

J. Bos
e-mail: johan.bos@rug.nl

1 Introduction

Ellipsis is a major topic in linguistics, yet there are no large, annotated corpora of ellipsis available. In this article we present an annotated corpus of verb phrase ellipsis (VPE) and related phenomena in English from the Wall Street Journal part of the Penn Treebank (Marcus et al. 1993).

VPE occurs when an auxiliary or modal verb abbreviates an entire verb phrase found elsewhere in the context, as in the following example:[1]

(1) <wsj_1996> How soon Wang will [$_{vp}$ stage a comeback], or if it **will** ~~stage a comeback~~ at all, are still matters of debate.

In example (1) the phrase *if it will at all* contains an elided verb phrase. Recovering the VP from the previous sentence, the full form would be *if it will stage a comeback at all*. Following the terminology introduced in Dalrymple et al. (1991), the clause containing the antecedent is termed the "source" clause (i.e., *Wang will stage a comeback*), and the clause with the elided verb phrase is called the "target" clause (*it will at all*). Phrases that are similar in structure and function occurring in both the source and target clause are called "parallel elements"; in (1) the subjects of the source and target clauses form parallel elements (*Wang* and *it*, respectively).

Analyzing VPE requires the use of contextual information to interpret its unexpressed meaning. Characterizing the constraints on when VPE is possible, and outlining how VPE examples should be interpreted is a challenging job, which is probably one of the main reasons why it has been such a prolific topic of research in formal syntax and semantics, as well as in computational linguistics. In fact, various aspects of VPE have been investigated in great detail: whether the level of resolution should take place on the syntactic level (Sag 1976; Hankamer and Sag 1976; Fiengo and May 1994; Merchant 2001), on the semantic level (Dalrymple et al. 1991; Hardt 1999a; Chierchia and McConnell Ginet 1991), or both (Kehler 1993), how VPE interacts with quantifier scope (Hirschbühler 1982; Shieber et al. 1996; Schiehlen 1999; Erk and Koller 2001) and rhetorical structure (Prüst 1992; Asher 1993; Kehler and Shieber 1997; Kehler 2002), and, especially, how to account for ambiguities resulting from the so-called sloppy and strict interpretations that occur when a VPE source clause contains anaphoric pronouns (Dahl 1973; Klein 1987; Dalrymple et al. 1991).

In contrast to the numerous theoretical studies, corpus studies on VPE are rare, as are implementations of theoretical VPE resolution algorithms in practical NLP applications. We believe that there are at least two reasons for this. First, from a purely practical perspective, automatically locating ellipsis and their antecedents is a hard task, not subsumed by ordinary natural language processing components. Recent empirical work (Hardt 1997; Nielsen 2005) indeed confirms that VPE identification is difficult. Second, most theoretical work begins at the point at which

---

[1] Throughout the paper we will mark the antecedent by enclosing it in square brackets, and only include syntactically reconstructed elided information (marked with a line through it) when this reconstruction involves simple copying and tense changes. In addition, the syntactic type of antecedent is indicated in subscript, and the VPE trigger is set in bold. A reference to the filename of the original Penn Treebank text is included in angled brackets.

the ellipsis example and the rough location of its antecedent are already identified, focussing on the resolution task. Moreover, the complex ellipsis phenomena described by theoretical work on ellipsis are not terribly frequent in ordinary texts such as newswire. Instead, there are many other linguistic phenomena associated with VPE requiring theoretical discussion and highlighting the more common problems is one of our aims.

We chose to semi-automatically annotate all the examples of VPE, the source and target clauses, as well as their syntactic forms, in the Wall Street Journal part of the Penn Treebank because this corpus is widely-used in natural language processing, and many other annotations are available for it. These include parse trees (Marcus et al. 1993), thematic roles, word senses and ontological links (Pradhan et al. 2007), co-reference relations (Weischedel and Brunstein 2005), discourse relation annotations based on Rhetorical Structure Theory (Carlson et al. 2001), discourse graphs in GraphBank (Wolf and Gibson 2005), and connective-based discourse annotation for the entire corpus in the Penn Discourse Treebank (Prasad et al. 2008). A thorough investigation of VPE in the Wall Street Journal texts is a much needed complement to these existing resources, and an ideal initial annotation project because of the possibility of utilizing existing resources together with the annotation in future work.

We believe that both theoretical approaches in formal semantics and practical work in computational linguistics can benefit from the results we present here, which is why we have made the stand-off annotation freely available for research purposes.[2] We hope our annotation will provide a solid basis for developing language technology to recognize and resolve ellipsis phenomena, as well as highlighting a number of frequent features of corpus ellipsis examples that might inform and inspire theoretical work.

The remainder of this article is organized as follows. In Sect. 2 we present earlier work that has been done on ellipsis annotation. Section 3 introduces a number of varieties of VPE and several related phenomena, and explains what we included and excluded in our annotation and why. In Sect. 4 we introduce our method of annotation, how we identified antecedents, and the different syntactic features that we included. Here we also present information about the reliability of the annotation, for which we compared inter-annotator agreement and also compared our results with existing annotations to the degree this was possible. Section 5 presents our results in the form of frequency information about the different types, highlighting a number of unexpected findings. Section 6 presents guidelines for evaluating ellipsis processing algorithms using our annotated data. Finally, Sect. 7 discusses how the results relate to theoretical work, how they could be used to develop automatic parsers for VPE, and gives suggestions for future work.

---

[2] The VPE stand-off annotation is available from http://www.let.rug.nl/bos/vpe/ and is also distributed with the C&C tools (Curran et al. 2007). The raw texts of the 25 sections of the Wall Street Journal are part of the Penn Treebank and are distributed via the Linguistic Data Consortium (LDC), http://www.ldc.upenn.edu.

2 Related work

Two earlier studies have annotated corpus examples of VPE to develop automatic resolution systems: Hardt (1997) and Nielsen (2005). We briefly review them here and discuss their merits as well as their shortcomings.

Hardt (1997) automatically found 644 examples of VPE in the parsed corpora included in the Penn Treebank (Marcus et al. 1993) and then manually annotated their source clauses. He found 260 instances of VPE in the Wall Street Journal (WSJ) and 384 in the Brown corpus. In the WSJ part of the treebank, VPE were located by searching for "empty VP" patterns in the gold-standard parse trees. In the Brown part of the treebank, which comes with a different labelling, VPE were detected by searching for sentences with an auxiliary but no main verb. We applied Hardt's method to the WSJ part of the corpus and found good precision paired with a relatively low recall (see Sect. 5). The low recall is caused by the absence of marking of empty VPs for many elliptical cases in the Penn Treebank. A case in point is (2), which isn't coded as an empty VP in the Penn Treebank, even though it contains an elided VP.

(2) <wsj_1391> China might stave off a crisis if it [$_{vp}$ acts as forcefully] as it **did** to arrest the 1985 decline, when Beijing slammed the brakes on foreign-exchange spending and devalued the currency.

By manually checking a small subset of the two corpora Hardt was able to estimate that his method yields a recall of 44% and a precision of 53% for finding VPE instances. Put differently, Hardt's method found roughly half of all VPE present in both corpora. Recall that Hardt used two different methods for automatically detecting VPE in the treebank: one for the WSJ sections and another one for the Brown corpus. As we will show in Sect. 5, the method applied only to the WSJ part of the treebank performs much better than these figures actually do suggest. Nevertheless, it is likely that the form of VPE instances found with Hardt's method differ significantly from those that were not found, given that the parser was not able to identify them as having an empty VP. As a consequence, factors found to be relevant for VPE identification and resolution based on these examples might not be optimal for all VPE instances. To have a clear picture of all forms of VPE, a more exhaustive corpus study is required. Additionally, Hardt's annotations haven't been made available, to the best of our knowledge.

Nielsen (2005) manually annotated VP Ellipsis on subsets of three different corpora: He used 14 sections of the British National Corpus (BNC) (444,000 words, 843 examples of VPE), seven sections of the WSJ (e.g. Sections 00, 01, 02, 03, 04, 10, 15, totalling 118 examples of VPE), and eight sections of the Brown corpus (513 examples of VPE). In total, he annotated ca. 1,500 instances of VPE. Nielsen noted that the rate of VPE differed substantially between the WSJ texts and the BNC and Brown texts. Nielsen found one VPE for every 77 sentences in the WSJ texts, which was approximately half the rate of VPE in the BNC and Brown texts.[3] The number of examples Nielsen (2005) has annotated is impressive, and Nielsen has made a

---

[3] Nielsen attributes his observation to the use of VPE being discouraged in journalistic writing.

stand-off version of these annotations available online.[4] However, for at least the WSJ files, the annotation is not easily reusable because it has been carried out on tokenized and parsed data, rather than on the texts in the form they are distributed by the LDC. This means that reusing the WSJ annotations requires re-identifying each example in the raw texts.

These two earlier studies differed as to what related phenomena they included. Choices related in part to the theory of VPE advocated by the researchers. Generally speaking, approaches to VPE can be roughly classified into two types, syntactic approaches or semantic approaches. Syntactic approaches have attempted to resolve VPE by using different copying mechanisms or syntactic reconstruction mechanisms, as there seem to be syntactic constraints on VPE. Nielsen (2005) believes ellipsis interpretation is mediated by syntax. However, Hardt (1999b) and others have shown that it is possible to have complex mismatches between the source clause and the target clause, including voice alterations and nominal antecedents. It is not clear how a reconstruction or syntax based approach could deal with the mismatches. These mismatches all suggest that a semantic approach is more appropriate. Hardt (1997) takes such a semantic approach. Whether or not ellipsis is treated as semantic or syntactic is irrelevant to the identification of standard VPE, but it does in part account for the differences in included examples between the two earlier annotation efforts.

In our work we aim to be theory-neutral and separate ellipsis identification from possible resolution algorithms. We marked up all 25 sections of the WSJ. The annotation was carried out semi-automatically on the raw text files. More than 500 instances of VPE and closely related phenomena with their respective antecedents were annotated (compared to 118 of Nielsen and 260 of Hardt in the WSJ part of their annotated corpora). In addition to annotating the antecedent and its syntactic form we also annotated the syntactic pattern connecting the source and target clause of each instance of VPE. All files and annotations were manually checked at least twice. The resulting annotation is publicly available in stand-off format. With respect to precision, recall, coverage, usability, and detail of annotation, this corpus is a considerable improvement on previous annotation efforts.

3 Selected phenomena

In this section we discuss the boundaries of our annotation efforts, and define which elliptical phenomena are annotated and which aren't. The technical aspects of annotation and the exact manner in which we annotate the selected phenomena will be presented in Sect. 4.

Our main concern will be classical VPE, where the target clause is reduced to one of the three auxiliary verbs *do* (3), *have* (4), or *be* (5), or a modal verb, as in example (6). We also include VPE triggered by the infinitival *to* (7). Table 1 lists all verb forms that can license VPE considered in our study.

---

[4] http://sourceforge.net/projects/vperesolver/.

Table 1 Auxiliary and modal verbs considered as VPE triggers

Type	Instances
aux	do don't does doesn't did didn't done doing
aux	am 'm are aren't ain't is 's isn't was wasn't were 're weren't be been
aux	have 've haven't has 's hasn't had 'd hadn't
modal	can cannot can't
modal	could couldn't
modal	may mayn't
modal	must mustn't
modal	might mightn't
modal	will won't
modal	would, wouldn't
modal	shall shan't
modal	should shouldn't
semi-modal	need needn't
semi-modal	dare daren't
semi-modal	ought oughtn't
infinitival	to

(3) <wsj_0036> The government [$_{vp}$ includes money spent on residential renovation]; Dodge **does**n't ~~include money spent on residential renovation~~.

(4) <wsj_0037> Mr. Katzenstein certainly would have [$_{vp}$ learned something], and it's even possible Mr. Morita would **have** ~~learned something~~ too.

(5) <wsj_0018> Not only is development of the new company's initial machine [$_{vp}$ tied directly to Mr. Cray], so **is** its balance sheet.

(6) <wsj_0585> On days when prices are tumbling, they must be willing to [$_{vp}$ buy shares from sellers] when no one else **will** ~~buy shares from sellers~~.

(7) <wsj_0039> He spends his days [$_{vp}$ sketching passers-by], or trying **to** ~~sketch passers-by~~.

Perhaps surprisingly, given the large body of literature on VPE, it is hard to find a crystal-clear definition of VPE. Given our aims of annotating a large corpus with occurrences of VPE, we need to be as precise as possible in choosing what will be subject to annotation and what won't. This isn't always easy: sometimes elliptical constructions have a VP as antecedent but we don't consider it VPE, and sometimes elliptical constructions look like VPE but don't have a VP as antecedent (such as predicative ellipsis, see below). To deal with this complex classification problem in a systematic way, we have tried to make a selection by taking prototypical VPE examples as our starting point, and including and excluding closely related phenomena based on choices that we believe are useful and related to earlier annotation work.

We leave out "gapping", which occurs when a main verb in the source clause is not repeated in the target clause and at least two constituents, none of them an auxiliary or modal verb, remain in the target clause, as in *Digital Equipment*

dropped 2 1/4 to 88 and Unisys 1/2 to 17. We also leave out "stripping", which occurs when the target clause contains only one constituent, usually accompanied by a discourse particle, as in *Everyone was disappointed. But not Richard Cottrel*. Finally, "sluicing" is not covered, which occurs when the target clause contains only an embedded WH- phrase and an inflectional phrase (a verb phrase with a finite form) is elided, as in *Apparently, we lost a lot of the Tiger votes. I don't know why*. These elliptical phenomena are, unlike VPE, specific to certain syntactic constructions such as coordination (Schiehlen 1999).

On the other hand, we include cases of predicative ellipsis, antecedent-contained deletion, comparative and equative deletion, pseudo-gapping, comparative subdeletion, subject-auxiliary inversion cases, and *do so* anaphora, and we describe these phenomena in detail below.

3.1 Predicative ellipsis

Some elliptical constructions resemble standard cases of VPE when just looking at the target clause, as in examples (8–10) below. Closer inspection reveals that the syntactic material that is deleted does not cover a full VP, but rather an adjectival phrase (8), a noun phrase (9), or prepositional phrase (10):

(8) <wsj_1146> The farmers stayed [$_{ap}$ angry]. They still **are** ~~angry~~.
(9) <wsj_0561> The ball he hit wasn't [$_{np}$ a strike]. If it had **been** ~~a strike,~~ he mighta hit it out.
(10) <wsj_0515> For just as the Arabs **were** ~~on the brink of global power and fame~~ in the 1960s, the farmers of Sidhpur are [$_{pp}$ on the brink of global power and fame].

Hence, strictly speaking, these aren't instances of VPE, even though in the VPE literature similar examples are mentioned together with other cases of VPE.[5] We will however include these cases of predicative ellipsis in our annotation efforts; the annotation scheme that we use allows us to separate these cases of predicative ellipsis from ordinary forms of VPE. Note that we have only found occurrences of predicative forms with the auxiliary *be* for these examples, and they also occur frequently in comparative or equative constructions (see below).

3.2 Antecedent-contained deletion

Usually, the source and target clauses of an ellipsis construction are different. In what is dubbed *antecedent-contained deletion* in the literature, the target clause is embedded in the source clause, by virtue of a relative clause (Bouton 1970; Sag 1976; May 1985). Consider example (11):

---

[5] For instance, in the context of VPE, Dalrymple et al. (1991), on p. 408, discuss their example (17) which seems to have an elided predicative adjective:

(17) It is obvious that Dan is happy, and George is too. Prüst (1992), on p. 84, presents the example (3–42) where a PP seems to have been elided:

(3–42) John is from Brazil (and) Bill is too.

(11) `<wsj_0433>` Do you think the British [tv know] something that we **do**n't ~~know~~?

The target clause in (11) is *we don't*, which is part of the VP in the source clause, to wit *know something that we don't*. On the surface level, the deleted material is not a VP but just the transitive verb *know*. Note that, in this example, the NP interpretation of *something* is shared between the source and target clause, which can be paraphrased as "something that the British know and we don't know."

3.3 Comparative and equative constructions

Comparative constructions often go hand-in-hand with ellipsis, which is usually referred to as *comparative deletion*. Comparative deletion covers a wide range of elliptical constructions. In our annotation efforts we only considered cases of comparative deletion involving elided verb phrases, where the target clause contains an auxiliary or modal verb as illustrated by (12). We also include the equative construction, as shown in (13).

(12) `<wsj_0445>` Moreover, Japanese offices tend to [vp use computers less efficiently] than American offices **do** ~~use computers efficiently~~.
(13) `<wsj_0456>` He did not [vp go as far] as he could **have** ~~gone far~~ in tax reductions; indeed he combined them with increases in indirect taxes.

3.4 Pseudo-gapping

In pseudo-gapping an auxiliary verb replaces a full verb form, as in VPE. But pseudo-gapping differs from VPE in that verbal arguments are retained, as in ordinary gapping. As a result, the syntactic category of the elided phrase corresponds to a transitive verb, as in the following example:

(14) `<wsj_0664>` He said traders should be on the lookout for how metals producers react to this rally. "I expect to see some selling, but will they [tv kill] this one as they **have** ~~killed~~ every rally in the recent past" by selling and locking in prices for their production?

Miller (1990) and Hoeksema (2006) argue that pseudo-gapping is best analyzed as a pro-form similar to the analysis for VPE given by Hardt (1999a). However, Hardt does not include pseudo-gapping constructions in his work because he does not consider them to be pro-forms. This is because the examples lack the typical pro-form features, that is, pseudo-gapping constructions do not allow cataphora, cannot be relativized, and do not allow an arbitrary amount of material to intervene between the source clause and the target clause (Hardt 1999a). Instead, Hardt considers pseudo-gapping, along with gapping and stripping, to be conjunction forms. These conjunction forms then behave the way they do because they require access to syntactic structures, while as a true pro-form, VPE instead only needs to refer to the discourse model. Nielsen (2005) included pseudo-gapping in his study of VPE because he believes that it requires syntactic reconstruction. Thus Hardt (1997)

excludes pseudo-gapping from his annotation for the same reason Nielsen (2005) includes it. As for our annotation efforts, we have included cases of pseudo-gapping, even though they turned out to be a rare phenomenon in WSJ newspaper texts.

3.5 Comparative subdeletion

Similar to pseudo-gapping are cases of comparative subdeletion (Bresnan 1973). Comparative subdeletion structures, as exemplified by (15), are the same as pseudo-gapping constructions in that they also elide a main verb but retain an auxiliary and an object argument. They differ from pseudo-gapping in that they always occur in comparative or equative constructions, and, moreover, they seem to be more widely applicable (see Bowers (1998), for examples).

(15) <wsj_2369> Don't say the TV sitcom, because that happens to be a genre that, in its desperate need to attract everybody and offend nobody, [$_{tv}$ resembles] politics more than it **does** ~~resemble~~ comedy.

We included comparative subdeletion structures in the corpus mark-up, but we did not distinguish between pseudo-gapping and comparative subdeletion explicitly. Neither Hardt (1997) nor Nielsen (2005) mention comparative subdeletion structures.

3.6 Subject-auxiliary inversion

Certain constructions involving VPE license an inversed order of the subject and auxiliary. There are five cases that we can distinguish here: subordinated clauses starting with *as* (16); pre-verbal *so* (17); comparative (and equative) structures (18); clauses headed by *neither* or *nor* (19); and tag questions (20).

(16) <wsj_0114> His wife also [$_{vp}$ works for the paper], as **did** his father.
(17) <wsj_1267> Someone with a master's degree in classical arts who works in a deli would [$_{vp}$ be ideal], Litigation Sciences advises. So **would** someone recently divorced or widowed.
(18) <wsj_1280> One analyst noted that the company often [$_{vp}$ has better store locations] than **do** its franchisees, thus aiding promotional efforts.
(19) <wsj_0239> See, the other rule of thumb about ballooning is that you can't [$_{vp}$ steer]. And neither **can** your pilot ~~steer~~.
(20) <wsj_1618> But you [$_{vp}$ knew that], **did**n't you ~~know that~~?

In some of these cases, to wit (16–18), the elliptical material cannot be easily reconstructed syntactically to yield a grammatical sentence. This observation applies to subject-auxiliary inversion accompanied by the adverbs *as* and *so*, and also to the comparative construction. Notable exceptions to this rule are when the subject-auxiliary inversion is part of a *neither/nor* clause or tag question; then syntactic reconstruction seems possible, as examples (19) and (20) illustrate.

3.7 So, likewise, same, opposite

A number of phenomena have VP antecedents and seem to share many of the characteristics of VPE. This is particularly true for a group of constructions that is sometimes termed *do X* anaphora (Culicover and Jackendoff 2005). Constructions formed by a combination of *do* and *so* are a particularly frequent subtype and seem at first sight genuine instances of VPE—they generally have a VP as antecedent and seem to use anaphoric *do*—yet there are several differences compared to the examples standardly classified as VPE. Consider example (21):

(21) <wsj_0041> Mr. Wilder did [_{vp} introduce such legislation 17 years ago], but he did **so** at the request of a constituent, a common legislative technique used by lawmakers.

A *do so* construction like (21) differs from standard VPE examples like (3), (11) and (12) in that the *do* in (21) has the characteristics of a main, transitive, verb (Hankamer and Sag 1976; Sag 1976; Miller 1990; Hardt 1997). This claim is backed up by the observations that the *do so* construction doesn't permit subject inversion in a question and cannot be negated like ordinary auxiliary verbs can. In addition, *do so* constructions can only refer to actions (Culicover and Jackendoff 2005), unlike VPE, which can refer to states such as *knowing*. For these reasons, the construction is often regarded as a phenomenon distinct from VPE.

In previous VPE annotation efforts, Hardt (1997) excluded *do X* forms like *do so*, considering them to be simply a very common variant of other *do X* forms, such as *do this*, *do the work*, and so on. Nielsen (2005) also excluded *do so* and the related form *so doing* because they do not require syntactic reconstruction.

Nonetheless, we actually did annotate a number of *do X* types, including in particular *do so*, in part because it is so frequent, and in part because of the many characteristics it shares with VPE. Like VPE, *so-anaphora* also allow mismatches between antecedent and target such as voice alterations, nominalizations, and split antecedents (Kehler and Ward 1999; Ward and Kehler 2005). Actually, these are the same characteristics that are used by e.g. Hardt (1997) to argue that VPE must be semantically treated. We also include *do the same* (22), *do the opposite* (23), and *do likewise* (24), but we clearly distinguish these phenomena from ordinary VPE in the annotation scheme that we propose.[6]

(22) <wsj_0331> Its plan, instead, is to [_{vp} spin off the remainder] of its real estate unit and to possibly do **the same** with its mining and energy assets.

(23) <wsj_1761> "[_{vp} Sell stocks that aren't doing well now, and that don't have good earnings prospects]," says Alfred Goldman, technical analyst at St. Louis-based A.G. Edwards & Sons. "Most people do just **the opposite**: They sell their winners and keep their losers."

(24) <wsj_1092> Americans [_{vp} place native or native speakers in charge of subsidiaries overseas]. European multinationals do **likewise**; even in America, their affiliates are usually run by American managers.

---

[6] As a result, we are able to distinguish the *do so* construction from the pre-verbal *so* construction mentioned in Sect. 3.6.

We have excluded *do it* and *do that*, where *it* or *that* can be interpreted as an abstract object anaphor, rather than as part of a VP anaphora. Certainly, it is not clear to what extent subgroups of *do X* constructions are similar to each other so a different selection could have been made, but we were not able to exhaustively annotate all phenomena that can take a VP antecedent. As a consequence, many potentially interesting examples (such as (25) where *do either* takes a conjoined VP antecedent that gets interpreted at the ellipsis site as two disjoint events), were excluded in part for practical reasons.

(25) <wsj_0858> One possibility is more active American recruitment of rebels who would [$_{vp}$ agree to hand Noriega to the U.S.] and [$_{vp}$ install the elected leaders he rebuffed]. The last coup plotters refused to do **either**.

3.8 Idiomatic occurrences of VPE

While annotating instances of VPE we stumbled upon a number of cases that could be classified as idiomatic expressions, as they seem to adhere to a non-productive syntactic structure. A case in point is the idiom in example (26), where the phrase "try as they might" has no literal meaning, but rather expresses a sense of regret or failure.

(26) <wsj_1146> [$_{vp}$ Try] as they **might** ~~try~~, the Communists could neither replace nor break him.

Idiomatic usage of VPE seems to go hand-in-hand with the use of a pronoun in the subject position of the target clause (Pullum 2000). Other cases are less clearcut. To avoid excluding non-idiomatic instances of VPE from the annotation, we treated all idiomatic suspects as if they were ordinary VPE, and included them in the annotation without explicitly marking them as such.

4 Annotation method

In this section we first introduce the annotation scheme, the tools we used to do the mark up, and the guidelines we used to determine what should be included in the antecedent. We also present a typology we used for identifying the ellipsis type, the syntactic type of the antecedent, and the relations between the source and target clause. Finally, we discuss inter-annotator agreement.

The corpus we used for annotation comprises the complete Wall Street Journal part (all 25 sections) of the Penn Treebank (Marcus et al. 1993). We took the raw files as the basis for stand-off annotation, which was carried out in three stages: the first stage proposes auxiliary and modal verbs that potentially trigger VPE; the second stage selects the antecedent for each annotated VPE in the previous stage; the third and final stage adds type information to all selected VPE with antecedent. Table 1 shows all morphological variants (including contractions) for the auxiliary, modal, and semi-modal verbs considered as triggers for VPE in the first stage of annotation.

4.1 Annotation scheme

We used stand-off annotation to record occurrences of VPE and their antecedents. This type of annotation has the advantage that it does not modify the raw files of the corpus; instead the annotations and references to the original texts are collected in a separate file. This ensures that combining multiple layers of annotation is a possibility, and it makes the annotation completely independent from further processing such as tokenization, chunking, parsing or other computational tasks. For example, the stand-off annotation could be used to insert XML tags in the raw text, which could then be tokenized while preserving the XML annotation, finally producing a new stand-off annotation on the basis of the tokenized data from the results.

Annotation of the VPE and corresponding antecedent were done at the character level. We adopted a simple annotation scheme to markup the VPE and its antecedent. To illustrate this scheme, consider the following excerpt of Section 01 of the WSJ, namely file raw/wsj/01/wsj_0112, with line numbers added in an additional initial column for convenience:

```
   01234567890123456789012345678901234567890123456789012345678901234567890123
70
71 As it stands now, Congress presumes after the Reagan administration ...
72 President Bush should set things straight.
73 If he does not, he will help realize Madison's fear in The Federalist ...
74
   01234567890123456789012345678901234567890123456789012345678901234567890123
```

In this textual fragment line 73 contains an elliptical VP (the phrase *If he does not*), and the antecedent (*set things straight*) is found in the previous line. The VPE is marked by the character positions of the auxiliary verb that licenses it. Here it's the word *does*, which starts at position 7 and ends at position 10; the corresponding character offsets in the raw file are 9215 and 9219. The antecedent is at line 72, starting at character position 23 and finishing at position 41, corresponding to offsets 9188–9207. Given this information, the stand-off annotation is straightforward, and is just the sequence of file name, VPE (start, end), and antecedent (start, end), all printed on one line and white space marking boundaries:

<div align="center">

`wsj_0112 9215 9219 9188 9207`

</div>

Note however that this straightforward way of marking VPE and antecedents does not permit a proper annotation of discontinuous phrases. This won't be a problem for the auxiliary or modal verb, as English grammar doesn't possess such discontinuous verbs. However, it could be a problem for discontinuous antecedents. As a matter of fact, Prüst (1992) and Asher (1993) argue that discontinuous VPs can act as split antecedents for VPE. Comparative deletion and antecedent-contained deletion can also give rise to discontinuous antecedents. But we believe that the actual number of cases is relatively low and does not warrant complicating the annotation scheme by incorporating discontinuous structures in the annotation.

4.2 Annotation tools

To simplify the process of annotation, we developed a series of minimalistic but efficient tools to assist the annotator. All tools were written in SWI Prolog with simple but effective keyboard-based shell interfaces, not requiring a graphical user-interface or mouse. This made the annotation process fast. Each tool corresponds to one of the three phases of annotation as outlined above. Coding the initial versions of the tools themselves took no longer than a day, and they were improved during the annotation efforts taking comments and requests of the annotators into account.

When given a text file, the first tool interactively asks for each occurrence of potential VPE trigger (all those listed in Table 1) in the texts whether it's classified as a VPE or not. For each of these queries the annotator just has to respond with a single keystroke to confirm or discard VPE candidates, resulting in a high-speed annotation process. On average it took about 2–3 h for an experienced annotator to work through one section of the WSJ for selecting occurrences of VPE. A second tool takes the annotated VPE found in the previous stage and adds the antecedent information. This was done by showing the line in which the VPE occurs and the previous (non-empty) line of the raw text.[7] In all but one case, a cascaded ellipsis[8] shown below in (27), this was sufficient context to select the antecedent.[9]

(27) `<wsj_1286>`
Why can't we [$_{vp}$ teach our children to read, write and reckon]?
It's not that we don't know how to, because we do.
It's that we don't want to.
And the reason we don't want **to** ~~teach our children to read, write and reckon~~ is that effective education would require us to relinquish some cherished metaphysical beliefs about human nature in general and the human nature of young people in particular, as well as to violate some cherished vested interests.

Antecedents were selected by identifying the start and end position using keystroke combinations while they were highlighted on screen. We thought this would be the most time-consuming annotation activity, but in fact, it turned out that it was even faster than identifying VPE in the previous annotation stage.

The third tool records additional, typological, information about previously selected VPE: the syntactic type of antecedent, and the syntactic environment of the source and target clause. This information is normalized into a predefined set of codes (see Sect. 4.4 for details) and simply added in extra columns to the stand-off annotation

---

[7] Note that we talk about *lines* here, not sentences or clauses. In the raw WSJ files, sentences do not always correspond to lines, and running text may contain blank lines as well.

[8] A cascaded ellipsis is defined in Dalrymple et al. (1991, p. 418) as a sentence "containing multiple elliptical clauses in which the interpretation of one elided constituent depends partially or entirely on the interpretation of another elided constituent". In example (27) there are four sentences, containing a series of four VPEs. It is debatable whether the interpretation of the last VPE depends on the third elided clause or on the textual explicit antecedent in the first sentence. In our annotation work we have chosen for the latter option, primarily for reasons of simplicity.

[9] This exceptional case was added manually by directly editing the stand-off annotation file.

file. An additional shell script generates HTML-code of the stand-off annotation where VPE instances and their antecedents taken from the original text files are clearly identified in the surrounding context and sorted by type or WSJ section. This allowed the annotators to check and revise their annotations when necessary.

4.3 Annotation guidelines for selecting antecedents

The context often provides more than one possible antecedent for a VPE: the source clause doesn't always immediately precede the target clause (even though this is usually the case), and the source clause might contain more than one possible VP that can serve as antecedent. In particular, cases of embedded VPs cause potential ambiguities, as the following example shows:

(28) `<wsj_1367>`
 The Volokhs were afraid that they'd end up like a friend of theirs who'd applied for a visa and [$_{vp}$ waited for 10 years], having been demoted from his profession of theoretical mathematician to shipping clerk. They **did**n't ~~wait for 10 year~~s.

(29) `<wsj_1367>`
 The Volokhs were afraid that they'd end up like a friend of theirs who'd [$_{vp}$ applied for a visa and waited for 10 years], having been demoted from his profession of theoretical mathematician to shipping clerk. They **did**n't ~~apply for a visa and waited for 10 year~~s.

(30) `<wsj_1367>`
 The Volokhs were afraid that they'd [$_{vp}$ end up like a friend of theirs who'd applied for a visa and waited for 10 years], having been demoted from his profession of theoretical mathematician to shipping clerk. They **did**n't ~~end up like a friend of theirs who'd applied for a visa and waite~~d ~~for 10 year~~s.

 In the above example it is, arguably, unclear whether the antecedent is properly identified in (28), (29), or (30). We have chosen to annotate the widest possible antecedent if the context doesn't clearly disambiguate, i.e. (30). This is an arbitrary decision, we could have as well as adopted the method of picking the smallest.

 Overall, in order to make the annotation consistent and promote inter-annotator agreement, we developed the following annotation guidelines for selecting antecedents:

1. Mark antecedents as they would appear in the non-elliptical variant of the same clause (to the degree that this is possible);
2. Do not include punctuation symbols or spaces at the start or end of an antecedent (except for quoted speech);
3. Exclude parallel elements when they occur at the start or end of an antecedent;
4. Mark antecedents as wide as possible in vague or ambiguous cases.

 Note that some antecedents are not themselves VPs. Also, in many cases a difference in polarity between source and target does not allow literal rephrasing of

negative polarity items. The same holds for verbal morphology introduced by tense. Perfect syntactic reconstruction isn't possible in such cases, but this wasn't considered an obstacle for annotating the intended antecedent.

Further note that we do not mark ambiguities caused by scope or pronouns, a frequent topic in the theoretical literature on VPE. This is because the identification of the antecedent itself doesn't disambiguate between different possible readings, thus for annotation of the surface antecedent itself the issue of ambiguity encoding is unimportant. Including such ambiguities in the annotation obviously requires a richer annotation scheme, including perhaps proposing a resolved version of the VPE. This goes beyond the scope of our current annotation efforts.

We defined additional guidelines for annotating the antecedent of VPE occurring with comparatives and equative constructions. Because they give rise to various complications, they are described in detail in the following subsections.

4.3.1 Comparative constructions

We found comparative constructions both with genuine VPE and with predicative ellipsis cases. English grammar exhibits quite a large variety of syntactic constructions to express comparisons, and this considerably complicates the annotation task when VPE are involved. The comparative construction can be generalized over the following patterns (this is not meant to be an exhaustive list licensed by English grammar, but is rather based on patterns found in the WSJ corpus):

$[_{vp}$ VP ADV-er]	than	BE less|more $[_{ap}$ AP]	than
$[_{vp}$ VP less|more ADV]	than	BE $[_{ap}$ AP-er]	than
$[_{vp}$ TV less|more NP]	than	BE $[_{ap}$ ADV-er AP]	than
$[_{vp}$ VP] less|more	than	BE $[_{ap}$ ADJ-er NP]	than
$[_{vp}$ TV ADJ-er NP]	than	BE $[_{pp}$ PREP a ADJ-er NP] than	

In addition, comparatives can be modified by intensifiers (such as *far*, *much*, *significantly*, *so much*, *a great deal*) and measure phrases identifying the difference of comparison (for instance, *at least 15 min*, *10 points*, and so on). In annotating the antecedent, the comparative construction adverb *than* is excluded, and so are the operators *less* and *more*, intensifiers, and measure phrases—but only if they are at the start or end position of the selected antecedent. Consider the following examples with annotations following the proposed guidelines:

(31) <wsj_0445> Moreover, Japanese offices tend to $[_{vp}$ use computers less efficiently] than American offices **do** ~~use computers efficiently~~.

(32) <wsj_0071> But consumers who buy at this level are also more $[_{ap}$ knowledgeable] than they **were** ~~knowledgeable~~ a few years ago.

(33) <wsj_0795> He said: "Inflation is $[_{ap}$ lower] than I think people expected it to **be** ~~low~~, and I think that's good news."

In (31), the token *less* is marked as part of the antecedent because it doesn't occur at the start of end of the antecedent, whereas in (32), *more* isn't marked as part of the

antecedent. Finally, if the comparative form is morphologically expressed as in (33), we include it as well (we don't annotate beyond the token level).

This annotation convention has the added advantage of making the antecedent annotation of standard VPE and predicative ellipsis uniform; in predicative ellipsis in comparative constructions, e.g. (32), the property compared *must* be treated as the antecedent in order to have an antecedent to identify.

4.3.2 Equative constructions

As with comparatives, we found equative constructions both with genuine VPE and with predicative ellipsis cases. The equative construction can be generalized over the following patterns (again, this is not an exhaustive list, but rather based on patterns found in the WSJ corpus):

$$[_{vp} \text{ TV as ADJ NP}] \quad \text{as} \quad \text{BE as } [_{vp} \text{ much VP}] \text{ as}$$
$$[_{vp} \text{ TV as much}] \quad \text{as} \quad \text{BE so } [_{vp} \text{ much VP}] \text{ as}$$
$$[_{vp} \text{ VP as ADJ}] \quad \text{as} \quad \text{BE as } [_{ap} \text{ ADJ}] \quad \text{as}$$
$$[_{vp} \text{ VP as much VP-ed}] \text{ as} \quad \text{BE as } [_{ap} \text{ ADJ NP}] \text{ as}$$

In marking the antecedent of a VPE occurrence in an equative construction, the second occurrence of *as* is not included. As with comparatives, intensifiers are excluded when they occur at the start or end of the antecedent. An example of this kind is shown below in (34):

(34) `<wsj_0114>` Many felt Hearst [$_{vp}$ kept the paper alive as long] as it **did**, if marginally, because of its place in family history.

4.4 A typology for VPE annotation

After VPE identification and antecedent annotation we also annotated the type of ellipsis trigger, the syntactic type of the antecedent, and the pattern between the source and the target clauses. This information is all added in the same stand-off annotation file, by adding new columns of information. Because this data lies outside the core annotation efforts, no inter-annotator agreement tests were performed. Instead, the task of adding this information was split between two annotators and subsequently verified and corrected.

4.4.1 Type of trigger

For each instance of VPE encountered in the corpus we added the type of trigger in a separate column in the raw annotation file. For the auxiliary and modal verbs the types are based on the corresponding lemmata. These are:[10]

---

[10] Semi-modals like *need* or *dare* were not found in the corpus in connection with VPE so we ignore them here.

```
do        can        may        will     shall       to
be        could      must       would    should
have                 might
```

For the cases of *so*-anaphora and related *do X* anaphoric phenomena (as discussed in Sect. 3.7) we used the following trigger types:

```
        so        same       likewise       opposite
```

Admittedly, some information will be lost by strictly following this scheme, because not all of these occur with the verb *do*, as (35) below demonstrates.

(35)   `<wsj_1578>` Among other stocks [_{vp} involved in restructurings] or rumored to **be** so: Holiday Corp. gained 1 7/8 to 73 and Honeywell rose 2 7/8 to 81 1/2.

However, this is rather an exception than the rule. On the other hand, it also demonstrates that a more extensive annotation project recording all aspects of *so*-anaphora and related phenomena might require a richer annotation scheme.

4.4.2 Syntactic type of antecedent

The majority of the elliptical phenomena considered in our study can be classified as VPE. However, some cases which seem to be clear examples of elided VPs have antecedents that correspond to categories other than VP, in particular when the auxiliary is a form of *to be*, discussed in Sect. 3.1 as predicative ellipsis. To distinguish these from ordinary VPE, we introduce the syntactic type to the annotation scheme. In the normal case this type is VP. In addition we have AP (8), NP (9), and PP (10) to cover the three different cases of predicative ellipsis.

Further, we also introduce the type TV (transitive verb) to mark up cases of pseudo-gapping (14) and antecedent-contained deletion (11) because in these cases only the transitive verb is elided. Even though we use the term *transitive verb* here, we do not necessarily refer to lexical items, but to any derived phrases that correspond to the category of transitive verb. For instance, in (4) the phrasal verb and its accompanying particle is elided, yielding the type TV.

(36)   `<wsj_0045>` In CAT sections where students' knowledge of two-letter consonant sounds is tested, the authors noted that Scoring High [_{tv} concentrated on] the same sounds that the test **does** ~~concentrate on~~ to the exclusion of other sounds that fifth graders should know.

Finally, we distinguish between VP antecedents that are realized as past participles as in example (37) below or present participles (7) and those that are not (corresponding to the standard case). We do this both for the full VP and the TV categories.

(37)   `<wsj_0437>` According to Fred Demler, metals economist for Drexel Burnham Lambert, New York, "Highland Valley has already [_{vp} started operating] and Cananea is expected to **do** so soon."

Summing up, this gives us a set of nine different types of syntactic antecedents: `vp` (verb phrase), `vp-ed` (past participle verb phrase), `vp-ng` (present participle verb phrase), `tv` (transitive verb), `tv-ed` (past participle transitive verb), `tv-ng` (present participle transitive verb), `np` (noun phrase), `pp` (prepositional phrase), and `ap` (adjectival phrase). This type information is added in a separate column in the stand-off annotation file.

4.4.3 Source–target patterns

In addition to type of trigger and syntactic antecedent type, we also annotated the surface structure connecting the source and target clause, which we refer to as *source–target patterns*. The aim of these patterns is to supply a different layer of annotation to classify VPE, for instance to facilitate a corpus search of VPE that occur in similar linguistic environments.

The resulting patterns, 34 in total, are listed according to their co-occurrence with the various VPE triggers in the Table 4, and again with their frequency and an example of each in the Appendix. The source–target patterns are mutually exclusive categories, and each instance of VPE can only belong to one class. Furthermore, for ease of classification, the patterns are generalized, and common processes such as possible subject elision, are not listed explicitly in the patterns.

Each source–target pattern is associated with an internal three-letter code which is added in the raw annotation file. The complete annotation of the example discussed in Sect. 4.1 is now `wsj_0112 9215 9219 9188 9207 do vp con` where the last three columns encode the VPE trigger (`do`), the syntactic type of the antecedent (`vp`), and the source–target pattern (`con`, corresponding to Pattern 9 in the Appendix).

4.5 Inter-annotator agreement

Three annotators (including the two co-authors) carried out the VPE identification and antecedent mark-up. Initially, each annotator took a number of sections from the WSJ files and did the annotation individually, asking the other annotators for input for difficult cases. We used WSJ section 19 to check for inter-annotator agreement, which was independently annotated by all three annotators. Section 19 has 21 VPE examples, and while two of the annotators found all 21 cases, one annotator only found only 18 of these 21. Even though this is a small sample size, this excellent inter-annotator agreement on VPE detection shows that the task of agreeing on VPE isn't particularly hard for human annotators (the average of the combined F-scores was 0.97).

After this evaluation, we decided that each section needed to be annotated at least twice to ensure no instances of VPE were missed. After the initial identification and antecedent identification annotation, the syntactic type and source–target connection was annotated by one annotator and checked by the other (the co-authors). Once again, dubious cases were discussed until agreement was reached.

We also manually compared our annotation results with those of Nielsen (2005), who manually annotated seven sections of the WSJ, and Hardt (1997)'s automatic method of detecting VPE examples in the Penn Treebank. We missed two VPE examples that Nielsen found, and Nielsen identified three examples that we didn't consider to be cases of VPE. Hardt's method revealed 14 cases that we missed. The missed cases were added to the final gold-standard annotation. Again, the relatively small number of mismatches demonstrates good inter-annotator agreement.

5 Annotation results

In this section we present some quantitative results from our annotation efforts: frequency of the various VPE triggers, the distribution of the syntactic category of VPE antecedents with respect to VPE triggers and source–target patterns, and the distribution of source–target patterns across VPE triggers.

5.1 Frequency of VPE

In total we found 487 instances of VPE and 67 examples of related phenomena in the 25 sections of the WSJ corpus. VPE frequency differed by section, but this can be attributed to the unequal length of the various sections rather than any type of genre difference among articles.[11] Given the 53,561 sentences (following our own tokenization and sentence boundary determination) present in all 25 sections of the WSJ, we can calculate an average of one VPE (excluding *do X* phenomena) for every 109 sentences.[12]

Just looking at the range of possible VPE triggers (Table 2), the auxiliary *do* was the most frequent one used (in 44% of all cases), followed by *be* (22%) and *have* (9%). Some of the modal verbs were rarely found to trigger VPE, and none of the semi-modal verbs were. Table 2 also lists the proportion of tokens that were cases of VPE for each trigger. Among all auxiliary and modal verbs, *do* is most frequently used as VPE trigger. Perhaps surprisingly, VPE examples with the modal verb *can* are relatively common compared to other triggers.

5.2 Syntactic category of antecedent

Table 3 shows the distribution of syntactic category of the antecedent split by VPE trigger type. Looking first at antecedent type, we see that VP is the most frequent

---

[11] Whether this claim is valid or not is subject to future work. For instance, Webber (2009) shows that genre in the WSJ corpus affects the senses of discourse connectors and non-lexically marked discourse relations.

[12] This fraction is considerably lower than what Nielsen (2005) found, who reports one VPE in every 77 sentences. Recall that Nielsen annotated only seven sections of the WSJ, sections 00–04, 10, and 15, where he found 118 examples of VPE. We counted 14,811 sentences in these seven sections, yielding a VPE rate of one in every 125 sentences.

Table 2 Frequency of VPE triggers in the Wall Street Journal, per token and per actual VPE

Trigger	Token frequency	VPE frequency	Percentage
be	30,124	108	0.36
to	29,868	29	0.10
have	11,249	44	0.39
will	4,385	26	0.59
do	3,103	213	6.86
would	2,954	14	0.47
could	1,465	11	0.75
can	1,171	29	2.48
may	1,049	1	0.10
should	580	7	1.21
might	464	4	0.86
must	327	1	0.31
shall	32	0	0.00

Table 3 Distribution of VP Ellipsis in the Wall Street Journal corpus of the Penn Treebank 3.0, sorted by antecedent syntactic type

	VP	VP-ng	VP-ed	AP	TV	NP	TV-ed	PP	TV-ng	Total
do	176	14	11	0	10	0	1	0	1	213
be	7	11	18	48	0	20	0	4	0	108
have	24	4	12	0	2	0	2	0	0	44
to	19	7	2	0	1	0	0	0	0	29
can	14	3	1	0	7	0	2	0	2	29
will	26	0	0	0	0	0	0	0	0	26
would	10	2	1	0	1	0	0	0	0	14
could	6	3	0	0	2	0	0	0	0	11
should	6	0	1	0	0	0	0	0	0	7
might	4	0	0	0	0	0	0	0	0	4
must	1	0	0	0	0	0	0	0	0	1
may	1	0	0	0	0	0	0	0	0	1
so	42	6	5	0	0	1	0	0	0	54
same	7	1	0	0	0	0	0	0	0	8
likewise	2	0	0	0	0	1	0	0	0	3
opposite	2	0	0	0	0	0	0	0	0	2
	347	51	51	48	23	22	5	4	3	554

syntactic category. Transitive verbs (TV) as antecedents are fairly infrequent, as expected, numbering a total of 31 examples. This category is almost entirely made up of cases of antecedent-contained deletion (29 times); there was only one case of pseudo-gapping and one case of comparative subdeletion.

Predicative ellipsis, comprising the types AP, NP and PP, all occur only with the auxiliary *be*, as expected (Table 3). There are two interesting cases with NP

antecedents: one is an example of a *be-so* anaphor;[13] the other is an interesting case of a *do likewise* anaphor (38), where the antecedent is something like "winning specific benefits", but such a phrase can only be obtained by reformulating the marked noun phrase, which in this case acts as antecedent.

(38) <wsj_1695> Second, it explains why voters hold Congress in disdain but generally love their own congressional representatives: Any individual legislator's constituents appreciate [np the specific benefits that the legislator wins] for them but not the overall cost associated with every other legislator doing **likewise** for his own constituency.

5.3 Source–target patterns

Table 4 presents source–target patterns split by VPE type. It is interesting to note that the four most frequent patterns (column numbers 1–4) all belong to the family of comparative constructions. However, these patterns are not the most frequent patterns for *do X* anaphora like *do so*, suggesting that source–target patterns may be another feature distinguishing these from standard VPE. Further note that the modal verb *can* occurs relatively often with antecedent contained deletion (column number 11)—future work on larger datasets could establish whether such differences are statistically significant or occur by chance.

Table 5 presents the source–target patterns split by antecedent type. The cases of the transitive verb (TV) category co-occur with only four of the patterns, of which two are antecedent-contained deletion patterns (column numbers 11 and 21). More than half of the AP (adjectival phrase) category instances coincide with comparative or equative constructions.

6 Evaluation of VPE processing algorithms

One of the aims of our gold standard annotation of VPE is to use it to test and compare algorithms developed for VPE processing. In this section we give some recommendations for how to use the data, and suggest some evaluation scores to measure the performance of algorithms. We point out a problem with earlier proposals for evaluating selected antecedents for VPE and propose an alternative that deals with this problem.

6.1 Preliminaries

We follow Nielsen (2005) and divide VPE processing into three tasks:

- VPE Detection
- VPE Antecedent Selection
- VPE Resolution

---

[13] Because this example occurs in the test set part of the Wall Street Journal, we refrain from publishing it here.

Table 4 Distribution of VPE source–target patterns across triggers

The patterns (columns 1–34) are:

1. [src NP VP] (.) (just|much) as|like [tgt NP AUX]]
2. [src NP VP X-er than [tgt NP AUX]]
3. [src NP VP as AP as [tgt NP AUX]]
4. [src NP VP] (.) as [tgt AUX NP]
5. [src NP VP] (.) but [tgt NP AUX]
6. [src NP VP] : [tgt NP AUX]
7. [src NP VP] (.|,|–) and [tgt NP AUX]
8. [src NP VP] ,|;|.|– (and) [tgt NP AUX not]
9. [src NP VP] (.|,|and|or) (even) if [tgt NP AUX]
10. [src NP VP] ,|. (and|but) so [tgt AUX NP]
11. [src NP TV NP (that) [tgt NP AUX]]
12. [src NP VP] ? [tgt NP AUX]
13. [src NP VP] ,|;| (and|but) once|before|after|when|until [tgt NP AUX]
14. [src NP VP] before|until|when|whenever [tgt NP AUX]
15. if [src NP VP] (.) [tgt NP AUX]
16. while|although|though [src NP VP] , [tgt NP AUX]
17. [src NP VP] , [tgt AUX NP]]
18. [src NP VP] (.) (in) the (same|exact) way [tgt NP AUX]
19. [src NP TV what [tgt NP AUX]]
20. [src NP VP] (.) (and) nor|neither [tgt AUX NP
21. [src NP VP] . MOD (.) [tgt NP AUX]
22. [src NP VP] (.) because|so [tgt NP AUX]
23. [src NP VP] (.) now|today [tgt NP AUX]
24. [src NP VP] (.) while [tgt NP AUX]
25. [src NP VP] – (and) [tgt NP AUX]
26. [src NP VP] (.|–) (or) [tgt AUX (NEG) NP]?
27. [src NP VP] : [tgt NP AUX]
28. [src NP who|that [src VP] (often|generally|usually) [tgt AUX] so
29. (but) (just) as [src NP VP] , [src NP VP]
30. [src NP VP] (.) (and) [tgt NP AUX] (.) too
31. [src NP who|that [src VP] TV NP who|that [tgt AUX]]
32. [src NP VP as AP as [tgt AUX NP]]
33.–34. (further patterns)

Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
do	42	20	11	18	6	11	6	10	10	13	7	2	1	7	4	0	2	8	1	7	4	1	4	3	3	2	2	3	1	0	3	1	0	0
be	13	21	10	11	5	3	4	11	5	3	0	0	4	0	1	1	4	1	1	0	0	2	1	0	1	0	1	1	2	0	0	0	1	1
have	11	7	3	5	2	1	1	1	2	0	3	0	0	0	2	0	0	0	0	0	0	0	1	2	1	0	0	0	1	1	0	0	0	0
to	1	4	2	0	3	1	4	0	1	0	1	3	2	0	2	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0
can	0	3	4	0	0	1	0	2	0	1	1	0	1	1	2	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
will	1	0	0	0	2	2	3	1	2	2	0	3	2	0	1	4	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
would	4	3	0	0	0	0	0	0	3	1	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
could	0	1	3	0	1	0	1	1	0	0	1	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
should	0	1	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
might	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
must	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
may	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
so	0	0	0	4	0	12	8	8	0	4	0	0	2	0	2	2	3	3	0	4	0	0	0	0	2	0	0	0	0	4	0	0	0	0
same	0	0	0	0	3	2	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
likewise	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
opposite	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	74	61	35	34	33	32	29	26	26	23	13	12	12	12	11	10	10	9	8	6	6	6	6	6	5	4	4	4	4	4	3	2	1	

The first task is a binary classification task: given a potential VPE trigger (i.e., an auxiliary or modal verb), are we dealing with an instance of VPE or not? The second task presupposes an identified VPE and a context, and asks for a string that matches with the textual antecedent. The third task is proper resolution of the elided verb phrase material, including anaphora resolution. We won't say anything about evaluating the third task, mainly because it is difficult to define it in any theory-neutral way, as resolution could take place at the surface level, at the level of syntax, or at the level of semantics. For the tasks of detecting VPE and selecting the antecedent we will instead introduce standard measures: accuracy, precision, and recall.

Before doing this, we need to decide what data to use for developing, training and testing. Depending on whether rule-based or statistical methods will be used, different options are available. In syntactic parsing tasks, Sect. 23 of the WSJ corpus is usually taken as the test corpus. However, because VPE is a relatively rare

Table 5 Distribution of source–target patterns across syntactic category of antecedent

The columns correspond to the following source–target patterns (Pattern numbers 1–34):

1. $[_{src}$ NP VP] (.) (just|much) as|like $[_{tgt}$ NP AUX]
2. $[_{src}$ NP VP X-er than $[_{tgt}$ NP AUX]]
3. $[_{src}$ NP VP as AP as $[_{tgt}$ NP AUX]]
4. $[_{src}$ NP VP] (.) as $[_{tgt}$ AUX NP]
5. $[_{src}$ NP VP] (.|,) but $[_{tgt}$ NP AUX]]
6. $[_{src}$ NP VP] · $[_{tgt}$ NP AUX]]
7. $[_{src}$ NP VP] (.|,|-) and $[_{tgt}$ NP AUX]
8. $[_{src}$ NP VP] (.|,|- (and) $[_{tgt}$ NP AUX not]
9. $[_{src}$ NP VP] (.|,|and|or) (even) if $[_{tgt}$ NP AUX]
10. $[_{src}$ NP VP] ,|. (and|but) so $[_{tgt}$ AUX NP]
11. $[_{src}$ NP TV NP (that) $[_{tgt}$ NP AUX]]
12. $[_{src}$ NP VP] before|until|when|whenever $[_{tgt}$ NP AUX]
13. $[_{src}$ NP VP] ? $[_{tgt}$ NP AUX]
14. $[_{src}$ NP VP] ,|.|; (and|but) once|before|after|when|until $[_{tgt}$ NP AUX]
15. $[_{src}$ NP VP] (.|,) or $[_{tgt}$ NP AUX]
16. if $[_{src}$ NP VP] (,) $[_{tgt}$ NP AUX]
17. while|although|though $[_{src}$ NP VP] , $[_{tgt}$ NP AUX]
18. $[_{src}$ NP VP X-er than $[_{tgt}$ AUX NP]]
19. $[_{src}$ NP VP] (.) $[_{tgt}$ NP AUX]
20. $[_{src}$ NP VP] , $[_{tgt}$ NP AUX]
21. $[_{src}$ NP VP] (.) (in) the (same|exact) way $[_{tgt}$ NP AUX]
22. $[_{src}$ NP TV what $[_{tgt}$ NP AUX]]
23. $[_{src}$ NP VP] (.|,) (and) nor|neither $[_{tgt}$ AUX NP]
24. $[_{src}$ NP VP] . MOD (.) $[_{tgt}$ NP AUX]
25. $[_{src}$ NP VP] − (and) $[_{tgt}$ NP AUX]
26. $[_{src}$ NP VP] (.) because|so $[_{tgt}$ NP AUX]
27. $[_{src}$ NP VP] (.) now|today $[_{tgt}$ NP AUX]
28. $[_{src}$ NP VP] (.) while $[_{tgt}$ NP AUX]
29. $[_{src}$ NP VP] (.|−) (or) $[_{tgt}$ AUX (NEG) NP]?
30. $[_{src}$ NP VP] ; $[_{tgt}$ NP AUX]
31. NP who| that $[_{src}$ NP VP] (often|generally|usually) $[_{tgt}$ AUX] so
32. (but) (just) as $[_{tgt}$ NP AUX] , $[_{src}$ NP VP]
33. NP who|that $[_{src}$ NP VP] TV NP who|that $[_{tgt}$ AUX]
34. $[_{src}$ NP VP as AP as $[_{tgt}$ AUX NP]]

Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
VP	49	31	20	20	23	25	20	15	17	19	0	11	8	10	7	10	6	9	8	5	0	3	3	5	3	4	3	3	3	3	3	1	0	0
VP-ng	12	8	4	1	1	1	2	3	4	0	0	2	0	0	2	0	1	0	0	3	0	0	1	1	1	0	0	0	0	1	0	1	2	0
VP-ed	6	5	3	7	2	4	6	0	2	3	0	0	0	2	2	0	0	0	0	0	0	3	1	0	2	1	0	0	0	0	1	1	0	0
TV	1	1	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
TV-ed	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TV-ng	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AP	3	16	8	4	2	1	0	5	0	1	0	0	3	0	0	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NP	2	0	0	2	4	1	1	2	3	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	0	0
PP	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	74	61	35	34	33	32	29	26	26	23	23	13	12	12	12	11	10	10	9	8	6	6	6	6	6	5	4	4	4	4	4	3	2	1

phenomenon, and a single section wouldn't provide sufficient examples for any meaningful evaluation, we propose to set aside the five sections 20, 21, 22, 23, 24 as test data, which comprise ca. 100 instances of VPE (translating to 20% of the total). Alternatively, one could apply cross-validation techniques when using statistical methods.

6.2 Evaluating VPE detection

The accuracy of VPE Detection can be measured by computing precision, recall, and the F-score, the harmonic mean of precision and recall (Nielsen 2005). Precision (P) is computed by dividing the number of correctly detected VPE by the number of all (correctly or incorrectly) detected VPE. Recall (R) is computed by dividing the number of correctly detected VPE by the number of VPE in the gold standard annotation. The harmonic mean or F-score is computed in the traditional way: $F = 2PR/(P + R)$.

We illustrate these measures with the automatic method that Hardt (1997) developed for detecting VPE examples in the Penn Treebank. We reproduced Hardt's method by searching for the pattern

$$(\text{VP} \ * \ ? *$$

in the parsed version of the WSJ sections. Our search gave 295 hits, of which 281 were judged by us to be genuine cases of VPE. This gives a precision of 0.95 (281/295) and a relatively low recall of 0.58 (281/487), and an *F*-score of 0.72.

6.3 Evaluating VPE antecedent selection

The accuracy of VPE antecedent selection can be measured by computing the average *F*-score on correctly identified tokens per antecedent. Here, the precision (*P*) of a selected antecedent is the number of correctly identified tokens divided by the number of proposed tokens, and recall (*R*) the number of correctly identified tokens divided by the number of tokens in the gold-standard antecedent. The *F*-score for one antecedent is computed by taking the harmonic mean of *P* and *R* (see above), and the average *F*-score for all computed antecedents is the final score denoting the accuracy of VPE antecedent selection.[14]

This measure for antecedent selection accuracy is rather crude, of course, as it doesn't discriminate between tokens. It treats all tokens within an antecedent equally, disregarding their length or linguistic importance. So it would assign the same accuracy scores to the two fictitious proposals in (39a) and (39b), even though intuitively, we would tend to give more credit to the former answer, constituting a verb, than the latter, covering just a noun.

(39)　a.　Admittedly last season's runaway hit, "Steel Magnolias," [helped] a lot, but so **did** cost cutting and other measures insisted on by the board.
　　　b.　Admittedly last season's runaway hit, "Steel Magnolias," helped a [lot], but so **did** cost cutting and other measures insisted on by the board.

Nonetheless, this way of measuring the accuracy of antecedent selection is relatively straightforward to compute, and moreover, it improves on previous proposals. Hardt (1997) proposes three measures: (1) **head overlap** between proposed and gold-standard antecedent, (2) **head match**, and (3) **exact match**. Hardt's exact match ("the system choice and code choice match word-for-word") is identical to our proposed measure with a perfect recall and precision. Hardt's head overlap ("either the head verb of the system choice is contained in the code choice, or the head verb of the coder choice is contained in the system choice") is not really useful, as systems could choose to return large portions of text and so fool the evaluation procedure.[15] Hardt's proposal for head match ("the system choice and coder choice have the same head") is the most useful of the three but requires agreeing on a definition of *head* in complex verb phrases. Our proposed method of using the harmonic mean for comparing the computed antecedent with the gold-standard antecedent doesn't run into these problems.

---

[14] One anonymous reviewer suggested to use the Dice coefficient to calculate the similarity between computed and gold-standard antecedent. This would then be equivalent to computing the *F*-score.

[15] In the extreme (but rather silly) case, a system could return the entire text as antecedent, thereby earning a perfect score.

7 Discussion and conclusions

7.1 Empirical vs. formal studies of ellipsis

Our frequency results were in some cases predictable. Comparative subdeletion and pseudo-gapping were not expected to be very frequent and in fact were not. But in contrast to the expectations raised by formal work, our frequency results are surprising in two ways: we found types that make up a considerably greater proportion of VPE cases than theoretical work suggests, and types discussed frequently in the theoretical work that are extremely infrequent or non-existent in the data annotated.

For example, comparative and equative constructions were frequent with VPE, 193 cases, or 31% of the total number of examples. Theoretical papers on VPE seldom use these constructions to illustrate processes in ellipsis yet these constructions seem to facilitate the use of VPE.

Actually, the standard example of VPE in theoretical work consists of two sentences conjoined with *and*, where the second sentence is marked by the presupposition trigger *too*. In such examples, the event in the target clause is linked presuppositionally with the source clause event. But among the 554 examples found, there were only five cases with *too*: three cases where the source–target pattern includes *too*, and two additional examples with *too* marking the source–target connection, while not being part of the connection itself. The dominance of this type of example in the theoretical literature gives the impression that presuppositional marking is perhaps an essential part of VPE licensing, and indeed work like Rooth (1992) and Bos (1994) use the identification of focus and redundancy communicated by focus particles like *too* as a part of their resolution algorithm. Because the Wall Street Journal texts are written, it is somewhat surprising to find so few overt markers of redundancy if its recognition is instrumental in ellipsis interpretation.

Another interesting finding was the lack of examples with bound pronouns in the source clause. VPE examples with quantifiers or pronouns in the source clause can lead to ambiguities in the target clause, leading to the well-studied strict or sloppy interpretations of pronouns in elided VPs. An example from the corpus is (40), where the possessive pronoun is co-indexed with its intended antecedent.

(40) `<wsj_0445>` IBM$_1$, though long a leader in the Japanese mainframe business, didn't [$_{vp}$ introduce its$_1$ first PC in Japan] until 5 years after NEC$_2$ **did** ~~introduce its$_2$ first PC in Japan~~, and that wasn't compatible even with the U.S. IBM standard.

In the corpus we encountered only nine such examples (less than 2% of all VPE found). All nine cases seemed to license a strongly preferred sloppy interpretation, as above. We also found no examples where VPE interacted with quantifier scope. Neither did we encounter any examples of split antecedents, i.e. combined VPs, as discussed by Prüst (1992) and Asher (1993). We did, however, find a few examples of coordinated VPs acting as antecedents:

(41) `<wsj_0842>` Goodwill planned to [ᵥₚ sell the property and pocket the proceeds], as it **had** in many similar cases.

Of course, even if a VPE type is infrequent, it may still be the source of important theoretical insights about language and language processing. But we also believe that the characteristics of very frequent VPE types offer important clues about the nature of VPE and how we interpret it. For this reason, the distributional information in the annotation can be a useful resource for theoretical and experimental work on VPE resolution.

7.2 Corpus utility

This work makes a number of contributions to the goal of automatic recognition, analysis and resolution of ellipsis. The stand–off format of the annotation makes it easily usable by researchers who want to combine the information with different tokenizers, taggers, or parsers because it makes the data independent of these choices. Further, the amount of data annotated, while limited, is large enough to give useful frequency information for developing heuristics for VPE resolution. The coverage, usability, and detail of annotation of this corpus should make it highly useful for the development of algorithms and we hope it will lead to more efforts to deal with VPE in language processing systems.

The source–target classification is a novel contribution and convincingly illustrates the utility of larger annotation projects. It is somewhat surprising to realize that, at least in newspaper texts, VPE occur within a limited set of syntactic contexts. This classification is particularly valuable from at least two perspectives.

First, it can be seen as a descriptive, linguistic study of VPE applied to a large corpus. Information about the frequency and distribution of syntactic patterns can give us more insights into what linguistic rules govern VP ellipsis. A case in point is VP-cataphora (where the antecedent comes after the VPE trigger) where all four cases fell under Pattern 31 shown in the Appendix. A small sample—true—but certainly an interesting observation from a linguistic point of view.

Second, the patterns could serve as a resource for automating the detection and resolution of VPE in NLP applications. Symbolic approaches could base their rules for detection and antecedent location on these patterns. Machine learning approaches could also select features on the basis of these syntactic patterns. We believe this is a promising approach because some of the patterns seem to be very good predictors of where the antecedent might be found, both in terms of proximity and direction. VP-cataphora mentioned above is one obvious example. Another example is Pattern 1 in Appendix. Instances of VPE found with this pattern require very close antecedents. These are just two examples of how the patterns could help in developing a rule-based antecedent identification method. There are many more.

Further, as far as we know, our annotation of a number of related forms is novel. It provides the first quantitative data available about the class of *do X* anaphora. Together with the standard VPE annotation, it gives a first overview of the syntactic types and the constructions these related forms occur with. Our inclusion of a

number of predicative ellipsis types is also useful, since these constructions are often included in discussions of VPE, but not specifically addressed, so this makes a first modest contribution to expanding our knowledge of these anaphoric devices.

7.3 Directions for future research

An obvious extension of this work is to annotate a number of closely related forms that we excluded. For example, There are cases of predicative ellipsis that occur with verbs other than auxiliaries or modals, such as (42) and (43).

(42) <wsj_0405> The sales pitch mightn't be as farfetched as it seems.
(43) <wsj_0972> So it's entirely possible that "Look Who's Talking" isn't as entertaining as it seems in comparison to the turgid other films opening now.

These examples only occur in comparative or equative sentences and they are also limited to reporting or parenthetical verbs ("verba dicendi"). It would be interesting to see what further patterns they display. In general we need much more empirical data about equative and comparative constructions, as well as some discussion about how they should be annotated. Comparative and equative constructions seem to strongly interact with a number of other constructions, and they seem to frequently license particular sub-types of anaphoric processes. These constructions also seem to interact with particular lexical-semantic classes, and more empirical work about this interaction would also be useful.

The current annotation could be extended to include more semantic or discourse information. The syntactic patterns could serve as a base for creating a more discourse relation based annotation. Alternatively, there already exists a discourse annotated version of the Wall Street Journal, the Penn Discourse Treebank (Prasad et al. 2008), and it would be useful to determine how instances of VPE correlate with rhetorical relations. Correlation of VPE with genre is another interesting direction to explore, even more so since annotation of genre for the WSJ exist (Webber 2009).

We also have made no attempt to resolve the ellipsis examples into some natural language form, the third ellipsis processing task as identified in Nielsen (2005). This is partly because many examples would require extensive paraphrasing, and necessitate using techniques from natural language generation (NLG). Nevertheless, the question of simple insertability of the surface antecedent at the elision point should be tested and evaluated. More useful than resolving VPE examples to some natural language form would be to make sure their informational contribution is correctly included in a semantic representation. This would forgo any NLG work and seems to be a more efficient approach, but of course, the authors are biased here.

Working on a more semantic level also gives another way to evaluate VPE resolution. The setup known from recognizing textual entailment (Dagan et al. (2006) extends naturally to this task: identifying valid entailments from a VPE requires correctly identifying the antecedent, e.g. <wsj_0445>: "*But early on, IBM offered its basic design to anybody wanting to copy it. Dozens of small companies*

did, swiftly establishing a standard operating system." should not entail that dozens of small companies *wanted* to copy IBM's design, but that they, in fact, *copied* it.

As a final note, we should admit that the endeavor took more time than we originally anticipated (pure annotation took between 2 and 3 h per WSJ file, per annotator, excluding discussion and development time), and this is certainly one of the reasons why large coverage manual annotation is done far less often than it should be. But we feel that the result was worth the effort. This type of comprehensive manual annotation is essential for the development of automatic methods for dealing with VPE, or any complex linguistic structure.

Appendix: Source–target patterns in VP ellipsis

This appendix lists all the patterns of linking between source and target clause for the examples found in the corpus. Each pattern contains the antecedent verb phrase (VP), the subject noun phrase in the target clause (NP) and auxiliary verb triggering the ellipsis (AUX). Vertical bars in the patterns indicate alternatives, and round brackets mark optional elements. In the corpus examples, the VPE antecedents are enclosed in square brackets, and the matched pattern is typeset in a boldface font. Patterns are ordered by frequency.

1. $[_{src}$ NP VP$]$ $(,)$ $(\text{just}|\text{much})$ as$|$like $[_{tgt}$ NP AUX$]$ (ASI, 74 cases)
    The parishioners of St. Michael and All Angels stop to chat at the church door, as members here always have. `<wsj_0089>`
2. $[_{src}$ NP VP X−er than $[_{tgt}$ NP AUX$]]$ (COM, 61 cases)
    And then, asset values of the funds could plunge more than they have so far. `<wsj_0768>`
3. $[_{src}$ NP VP as AP as $[_{tgt}$ NP AUX$]]$ (EQU, 35 cases)
    Department economists don't expect 1989 to be as good a year as 1988 was. `<wsj_0113>`
4. $[_{src}$ NP VP$]$ $(,)$ as $[_{tgt}$ AUX NP$]$ (ASA, 34 cases)
    Both banks have been battered, as have other Arizona banks, by falling real estate prices. `<wsj_1299>`
5. $[_{src}$ NP VP$]$ $(.|,)$ but $[_{tgt}$ NP AUX$]$ (BUT, 33 cases)
    He has not changed, but those around him have. `<wsj_0747>`
6. $[_{src}$ NP VP$]$. $[_{tgt}$ NP AUX$]$ (FUL, 32 cases)
    But early on, IBM offered its basic design to anybody wanting to copy it.

Dozens of small companies did, swiftly establishing a standard operating system. <wsj_0445>

7. [$_{src}$ NP VP] (.|,| − −) and [$_{tgt}$ NP AUX] (AND, 29 cases)
Things were supposed to change when Vietnam's economic reforms gathered pace, and for awhile they did. <wsj_1603>

8. [$_{src}$ NP VP] .|; |.| − − (and) [$_{tgt}$ NP AUX not] (NOT, 26 cases)
The onslaught of the program selling dashed any hopes that some of the big program trading firms would hold off until the market stabilized. They didn't. <wsj_2232>

9. [$_{src}$ NP VP] (.|,|and|or) (even) if [$_{tgt}$ NP AUX] (CON, 26 cases)
It isn't yet clear whether the 200-ruble limit will be lifted. If it isn't, the black market for dollars probably will continue to thrive. <wsj_0934>

10. [$_{src}$ NP VP] ,|. (and|but) so [$_{tgt}$ AUX NP] (SOA, 23 cases)
Admittedly last season's runaway hit, "Steel Magnolias," helped a lot, but so did cost cutting and other measures insisted on by the board. <wsj_0819>

11. [$_{src}$ NP TV NP (that) [$_{tgt}$ NP AUX]] (REL, 23 cases)
Maybe we recognize values the other guys don't," Mr. Pilson says. <wsj_1057>

12. [$_{src}$ NP VP] before|until|when|whenever [$_{tgt}$ NP AUX] (WHE, 13 cases)
On days when prices are tumbling, they must be willing to buy shares from sellers when no one else will. <wsj_0585>

13. [$_{src}$ NP VP] ? [$_{tgt}$ NP AUX] (QUM, 12 cases)
Can Sihanouk and Hun Sen knock off the Khmer Rouge still supported by China? He can't. <wsj_0296>

14. [$_{src}$ NP VP] .|,|; (and|but) once|before|after|when|until [$_{tgt}$ NP AUX] (ONC, 12 cases)
It added another 5% Monday before stocks rallied. When they did, the firm reduced those stock holdings to about 70%. <wsj_2057>

15. [$_{src}$ NP VP] (.|,) or [$_{tgt}$ NP AUX] (DIS, 12 cases)
"You either believe Seymour can do it again or you don't." <wsj_0018>

16. if [$_{src}$ NP VP] (,) [$_{tgt}$ NP AUX] (THE, 11 cases)
If I need to ask for money up front later, I will. <wsj_1002>

17. while|although|though [$_{src}$ NP VP] , [$_{tgt}$ NP AUX] (WHN, 10 cases)
While the theme is compelling, the plot and characters are not. <wsj_0790>

18. [$_{src}$ NP VP X − er than [$_{tgt}$ AUX NP]] (COI, 10 cases)
One analyst noted that the company often has better store locations than do its franchisees, thus aiding promotional efforts. <wsj_1280>

19. [$_{src}$ NP VP] , [$_{tgt}$ NP AUX] (CMA, 9 cases)
Even the Soviet Union has Peter the Great to rediscover, should it choose to. <wsj_1929>

20. [$_{src}$ NP VP] (,) (in) the (same|exact) way [$_{tgt}$ NP AUX] (WAY, 8 cases)
A Japanese apple called the Fuji is cropping up in orchards the way Hondas did on US roads. <wsj_1128>

21. [$_{src}$ NP TV what [$_{tgt}$ NP AUX]] (WHA, 6 cases)
Please contribute what you can," the ad said. <wsj_0453>

22. $[_{src}$ NP VP] (.|,) (and) nor|neither $[_{tgt}$ AUX NP (NOR, 6 cases)
    See, the other rule of thumb about ballooning is that you can't steer. And
    neither can your pilot. <wsj_0239>

23. $[_{src}$ NP VP] . MOD (,) $[_{tgt}$ NP AUX] (MOD, 6 cases)
    It's a mystery how this could have escaped the notice of Nebraska coaches.
    Probably, it didn't. <wsj_0966>

24. $[_{src}$ NP VP] $--$ (and) $[_{tgt}$ NP AUX] (DAS, 6 cases)
    But even if he agrees, Mr. Kemp doesn't write the nation's housing law—
    Congress does. <wsj_0309>

25. $[_{src}$ NP VP] (,) because|so $[_{tgt}$ NP AUX] (BEC, 6 cases)
    They say insurance companies use policies aimed at excluding bad risks
    because their competitors do. <wsj_0518>

26. $[_{src}$ NP VP] (.|,) now|today $[_{tgt}$ NP AUX] (NOW, 5 cases)
    The sales job seems to be paying off: When he bought the team, only six of the
    suites had been sold. Today, 30 have. <wsj_1411>

27. $[_{src}$ NP VP] (,) while $[_{tgt}$ NP AUX] (WHI, 4 cases)
    But from early on, Tiger's workers unionized, while Federal's never have.
    <wsj_1394>

28. $[_{src}$ NP VP] (,|$--$) (or) $[_{tgt}$ AUX (NEG) NP]? (TAG, 4 cases)
    But you knew that, didn't you? <wsj_1618>

29. $[_{src}$ NP VP] ; $[_{tgt}$ NP AUX] (SEM, 4 cases)
    In 1980, 18% of federal prosecutions concluded at trial; in 1987, only 9% did.
    <wsj_0617>

30. NP who|that $[_{src}$ VP] (often|generally|usually) $[_{tgt}$ AUX] so (RSO, 4 cases)
    Moreover, they note, those who manage to pay their own way often do so only
    by selling their homes, using up life savings or drawing heavily on children
    and other relatives. <wsj_0595>

31. (but) (just) as $[_{tgt}$ NP AUX] , $[_{src}$ NP VP] (CAT, 4 cases)
    As she has done in the past, she stated her support for Mr. Lawson but insisted
    on keeping on an adviser who opposed and disparaged his policies.
    <wsj_0883>

32. $[_{src}$ NP VP] (.|,) (and) $[_{tgt}$ NP AUX](,)too (TOO, 3 cases)
    Dean Steinkuhler, a spectacularly bulked-up former lineman, confesses that he
    used 'em, and says other Huskers did too. <wsj_0966>

33. NP who|that $[_{src}$ VP] TV NP who|that $[_{tgt}$ AUX] (SCL, 2 cases)
    Firms that are paying employment taxes also provide leads to competitors that
    aren't, he says. <wsj_1570>

34. $[_{src}$ NP VP as AP as $[_{tgt}$ AUX NP]] (EQI, 1 case)
    He isn't as well-known to clients as is Mr. Roman. <wsj_0554>

References

Asher, N. (1993). *Reference to abstract objects in discourse*. Dordrecht: Kluwer.

Bos, J. (1994). Presupposition & VP-ellipsis. In *Proceedings of the 15th international conference on computational linguistics (Coling 1994), Kyoto, Japan*, pp. 1184–1190.

Bouton, L. (1970). Antecedent-contained pro-forms. In *Papers from the 6th regional meeting of the Chicago Linguistic Society, Chicago, IL*, pp. 154–167.

Bowers, J. (1998). *On pseudo-gapping*, ms., Cornell University.

Bresnan, J. (1973). Syntax of the comparative construction in English. *Linguistic Inquiry, 4*(3), 275–345.

Carlson, L., Marcu, D., & Okurowski, M. E. (2001). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the 2nd SIGdial workshop on discourse and dialogue, association for computational linguistics, Morristown, NJ*, pp. 1–10.

Chierchia, G., & McConnell-Ginet, S. (1991). *Meaning and grammar. An introduction to semantics*. Cambridge: The MIT Press.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press.

Curran, J., Clark, S., & Bos, J. (2007). Linguistically motivated large-scale nlp with c&c and boxer. In *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions, Prague, Czech Republic*, pp. 33–36.

Dagan, I., Glickman, O., & Magnini, B. (2006). The PASCAL recognising textual entailment challenge. In *Lecture notes in computer science* (Vol. 3944, pp. 177–190).

Dahl, O. (1973). On so-called 'sloppy identity'. *Synthese, 26*, 81–112.

Dalrymple, M., Shieber, S. M., & Pereira, F. C. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy, 14*(4), 399–452.

Erk, K., & Koller, A. (2001). VP ellipsis by tree surgery. In *Proceedings of the 13th Amsterdam colloquium, Amsterdam*, pp. 85–90.

Fiengo, R., & May, R. (1994). *Indices and identity*. Cambridge, MA: MIT Press.

Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry, 7*(3), 391–428.

Hardt, D. (1997). An empirical approach to VP ellipsis. *Computational Linguistics, 23*(4), 525–541.

Hardt, D. (1999a). Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy, 22*(2), 187–221.

Hardt, D. (1999b). VPE as a proform: Some consequences for binding. In F. Corblin, C. Dobrovie-Sorin, & J. M. Marandin (Eds.), *Empirical issues in formal syntax and semantics 2*. Selected Papers from the Colloque de Syntaxe et Sémantique à Paris (pp. 215–232). The Hague: Thesus.

Hirschbühler, P. (1982). VP deletion and across-the-board quantifier scope. In J. Pustejovsky & P. Sells (Eds.), *Proceedings of the 12th annual meeting of the North Eastern Linguistic Society, Amherst, MA* (pp. 132–139).

Hoeksema, J. (2006). Pseudogapping: Its syntactic analysis and cumulative effects on its acceptability. *Research on Language and Computation, 4*(4), 335–352.

Kehler, A. (1993). The effect of establishing coherence in ellipsis and anaphora resolution. In *Proceedings of the 31st annual meeting of the association for computational linguistics, Columbus, OH*, pp. 62–69.

Kehler, A. (2002). *Coherence, reference and the theory of grammar*. Stanford: CSLI Publications.

Kehler, A., & Shieber, S. (1997). Anaphoric dependencies in ellipsis. *Computational Linguistics, 23*(3), 457–466.

Kehler, A., & Ward, G. (1999). On the semantics and pragmatics of identifier so. In *The semantics/pragmatics interface from different points of view*, pp. 233–256. Elsevier.

Klein, E. (1987). VP ellipsis in DR theory. In J. Groenendijk, D. de Jongh, & M. Stokhof (Eds.), *Studies in discourse representation theory and the theory of generalised quantifiers* (pp. 161–187). Dordrecht, The Netherlands: Foris Publications.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2), 313–330.

May, R. (1985). *Logical form*. Cambridge, MA: MIT Press.

Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.

Miller, P. (1990). Pseudogapping and *do so* substitutions. In *Papers from the 26th regional meeting of the Chicago Linguistic Society, Chicago, IL*, pp. 293–305.

Nielsen, L. A. (2005). *A corpus-based study of verb phrase ellipsis identification and resolution*. PhD thesis, King's College London.

Pradhan, S. S., Hovy, E. H., Marcus, M. P., Palmer, M., Ramshaw L. A., & Weischedel, R. M. (2007). Ontonotes: A unified relational semantic representation. In *Proceedings of the 1st IEEE international conference on semantic computing (ICSC 2007), Irvine, CA*, pp. 517–526.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., et al. (2008). The Penn Discourse TreeBank 2.0. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, et al.

(Eds.), *Proceedings of the 6th international conference on language resources and evaluation (LREC 2008), Marrakech, Morocco* (pp. 2961–2968).

Prüst, H. (1992). *On discourse structuring, VP anaphora and gapping*. PhD thesis, University of Amsterdam.

Pullum, G. (2000). Hankamer was! In S. Chung, J. McCloskey, & N. Sanders (Eds.), *Jorge Hankamer WebFest*. California: University of California

Rooth, M. (1992). Ellipsis redundancy and reduction redundancy. In S. Berman, & A. Hestvik (Eds.), *Proceedings of the Stuttgart Ellipsis workshop*. Arbeitspapiere des Sonderforschungsbereichs 340 (Vol. 29). Heidelberg: IBM Germany.

Sag, I. (1976). *Deletion and logical form*. PhD thesis, Massachusetts Institute of Technology.

Schiehlen, M. (1999). *Semantikkonstruktion*. PhD thesis, Universität Stuttgart.

Shieber, S. M., Pereira, F. C. N., & Dalrymple, M. (1996). Interactions of scope and ellipsis. *Linguistics and Philosophy, 19*, 527–552.

Ward, G., & Kehler, A. (2005). Syntactic form and discourse accessibility. In A. Branco, T. McEnery, & R. Mitkov (Eds.), *Anaphora processing: Linguistic, cognitive and computational modelling* (pp. 365–384). John Benjamins.

Webber, B. (2009). Genre distinctions for discourse in the penn treebank. In *Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP, Suntec, Singapore*, pp. 674–682.

Weischedel, R., & Brunstein, A. (2005). *BBN pronoun coreference and entity type corpus*.

Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics, 31*(2), 249–287.