Anne Abeillé* The empirical turn and its consequences for theoretical syntax

https://doi.org/10.1515/tl-2024-2011

Abstract: In a pioneer paper, Featherston (Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. Theoretical Linguistics 33. 269-318) advocated the use of better controlled data in theoretical linguistics. Despite diverging on many aspects, most syntactic theories are now testing their hypotheses with more data than a few linguists' intuitions. I will examine the consequences of this empirical turn on two syntactic phenomena: long-distance dependencies (LDD) and ellipsis. In a series of recent experiments (Liu, Yingtong, Elodie Winckel, Anne Abeillé, Barbara Hemforth & Edward Gibson. 2022. Structural, functional and processing perspectives on linguistic islands effects. Annual Review of Linguistics 8. 495–525), most of the syntactic constraints ('island constraints') on LDD have shown less crosslinguistic variation and more cross-construction variation than previously thought. Corpus and experimental data have also shown elliptical clauses to be more flexible than expected under deletion-under-identity theories (Poppels, Till. 2022. Explaining ellipsis without identity. The Linguistic Review 39. 341–400). These are challenges for most syntactic theories, which call for taking discourse factors more seriously into account.

Keywords: corpus; experiments; locality constraints; ellipsis; French; English

1 Introduction

In *Theoretical Linguistics*, Featherston (2007: 33) pioneered a warning against "the intellectually corrosive effect of paying insufficient attention to data" and reminded us that "data is a pre-condition for theory, and the quality of a theory can never exceed the quality of the data set which it is based on." He advocated using controlled experiments, with multiple informants, and multiple lexical variants of the structures at stake, moving away from the one-item one-informant method, which Gibson and Fedorenko (2010) called 'weak standards' in syntax and semantics.

At the same time, large corpora have become available for a variety of languages, and syntactically annotated databases (treebanks) and annotation tools (parsers)

^{*}Corresponding author: Anne Abeillé, Laboratoire de Linguistique Formelle, Université Paris Cité, Paris, France, E-mail: anne.abeille@u-paris.fr. https://orcid.org/0000-0002-9187-2298

have made it possible to search for syntactic structures (Abeillé 2003; Nivre et al. 2016). While not providing ungrammatical variants, nor acceptability degrees, such corpora may provide naturally occurring examples with appropriate lexical choice, prosody (if spoken) and context, that may escape the experimenter's intuition and thus usefully complement experiments.

Despite diverging on many crucial aspects (such as the role of the lexicon, the status of grammatical functions, the status of invisible elements, the need for constructions as non-compositional form-meaning pairs, etc.), most syntactic theories now aim at testing their hypotheses with more and better controlled data. I will briefly present how new data and experimental results have called for revising analyses of two major syntactic topics: locality constraints (*Theoretical Linguistics* vol. 33 and 36) and ellipsis (*Theoretical Linguistics* vol. 42).

2 Locality constraints and the empirical turn

Ross 1967's work was seminal in proposing a number of locality constraints on whdependencies and coining the term 'island' for configurations blocking movement: sentential subject island (1a), relative clause island (1b), and coordinate structure island (1c).¹

- (1) a. * The teacher who [that the principal would fire _] was expected by the reporters is a crusty old battleax. (Ross 1967)
 - b. * What do you know a man [who saw _]?
 - c. * What did John buy [a shirt and _] / [_ and a shirt]?

For syntactic theories, these constraints are innate and part of universal grammar, and more constraints have been added over the years, on all subjects (2a) by Chomsky (1973), on all adjuncts (2b) by Huang (1982), and on various kinds of complements: *wh*-clauses (2c), *that*-clauses after factive verbs, complements of nouns, etc.

- (2) a. * Who did [stories about _] terrify John? (Chomsky 1973: 249)
 - b. * What did you worry [if John bought _]? (Sprouse et al. 2016)
 - c. * What did you wonder [whether John bought _]? (Sprouse et al. 2016)

Over the years, counterexamples have also been found to most locality constraints, for instance, to the complex NP constraint (3a), the adjunct constraint (3b), or the relative clause constraint (3c):

¹ I represent the canonical position with _, for clarity, but this does not mean that the fronted element depends on a 'gap'. For theories such as LFG (Bresnan et al. 2015), HPSG (Pollard and Sag 1994; Müller et al. 2021) or Construction Grammar (Goldberg 2013), the dependency is between the fronted element and a head ('fire' or 'saw' in 1a,b).

- (3) a. Which Middle East country did you hear [rumors [that we had infiltrated _]]? (Pollard and Sag 1994: 206)
 - b. This is the watch that I got upset [when I lost _]. (Truswell 2011: 175)
 - violence is something that there are many Americans [who condone _]. (McCawley 1981: 108)

Different syntactic theories have been proposed: subjacency (Chomsky 1973), barriers (Chomsky 1986), freezing (Wexler and Culicover 1980), and phase theory (Chomsky 2008; den Dikken 2007). Freezing is supposed to account for subject islandhood if the subject moves outside the VP (2a) but may be problematic if subjects of passive verbs are more permissive (4a). Phase theory (Chomsky 2008) was designed to account for CP and DP islands, but as observed by den Dikken (2007), only definite DPs (4b) should be considered phases and thus opaque for extraction:

- (4) a. *Of which car was [the driver _] awarded a prize?* (Chomsky 2008: 147)
 - b. * *Who did you read* [*the/that book about* _]? (den Dikken 2007: 35)
 - c. Who did you read [a book about _]?

However, such approaches mostly predict a binary contrast between grammatical (4c) and ungrammatical (4b) LDD. In a series of experiments, Keller (2000) has found that extraction out of object NPs is ameliorated with indefinites and penalized with definites but with gradient acceptability:

- (5) a. Which actress did you buy [a picture of]?
 - b. ? Which actress did you buy [the picture of _]?
 - c. ?? Which actress did you buy [that picture of _]?

Gradient grammaticality is accounted for by other approaches to locality constraints. For processing theories (Hofmeister and Sag 2010; Kluender 1991, etc.), islands effects arise from (cumulative) processing overload and may come from working memory limitation (Hofmeister and Sag 2010), interference (Gibson 1998) and surprisal (Chaves 2013; Chaves and Dery 2019). For discourse-based theories, extraction out of 'backgrounded' constituents is difficult because they are not accessible when looking for a gap (Erteschik-Shir 1973; Erteschik-Shir and Lappin 1979) or their discourse status conflicts with that of the filler (Abeillé et al. 2020; Goldberg 2013). For non-syntactic approaches, unacceptable long-distance dependencies may not be ungrammatical and may be ameliorated by lexical choice and context.

As shown by Kluender (1991) and Hofmeister and Sag (2010), with controlled experiments, extraction out of a *wh*-clause yields outputs of gradient acceptability across languages, with the nature of the extracted element and the finiteness of the embedded clause playing a crucial role:

- (6) a. ?? [What] do you wonder [whether John has read _]?
 - b. ? [What] do you wonder [whether to read _]?
 - c. [Which book] do you wonder [whether to read _]?

The interference of a nominal subject ('John') with a finite verb may explain the penalty of (6a) over (6b). If nominal *wh*-phrases ('which book') are better anchored in memory than pronominal ones ('what'), the amelioration of (6c) compared to (6b) may be explained.

2.1 The 'adjunct island' constraint and the empirical turn

More recently, a quantitative approach has been proposed such that the unacceptability of an 'island' sentence goes beyond the additive badness of its complexity components (Hofmeister and Sag 2010; Sprouse et al. 2016, among others). Specifically, Sprouse et al. (2016) suggest that an 'island effect' is a super-additive interaction between complexity and distance. For example, they tested the adjunct island constraint with the following conditions: comparing short (7a,b) and long (7c,d) dependencies and adjunct *if*-clauses (7b,d) and complement *that*-clauses (7a,c):²

- (7) a. short-cplt: Who _ thinks [that the lawyer forgot his briefcase at the office]?
 - b. short-adjt: Who _ worries [if the lawyer forgot his briefcase at the office]?
 - c. long-cplt: What do you think [that the lawyer forgot _ at the office]?
 - d. long-adjt: What do you worry [if the lawyer forgets _ at the office]? (Sprouse et al. 2016)

They found an interaction, i.e., a super-additive penalty for (7d). However, applying the same design to relative clauses, they did not find such a penalty for the same adjunct clause (8b), compared with the same complement clause (8a). In a corpus study, Müller and Eggers (2022) found naturally occurring examples with relative clauses (8c), not with *wh*-questions:

- (8) a. I called the client who the secretary thought [that the lawyer insulted _].
 - b. *I* called the client who the secretary worries [if the lawyer insults _].
 - c. Many of the exercises are ones that I would be surprised [if even 1 percent of healthy women can do _]. (COCA) (Müller and Eggers 2022)

The same difference between *wh*-questions and relative clauses was found with a similar experimental paradigm in Norwegian (Kobzeva et al. 2022; Kush et al. 2017).

² As pointed out by Liu et al. (2022), it is also important to compare putative island cases with ungrammatical controls, since an 'island' is supposed to be an ungrammatical LDD.

This difference between constructions is not predicted by syntactic nor processing theories.³ Following Abeillé et al. (2020), Liu et al. (2022) proposed a discourse-based explanation: if the adjunct-clause is more backgrounded than the complement clause, focalizing an element out of an adjunct with a *wh*-question may lead to a discourse clash, which is not the case with a relative clause, which is not a focalizing construction.

2.2 The 'subject island' constraint and the empirical turn

A similar discourse-based approach may also account for recent findings on nominal 'subject islands'.

To account for crosslinguistic variation, Rizzi (1982) suggested that null-subject languages, such as Spanish or Italian, may be more permissive in not having a fixed subject position. However, French is not a null subject language, but has been claimed to allow for extraction out of subjects, at least with *dont* ('of which') (9a). For other relativizers such as *de qui* ('of whom'), Godard (1988) (9b) and Tellier (1990) (9c) gave contradictory judgments:

(9) a. la jeune femme dont [le portrait] est à la fondation Barnes the young woman of.which the portrait is at the foundation Barnes (Godard and Sag, 1996: 63)

'this young lady, of which the portrait is at the Barnes foundation'

- b. *un homme de qui [la force de travail _] est étonnante* (Godard, 1988: 56) 'a man of whom the power of labour is astonishing'
- c. * *le diplomate de qui* [*la secrétaire*] *a téléphoné* (Tellier, 1990: 307) 'the diplomat of whom the secretary has called'

To settle the data, we conducted several corpus studies and experiments. In large written and spoken corpora (Frantext after 1900, French Treebank, CFPP2000), we found a plethora of relative clauses relativizing the complement of the subject. In written French, it is the most common use of both *dont* and *de qui* relative clauses, much more frequent than relativizing the complement of the object (Abeillé and Winckel 2020).

³ Cinque (1996: 241) proposed that movement of referential phrases is more permissive than that of non-referential ones, thus accounting for the amelioration of (6c) over (6b). However, Sprouse et al. (2016) found the same adjunct penalty with a 'referential' *wh*-phrase such as 'which briefcase' in (7d), while the corresponding relative clause (8b) did not show any such penalty.

(10) a. Les premiers étaient des coopératives dont [les membres] the first were some cooperations of which the members lopins de terre. exploitaient sous forme privée des exploited under form private some parts of land (FTB) (Abeillé et al. 2020) 'The first ones were cooperations of which the members exploited on a private basis some land parts' h

[...] *un* des responsables, de ses amis, de qui of.the responsible, of friends, of who one his [le père] a ses entrées dans la police the father has his entries in the police (Frantext, Garat, 2010) (Abeillé and Winckel 2020) 'one of the persons in charge, a friend of his, of whom the father has connections with the police'

No such cases were found with *wh*-questions. Comparing extraction from subjects and from objects, experimentally, we found a subject penalty with *wh*-questions, but not with relative clauses (Abeillé et al. 2020; Winckel 2024), which confirms the contrast found in French corpora. Experiments on Italian (Sprouse et al. 2016) and Spanish (López Sancio and Laka 2019; Pañeda and Kush 2021) found the same difference between constructions. In English also, Abeillé et al. (2020) found a subject penalty with *wh*-questions (10c), and not in relative clauses (10a), with pied-piping:

- (10) a. The dealer sold a sportscar, of which [the color] delighted the baseball player because of its surprising luminance.
 - b. The dealer sold a sportscar, of which the baseball player loved [the color] because of its surprising luminance.
 - c. ?? Of which car did [the color _] delight the baseball player?
 - d. ? Of which car did the baseball player love [the color _]?

To account for this contrast, we proposed the focus-background conflict (FBC) constraint: it is unfelicitous to focus part of a backgrounded constituent (Abeillé et al. 2020). If the subject is (by default) more topical than the object, hence more backgrounded, a penalty arises when focalizing part of it with a *wh*-question, but not when relativizing part of it, since relative clauses are not focalizing constructions.

Recent empirical results did not confirm structural approaches to locality constraints, as configurations banning any kind of *wh*-movement. As Liu et al. (2022) have it: "the current set of results from corpus studies and experiments suggests a more nuanced view, with gradient acceptability and cross-construction variation, suggesting an important role of discourse, frequency, and memory constraints in explaining island phenomena as well as their counterexamples. We speculate that all of these island structures may eventually be fully explained in terms of discourse, frequency, and memory constraints."

What is left for syntactic theory? As Featherston (2007) has it: "A theory is always a theory about a set of data. If the data can be accounted for by known surface factors (weight, focus, ...) without additional theory, then no additional theory is necessary." Following Chaves (2012), Liu et al. (2022) suggested that only coordinations (1c) are best explained by a syntactic/semantic constraint, if there must be at least two coordinates. Chaves and Putnam (2021) also suggested a morphosyntactic constraint on complementizers (and prepositions in many languages), so they cannot be stranded: *That she will come, I am sure. * She will come, I am sure that* _.

3 Ellipsis and the empirical turn

Ellipsis (a reduced syntax with a clausal meaning) comes in various types: VP ellipsis (with an auxiliary) (11a), gapping (verb ellipsis with two remnants) (11b), stripping (with an argument remnant) (11c), sluicing (with a *wh*-remnant) (11d) (Ross 1969), right-node-raising (RNR), or peripheral ellipsis (11e) (Postal 1974):⁴

- (11) a. John drank scotch [but Bill did not].
 - b. John drank scotch [and Bill bourbon].
 - c. John drank scotch [and Bill too].
 - d. Someone drank scotch but I don't know [who].
 - e. [John bought] and Bill drank scotch.

Two main approaches have been proposed: structural ones positing a full underlying syntax (Merchant 2001; Ross 1969) and referential ones positing a direct interpretation (Culicover and Jackendoff 2005; Ginzburg and Sag 2000). For the former, elliptical clauses have the same distribution as their non-elliptical counterparts, while the latter allow for various mismatches, and considers that ellipsis is governed by the same mechanisms of discourse reference that support other anaphoric expressions (Hardt 1993; Kehler 2000; Poppels 2022, a.o.).

Hankamer and Sag (1976) proposed an influential distinction between ellipsis ('surface anaphora') and (pronominal) anaphora ('deep anaphora'): while the latter may allow for exophoric (non-linguistic) antecedents and various mismatches with the antecedent, the former may not. They thus claimed exophora (12a) and voice mismatch (12b,c) to be reserved to VP anaphora and to be impossible with VP ellipsis (12a,b) and sluicing (12c):

⁴ Corpus data have provided an even wider variety, see for example Ginzburg and Fernández (2002)'s classification, based on the British National Corpus (BNC).

- (12) a. [Hankamer attempts to stuff a 9-inch ball through a 6-inch hoop]Sag: It's not clear that you'll be able to do it. / # that you'll be able to.
 - b. The oats had to be taken down to the bin, so Bill did it./ *so Bill did.
 - Maggie was made fun of, but she couldn't see who did it. / # who. (Hankamer and Sag 1976)

Kempson et al. (2016) proposed 'dynamic syntax' (based on under-specification and co-construction of utterances in interaction), to give a unified account of fragments, disfluencies, anaphora and ellipsis, found in spoken corpora. Kobele and Merchant (2016) pointed out that "the question of whether to analyze a particular sentence fragment in terms of ellipsis should be influenced by empirical considerations."

For most ellipsis types, a number of corpus-based and experimental studies have shown that elliptical constructions do not always have a verbal antecedent, and allow for more mismatches than previously thought when they do.

3.1 Exophoric uses of ellipsis

Examples of exophoric uses have been provided for sluicing (13a) and VPE (13b):

- (13) a. [in a lift] *Which floor?* (Ginzburg and Sag 2000)
 - b. [A standing beside son with surfboard in hand checking the waves] A: *I wouldn't if I were you. Notice the red flag.* (Kempson et al. 2016: 261)

A series of corpus studies highlighted that such exophoric uses require special contexts: an alternative between p and not(p) available in the situation, which makes a salient QUD for VPE (14a) (Miller and Pullum 2013), a proposition with contrasting subjects (or topics) for stripping (14b) (Abeillé and Kim 2022).

- (14) a. The aisles at the Lakewood Wal-Mart are surprisingly packed at 11 p.m.
 Can we? Can we? Vanessa tugs at her mother, pointing to a rack of
 "Lady and the Tramp" DVDs. (COCA) (Miller and Pullum 2013)
 - b. At the Riverside Caf, Kyle orders something that sounds like chicken-frystick and I nod and say, *"Me too."* (COCA 2019 FIC) (Abeillé and Kim 2022)

As pointed out by Ginzburg and Kim (2023), exophoric exclamative sluices are the most frequent case (in the BNC and the COCA) (15a), often without a verbal equivalent, and it is the same for emotive interrogative sluices (15b) (Kim and Kim 2023).

- a. Doctor: It's a limited time in your life, and you will get through it.
 Ms. Levy: What a relief (COCA 1993 SPOK) (Ginzburg and Kim 2023)
 - b. Suddenly a human hand punches through the wall. Chris: *What the hell ?* (COCA 2002 FIC) (Kim and Kim 2023)

3.2 Ellipsis with non-verbal antecedents

Various ellipsis types (with a sentential content) also allow for non-verbal antecedents: VPE (Miller and Hemforth 2024), stripping (Abeillé and Kim 2022), (16b) meaning 'did you have a hard day too?', and sluicing (Poppels 2022), (16c) meaning 'when he will be impeached'.

- (16) a. Mubarak's survival is impossible to predict and, even if he does, his plan to make his son his heir apparent is now in serious jeopardy.
 (COCA: CBS Evening News) (Miller and Hemforth 2024)
 - b. Dure journée. Toi aussi? (Frantext: Bouillier 2018) (Abeillé and Kim 2022: 17) 'Hard day. You too?'
 - c. *Regarding Trump's impeachment, the only question is when.* (Poppels 2022: 42)

As pointed out by Miller and Hemforth (2024), cases such as (16a) are rare because they require what they call a 'polar noun' like *survival*: a noun that in this context is interpreted as a polar question ('whether he will survive or not').

3.3 Voice mismatch and ellipsis

Thought to be ungrammatical by Hankamer and Sag (1976) (12b), voice mismatch is in fact possible for VPE in the right discourse context (corpus data by Hardt 1993, experiments by Kertz 2013, Poppels and Kehler 2019 a.o.):

- (17) a. The information could have been released but Gorbachev chose not to (release the information). (Hardt 1993)
 - b. Actually I have implemented it [=a computer system] with a manager, but it doesn't have to be (implemented with a manager). (Kehler 2000: 549)
 - c. This problem was to have been looked into, but obviously nobody did (look into the problem). (Vincent Della Pietra, in conversation, cited in Kehler 1993)

Are such examples grammatical? In a series of experiments, Arregui et al. (2006) found that passive-active mismatch (18a) was rated higher than active-passive mismatch (18b) and suggested that voice mismatch is not grammatical but can be 'repaired': passive VP are recycled (by the human parser) as active antecedents for ellipsis, since we are more likely to misrecall a passive as an active than vice versa.

14 — Abeillé

- (18) a. The dessert was praised by the customer after the critic did already. (Arregui et al. 2006)
 - b. *The customer praised the dessert after the appetizer was already.* (Arregui et al. 2006)

But such experimental items do not follow the discourse conditions brought to light by Kertz (2013) and Miller and Pullum (2013): VPE should either have subject-focus (different subjects which are contrastive topics) or auxiliary-focus (with the same subject). As shown by Kertz (2013) in another series of experiments, voice mismatch is more felicitous with auxiliary focus (17); otherwise, a penalty arises since the dessert and the critic are not contrastive topics in (18a) nor are the customer and the appetizer in (18b). Topic continuity is a discourse condition also observed outside ellipsis, and there may also be other factors that make voice mismatch less felicitous than match: a preference to coordinate similar clauses (Frazier et al. 1984) or a passive ellipsis penalty (Poppels and Kehler 2019).

To reconcile voice mismatch with identity theories of ellipsis, Merchant (2013) proposed a more abstract syntax, with a VoiceP, that matches both passive and active VPs. This hypothesis predicts that clausal ellipsis (above VoiceP) should not allow for voice mismatch, thus prohibiting it with sluicing (19a) and gapping (19b) for instance.

- (19) a. *Joe was murdered, but we don't know who murdered him.
 - b. *Some brought roses and lilies were brought by others.

But it also predicts pseudogapping to disallow voice mismatch, contrary to fact. Pseudogapping (Levin 1986) is a construction sharing properties with both VPE and gapping. Corpus studies (Miller 2014) show that pseudogapping mostly occurs in comparative contexts, not in coordination, with a preference for the same subject (20a). In the right context, it may also allow for voice mismatch (20b):

- (20) a. It hurt me [as much as it did her]. (COCA) (Miller 2014)
 - b. *These savory waffles are ideal for brunch, served with a salad* [<u>as you</u> would a quiche]. (COCA Mag) (Miller 2014)

Voice mismatch was also found possible for Right-node-raising (RNR), which is another case of clausal ellipsis: in French, Abeillé et al. (2023) found corpus examples such as (21a). In two experiments, they also found mismatch to be as acceptable as match (21b) in the appropriate semantic context (with the same subject and contrastive agents).

(21) a. Ce pharmacien doit des explications à ceux qui se sont ou qui ont été mobilisés pour lui. (2013/10/03, www.ipreunion.com) (Abeillé et al. 2023) 'This pharmacist owes explanations to those who have themselves or who have been rallied for him.'

b. Ce pharmacien doit des explications à ceux qui se sont ou qu'on a <u>mobilisés pour lui</u>

'This pharmacist owes explanations to those who have themselves or one has rallied for him.'

3.4 Other cases of mismatch between ellipsis and antecedent

Different kinds of ellipsis allow for other mismatches. Polarity mismatch is possible with VPE (22a) (Merchant 2013), sluicing (22b) (Anand et al. 2021; Poppels 2022), gapping (22c) (Repp 2009), and stripping (22d) (Abeillé and Kim 2022). (22b) means 'why you did not do it by Dec 15'.

- (22) a. John didn't see anyone, but Mary did (see someone).
 - b. Senator McCain said, 'Do it by December 15 or explain why' (NYT, Anand et al. 2021)
 - c. Pete wasn't called by Vanessa but John (was called) by Jessie.
 - d. A: We reserve the right to delete any comment for any reason at any time.B: Me either. (COCA 2012 WEB) (Abeillé and Kim 2022)

RNR has a first elliptical element, followed by a full one, which serves as its antecedent (Chaves 2014). Corpus studies showed that examples such as (11e) are not representative; see RNR data from the Penn Treebank (PTB) in Chaves (2014) and French corpus data in Abeillé and Mouret (2010). Most of the RNR occurrences in the PTB are subclausal, targeting a verbal phrase (23a) or a nominal phrase (23b).

- (23) a. Motorola [either denied] or would not comment on the various charges. (wsj-28924) (Chaves 2014)
 - b. ... this was [a formal] or an informal dinner party? (swbd-132959) (Chaves 2014)

Polarity mismatch is possible with RNR (Cann et al. 2005; Shiraïshi et al. 2019), and experiments on French have shown it to be perfectly acceptable: in (24a) the full counterpart of the first clause would be 'has read most of my books' and in (24b) 'qu'elle ait une dernière adresse connue' ('whether she has a last known address').

- (24) a. John has read, but he hasn't understood, <u>any of my books</u>. (Kayne 1994: 146)
 - b. Les textes actuels permettent de citer à l'audience une personne, qu'elle ait ou qu' elle n' ait pas <u>de dernière adresse connue</u>. (Avis Conseil d'Etat, 2007/02/13)
 'The current texts allow one to cite in court a person, whether she has or she doesn't have any known address.' (Shiraïshi 2018)

Verb form mismatch was found as well in English and French, in large corpora (Ententen, COCA and frenten), contra Pullum and Zwicky (1986), and shown to be acceptable in experiments, without mismatch penalty: in (25a), the unelided form would be 'saved' and in (25b) 'impacté' ('impacted').

- (25) a. I think she's going to be remembered as a young girl who has, and who will, <u>save an awful lot of lives</u>. (news.bbc.co.uk/2/hi/entertainment) (Shiraïshi et al. 2019)
 - b. Parler de sujets scientifiques, des innovations qui ont ou qui Talking of subjects scientific, of.del.PL innovations Rel have or Rel vont impacter le quotidien du grand public.
 will impact the daily.life of.det.M.SG broad public
 'Talking about scientific topics, innovations that have or that will impact the daily life of the population at large.' (www.cnrs.fr/centre-est) (Shiraïshi et al. 2019)

Such mismatches challenge different analyses of RNR, such as raising (Postal 1974) or multidominance (McCawley 1988), because they assume the peripheral element to match both conjunct clauses. We propose lexeme identity and not word identity, as a condition for peripheral ellipsis (Shiraïshi et al. 2019).

A general picture for different ellipsis types is the following (Abeillé et al. 2023; Poppels 2022):

	VPE	sluicing	stripping	gapping	RNR
Exophora	+	+	+	-	-
Non-verbal antecedents	+	+	+	-	+
Voice mismatch	+	-	-	-	+
Polarity mismatch	+	+	+	+	+

A conclusion is that not all elliptical constructions obey the same constraints and that ellipsis may be a misnomer for different phenomena.

Gapping seems to be more constrained than other ellipsis types. Johnson (2009) has proposed a raising analysis based on VP coordination, which disallows embedding gapping (26a). However, a number of experiments have shown that the gapped clause can be embedded under non-factive predicates in various languages (26b) (Spanish, Romanian, and Persian) (Bîlbîie and de La Fuente 2019; Bîlbîie et al. 2021) and without a complementizer in English (26c) (Bîlbîie et al. 2023). Park (2016) provided some corpus examples (26d).

- (26) a. *Some had eaten mussels and she claims [that others shrimp]. (Johnson 2009: 293)
 - b. Pero el chico la ama y dicen [que ella a él]. (CORLEC)
 (Bîlbîie and de La Fuente 2019)
 But the boy her loves and say that she DOM him
 'But the boy loves her and they say she him.'
 - c. At the bar, Paul ordered a beer and I suspect [John a whisky].
 (Bîlbîie et al. 2023)
 - d. *Truth is YOU will be in a position to hire ME, before I, YOU.* (Park 2016: 300)

This is compatible with a construction-based analysis of gapping, appealing to a dedicated asymmetric coordination, that maps a headless structure to a clausal meaning (see a.o. Abeillé et al. 2014; Bîlbîie 2017; Culicover and Jackendoff 2005; Sag et al. 1985).

The use of corpus data and controlled experiments have shown more mismatches to be possible than previously thought, giving referential theories of ellipsis a clear empirical advantage over identity theories. While structural theories are asking whether ellipsis is governed by some yet-to-be-discovered identity condition, referential theories are asking about referent accessibility and, relatedly, why inferential reference resolution succeeds in some cases but not others.

As Poppels (2022) put it: "Structural accounts of ellipsis require various degrees of structural match (e.g. Merchant 2001, 2013), having abandoned a full structural identity requirement once assumed in proposals like Hankamer and Sag (1976)." Nethertheless, they assume that "syntax is the pivot between form and meaning" (Kobele and Merchant 2016). "By contrast, referential theories of ellipsis are based on the fundamental architectural assumption that ellipsis is governed by independently motivated machinery, namely the system of discourse reference, which makes referential theories inherently more parsimonious." (Poppels 2022)

4 Conclusions

Locality constraints and ellipsis, which have been at the heart of syntactic theories for decades, have been renewed, thanks to naturally occurring corpus data (with context) and controlled experiments. Both topics have been shown to be more permissive than previously thought and to be sensitive to processing and discourse factors. They have also been shown to come in different constructions, with different discourse conditions.

As Featherston (2007: 34) had it "Constraining theory by such standards as 'elegance' and 'economy' is hardly any constraint at all since we have so little idea how to identify elegance and economy in a theory, and even if we did, this would constrain theory to be elegant, rather than constraining it to be accurate. Data gathered under controlled conditions, on the other hand, is very restrictive and unyielding, but permits the strongest theory of all, that which accounts for the facts."

A welcome consequence of taking the empirical turn is not only a better theory but also that results become reproducible, and data are made available for further research.

References

Abeillé, Anne (ed.). 2003. Treebanks: Building and using parsed corpora. Dordrecht: Kluwer.

- Abeillé, Anne & Jong-Bok Kim. 2022. *Me too* fragments in English and French: A direct interpretation approach. *The Linguistic Review* 39. 495–524.
- Abeillé, Anne & François Mouret. 2010. Quelques contraintes sur les coordinations elliptiques en français. *Revue de Sémantique et de Pragmatique* 24. 177–207.
- Abeillé, Anne & Elodie Winckel. 2020. *dont* and *de qui* relative clauses: an empirical study. *Journal of French Language Studies* 30. 275–300.
- Abeillé, Anne, Gabriela Bîlbîie & François Mouret. 2014. A Romance perspective on gapping constructions. In Hans Boas & Francisco Gonzalvez Garcia (eds.), *Romance in construction grammar*, 227–267. Amsterdam: John Benjamins.
- Abeillé, Anne, Barbara Hemforth, Elodie Winckel & Edward Gibson. 2020. Extraction from subject: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204. https://doi.org/10.1016/j.cognition.2020.104293.
- Abeillé, Anne, Barbara Hemforth & Aoi Shiraishi. 2023. Voice mismatch and contrast in French peripheral ellipsis. *Journal of Linguistics* 59. 361–388.
- Anand, Pranav, Daniel Hardt & James McCloskey. 2021. The Santa Cruz sluicing data set. *Language* 97. e68–e88.
- Arregui, Ana, Clifton Charles, Lyn Frazier & Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55. 232–246.
- Bîlbîie, Gabriela. 2017. Une grammaire des constructions elliptiques. Berlin: Language Science Press.
- Bîlbîie, Gabriela & Israël de la Fuente. 2019. Can gapping be embedded? Experimental evidence from Spanish. *Glossa* 4(1). 110. 1–39.
- Bîlbîie, Gabriela, Israël de la Fuente & Anne Abeillé. 2021. L'ellipse verbale enchâssée dans les langues romanes. *Langages* 223. 61–80.
- Bîlbîie, Gabriela, Israël de la Fuente & Anne Abeillé. 2023. Factivity and complementizer omission in English embedding gapping. *Journal of Linguistics* 59. 389–426.

- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2015. *Lexical functional syntax*. Oxford: Wiley Blackwell.
- Cann, Ronnie, Ruth Kempson, Lutz Marten & Masayuki Otsuka. 2005. Right node Raising, coordination and the dynamics of language processing. *Lingua* 115. 503–525.
- Chaves, Rui P. 2012. On the grammar of extraction and coordination. *Natural Language and Linguistic Theory* 30. 465–512.
- Chaves, Rui P. 2013. An expectation-based account of subject islands and parasitism. *Journal of Linguistics* 49. 285–327.
- Chaves, Rui P. 2014. On the disunity of right-node raising phenomena: Extraposition, ellipsis, and deletion. *Language* 90. 834–886.
- Chaves, Rui P. & Jerry E. Dery. 2019. Frequency effects in subject islands. Journal of Linguistics 55. 475–521.
- Chaves, Rui P. & Michael Putnam. 2021. Unbounded dependency constructions. Oxford: Oxford University Press.
- Chomsky, Noam. 1973. Conditions on transformations. In Steve Anderson & Paul Kiparsky (eds.), A *Festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart, & Winston.
- Chomsky, Noam. 1986. Barriers. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory*, 133–165. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 1996. 'Long' Wh-movements and referentiality. In Robert Freidin (ed.), *Current issues in comparative grammar*, 226–248. Dordrecht: Kluwer.
- Culicover, Paul & Ray Jackendoff. 2005. Simpler syntax. Oxford: Oxford University Press.
- den Dikken, Marcel. 2007. Phase extension contours of a theory of the role of head movement in phrasal extraction. *Theoretical Linguistics* 33. 1–41.
- Erteschik-Shir, Naomi. 1973. On the nature of island constraints. Cambridge, MA: MIT PhD thesis.
- Erteschik-Shir, Naomi & Shalom Lappin. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6. 41–86.
- Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33. 269–318.
- Frazier, Lyn, Lori Taft, Tom Roeper, Charles Clifton, Jr. & Kate Ehrlich. 1984. Parallel structure: A source of facilitation in sentence comprehension. *Memory and Cognition* 12. 421–430.
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. Cognition 68. 1–76.
- Gibson, Edward & Eve Fedorenko. 2010. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences* 14. 233–234.
- Ginzburg, Jonathan & Raquel Fernández. 2002. A corpus study of non-sentential utterances in dialogue. *Traitement Automatique des Langues* 43. 12–43.
- Ginzburg, Jonathan & Jong-Bok Kim. 2023. Exclaiming non-sententially: Exclamative sluices and the (nonelliptical) nature of ellipsis constructions. *Glossa* 8(1).
- Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations: The form, meaning, and use of English interrogatives*. Stanford, CA: CSLI Publications.
- Godard, Danièle. 1988. La syntaxe des relatives en français. Paris: Editions du CNRS.
- Godard, Danièle & Ivan A. Sag. 1996. Quels compléments de nom peut-on extraire en français? *Langages* 122. 60–79.
- Goldberg, Adele. 2013. Backgrounded constituents cannot be extracted from. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 221–238. Cambridge: Cambridge University Press.
- Hankamer, Jorge & Ivan A. Sag. 1976. Deep and surface anaphora. Linguistic Inquiry 7. 391-428.

- Hardt, Daniel. 1993. Verb phrase ellipsis: Form, meaning, and processing. University of Pennsylvania PhD. thesis.
- Hofmeister, Philip & Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86. 366–415. Huang, C-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. MIT PhD. thesis.
- Johnson, Kyle. 2009. Gapping is not (VP)-ellipsis. Linguistic Inquiry 40. 289-328.
- Kayne, Richard S. 1994. The antisymmetry of syntax. Cambridge, MA: The MIT Press.
- Kehler, Andrew. 1993. The effect of establishing coherence in ellipsis and anaphora resolution. *ACL* 31. 62–69.
- Kehler, Andrew. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23. 533–575. Keller, Franck. 2000. *Gradience in grammar*. University of Edinburgh PhD thesis.
- Kempson, Ruth, Ronnie Cann, Eleni Gregoromichelaki & Stergios Chatzikyriakidis. 2016. Language as mechanisms for interaction. *Theoretical Linguistics* 42. 203–276.
- Kertz, Laura. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89. 390–428. Kim, Jungsoo & Jon-Bok Kim. 2023. Aggressively non-D-linked construction and ellipsis: A direct
 - interpretation approach. Journal of Linguistics 59. 257–291.

Kluender, Robert. 1991. *Cognitive constraints on variables in syntax*. University of California, San Diego PhD thesis.

Kobele, Gregory M. & Jason Merchant. 2016. The dynamics of ellipsis. *Theoretical Linguistics* 42. 291–296. Kobzeva, Anastasia, Charlotte Sant, Parker Robbins, Myrte Vos, Terje Lohndal & Dave Kush. 2022.

- Comparing island effects for different dependency types in Norwegian. *Languages* 7. 197.
- Kush, Dave, Terje Lohndal & John Sprouse. 2017. Investigating variation in island effects. Natural Language and Linguistic Theory 36. 1–37.

Levin, Nancy S. 1986. Main-verb ellipsis in spoken English. New York: Garland.

- Liu, Yingtong, Elodie Winckel, Anne Abeillé, Barbara Hemforth & Edward Gibson. 2022. Structural, functional and processing perspectives on linguistic islands effects. *Annual Review of Linguistics* 8. 495–525.
- López Sancio, Sergio & Itziar Laka. 2019. Dependency type modulates island effects: Evidence from Spanish. *Presentation at the 93rd Annual Meeting of the Linguistic Society of America*. New York.

McCawley, James D. 1981. The syntax and semantics of English relative clauses. Lingua 53. 99-149.

- McCawley, James D. 1988. *The syntactic phenomena of English*. Chicago: University of Chicago Press. Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and identity in ellipsis*. Oxford: Oxford
 - University Press.
- Merchant, Jason. 2013. Voice and ellipsis. Linguistic Inquiry 44. 77-108.
- Miller, Philip. 2014. A corpus study of pseudogapping and its theoretical consequences. *Empirical Issues in Syntax and Semantics* 10. 73–90.
- Miller, Philip & Barbara Hemforth. 2024. Verb phrase ellipsis with nominal antecedents: The case of polar nouns. *Glossa* 9.
- Miller, Philip & Geoffrey K. Pullum. 2013. Exophoric VP ellipsis. In Philip Hofmeister & Elisabeth Northcliffe (eds.), *The core and the periphery, data-driven perspectives on syntax inspired by I. Sag*, 5–32. Stanford, CA: CSLI Publications.
- Müller, Christine & Clara U. Eggers. 2022. Island extractions in the wild: A corpus study of adjunct and relative clause islands in Danish and English. *Languages* 7. 125.
- Müller, Stefan, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.). 2021. *Head-driven phrase structure grammar*. Berlin: Language Science Press.

Nivre, Joachim, Marie-Catherine De Marneffe, Filip Ginter, Yoad Goldberg, Jan Hajic,

Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty

& Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings* of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 1659–1666.

Park, Sang-Hee. 2016. Towards a QUD-based analysis of Gapping constructions. *PACLIC* 30. 297–305. Pañeda, Claudia & Dave Kush. 2021. Spanish embedded question island effects revisited: An experimental

study. Linguistics 60. 463-504.

Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Stanford, CA: CSLI Publications. Poppels, Till. 2022. Explaining ellipsis without identity. *The Linguistic Review* 39. 341–400.

Poppels, Till & Andrew Kehler. 2019. Reconsidering asymmetries in voice-mismatched VP-ellipsis. *Glossa* 4(1), 60.

Postal, Paul M. 1974. On raising. Cambridge, MA: MIT Press.

- Pullum, Geoffrey K. & Arnold M. Zwicky. 1986. Phonological resolution of syntactic feature conflict. Language 64. 751–773.
- Repp, Sophie. 2009. Negation in gapping. Oxford: Oxford University Press.

Rizzi, Luigi. 1982. Issues in Italian syntax. Dordrecht: Foris.

- Ross, John R. 1967. Constraints on variables in syntax. MIT PhD thesis.
- Ross, John R. 1969. Guess who? CLS 5. 252-286.
- Sag, Ivan A., Gerald Gazdar, Tom Wasow & Steven Weisler. 1985. Coordination and how to distinguish categories. *Natural Language and Linguistic Theory* 3. 117–171.
- Shiraïshi, Aoi. 2018. Discordances dans l'ellipse périphérique en français. Université de Paris PhD thesis.
- Shiraïshi, Aoi, Barbara Hemforth, Philip Miller & Anne Abeillé. 2019. Verbal mismatch in right-node raising. Glossa 4(1). 114.
- Sprouse, John, Ivan Caponigro, Ciro Greco & Carlo Cechetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34. 307–344.
- Tellier, Christine. 1990. Subjacency and subject condition violations in French. *Linguistic Inquiry* 21. 306–311.
- Truswell, Robert. 2011. Events, phrases and questions. Oxford: Oxford University Press.
- Wexler, Kenneth & Peter W. Culicover. 1980. Formal principles of language acquisition. Cambridge, MA: MIT Press.
- Winckel, Elodie. 2024. *French subject islands: Empirical and formal approaches*. Berlin: Language Science Press (to appear).