Parsing out in English and Vietnamese

Nigel Duffield
Department of English and American Literature and Language
Konan University
Kobe, Japan
duffield@konan-u.ac.jp

***Abstract*—Here I outline an alternative approach to two types of minimal contrast in sentence acceptability that have previously been described (and explained) in terms of abstract formal grammatical mechanisms, namely, *that*-trace effects in English, and *wh*-island effects in Vietnamese. This alternative treatment—parsing out—is argued to be preferable to standard analyses, inasmuch as it makes use of independently needed parsing principles to yield equivalent or even superior empirical coverage.**

Keywords—Source Ambiguity, cross-linguistic variation, *that-trace effects; Wh-island effects, Parsing Out

I. Introduction

Chomsky noted in 1977: 'We may make an intuitive judgment that some linguistic expression is odd or deviant. But we cannot in general know, pre-theoretically, whether this deviance is a matter of syntax, semantics, pragmatics, belief, memory limitations, style, etc. [1].' This frequently cited observation, now over forty years old, gets to the heart of a debilitating problem in generative research: we know that speakers of a language are able to judge some utterances as being more acceptable than others, but we don't know what determines these judgments. Moreover, in spite of what Chomsky implies, this is not only *pre*-theoretical, it is also a *post*-theoretical problem: it persists long after we have developed theoretical tools to analyse it [2-5]. Indeed, if anything, the problem of explaining unacceptability in a particular case becomes more acute after a grammatical analysis has been proposed, when the cachet of a fashionable treatment may lead us to disregard other diagnoses—especially, the idea that the unacceptability is not due to grammatical factors at all.

Whilst there are many possible non-syntactic explanations for the relative unacceptability of an utterance in a given context—lexical, pragmatic, dialectal, (?)registral, and so forth—the alternative explored in this talk relates to *parsing* in spoken as well as written language comprehension: that is, to whether the perceived deviance of a particular utterance is due to formal featural properties of the analyzed sentence (grammar) or to the parse assigned to the utterance string. This choice between syntax and processing has recently been dubbed the 'Source Ambiguity problem': in a series of papers, Ivan Sag and his co-authors have suggested that many paradigm instances of putatively syntactic anomalies are better explained in terms of an interaction between independently-motivated processing factors—'super-additive factors'—than as resulting from purely formal constraints (such as the ECP, or the Subjacency condition, in earlier versions of generative theory). These anomalies include putative violations of the Superiority Condition, 'Complex NP'- and '*Wh*-island'-Constraints in (1b)-(3b), respectively [6-8] .

(1) a. Who *t* said what?
b. ?What did who *t* finish *t*?

(2) a. Which old piece of junk did you say that the guy had bought *t* ?
b. ??Which old piece of junk do you know the guy that bought *t* ?
c. ??Which old piece of junk do you know that the guy bought *t* ?

(3) a. What did Craig say [that doctors knew *t*]?
b. ?What did Craig wonder [if doctors knew *t*]?
c. ??What did Craig wonder [how doctors knew *t*]?

Notice in passing that (2b)—the alleged CNPC violation—appears no more deviant than the 'factive island' in (2c); thus, it is difficult to tell in this particular instance whether the unacceptability of (2b) compared to (2a) is due to the difference in verb type (think *vs.* know) or due to a difference in complement type (clausal *vs.* nominal complement). Likewise, in the case of indirect questions, both (3b) and (3c) are analyzed as grammatical (Subjacency) violations, yet—as is well-known—(2b) is markedly more acceptable than (2c) [9, 10], *cf.* [11].

Sag and his colleagues have stressed that the goal of the exercise is not to eliminate formal factors entirely from explanations of unacceptability, but rather to assign relative responsibility for deviance in a given case as honestly and economically as possible: if unacceptability in problematic utterances, such as the (b) examples above, can be shown to be due to an interaction among factors that explain increased processing difficulty in utterances that are hard to parse in spite of being grammatical, as this one is (!), then Occam's razor should lead us to prefer a processing diagnosis over a grammatical one.

In this paper, I first show how 'parsing out' can account for a frequently studied phenomenon in English, the so-called *that*-trace effect [12], [13]. I then consider how this analysis may be extended to explain apparent violations of island effects in Vietnamese, without recourse to abstract formal mechanisms [14], [15, 16].

II. Parsing out in English: **that*-trace effects

Consider the paradigm in (4), variants of which are used to illustrate a minimal contrast in English between 'long-distance' object *vs.* subject *wh*-extraction: complementizer deletion seems to be optional in the former case (a) *vs.* (b), but obligatory in the latter, where the complementizer is linearly adjacent to the trace of the extracted subject (hence, the term **that*-trace) [17].

(4) a. Who did you say Kate really likes *t* ?
b. Who did you say that Kate really likes *t* ?
c. Who did you say *t* really likes Kate?
d. *Who did you say *t* that really likes Kate?

Viewed in isolation, this paradigm indicates a dysfunctional—or at least *non-functional*— explanation for the unacceptability of (4d): certainly, the complementiser

seems to be no more disruptive to the process of obtaining a long-distance interpretation for the *wh*-phrase here than in the case of object extraction. What's more, by signalling the start of the subordinate clause, *that* apparently serves a useful parsing function, making the unacceptability of (*4d) even more surprising.

This is not the whole picture, however. In order to properly appreciate the nature of the **that*-trace effect requires that at least four additional factors be considered: when taken together, these considerations suggest that **that*-trace effects are largely, perhaps entirely, epiphenomenal. First is the fact that *that*-trace sequences are fully acceptable whenever *that* introduces a relative clause (that is to say, when it functions as a relativiser). Indeed, as the paradigm in (5) below demonstrates, relativiser *that* must *not* be deleted in subject relatives:[1] (5c) is unacceptable in most English varieties [18],[19, 20] .[2]

(5) a. Who is the person Kate really likes *t* ?
b. Who is the person that Kate really likes t ?
c. *Who is the person *t* really likes Kate?
d. Who is the person that *t* really likes Kate?

To account for the contradictory judgments in (4) and (5), those who wish to provide a grammatical account of **that*-trace are forced to stipulate that these two *that*s are (accidentally) homophonous lexical items, with distinct formal features; see [21], for example. This seems unintuitive, to say the least.

Second, it has been observed that **that*-trace effects, which are absent from some non-standard English varieties [22-24], are also reduced or eliminated in *all* varieties by the presence of an intervening adverbial phrase; this is demonstrated in (6) [25, 26]. This amelioration is unexpected on most formal analyses (in which string-adjacency is assumed to play no explanatory role).

(6) a. Lee forgot which dishes Leslie had said that *(under normal circumstances) should be put on the table.
b. Which kinds of medicines did you say that *(without proper testing) had been released onto the market?

Conversely, as is illustrated by the examples in (7), '*that*-trace' effects are observable even where the 'trace' position is a non-subject. In their discussion of ECP effects [27] acknowledge that for many speakers the presence of the complementiser inhibits construal of *wh*-adjuncts (*when, where, why,* etc.) with positions inside the embedded clause: as a result, whereas (7a) is ambiguous with respect to the interpretation of *why*, (7b) only permits a main clause interpretation. The blocking effects of *that* increase as more embedded clauses are added ((7b) > > (7c) > (7d) [> = 'more acceptable than']). The gradient nature of this effect again suggests a processing source for the unacceptability [3].

7. a. Why do you think Peter bought a new car *t* ?
b. ?Why do you think that Peter bought a new car *t* ?
c. ??Why do you think Peter said that Mary had bought a new car *t* ?
d. ??*Why do you think that Peter said that Mary had bought a new car *t* ?

A third consideration is that clausal complements with *that* are generally dispreferred when compared to the same complements without an initial complementizer, quite independently of the presence of any movement trace. This is evidenced by a Google NGram comparison of the strings '*who said say he/she/they*' vs. *'who said say that he/she/they'* which reveals a consistent preference for all of the former strings (i.e., without *that*) throughout the 20th century; see [16]. Thus, even before considering whether sentence (4d) merits an asterisk (*), it should be acknowledged that part of its deviance lies in the simple presence of the complementizer.

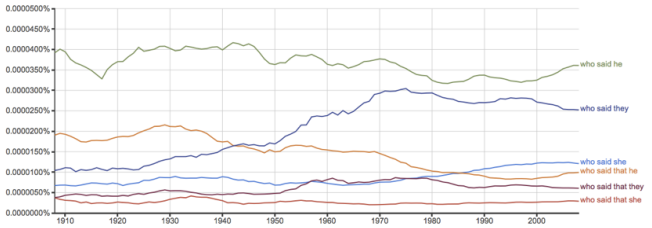

Figure 1. Complement Clauses with and without *that*

Finally—most relevantly to the present discussion—it is simply not self-evident that example (4c), the supposed baseline utterance, involves any long-distance extraction, since there exists a parallel analysis in which the string *did you say* is treated as a parenthetical expression. To see this more clearly, (4c) should be compared with the sentences in (8), where commas and/or extraposition of the parenthetical string to the utterance edge, reveal that *likes,* rather than *say,* is the main clause predicate; hence, no long-distance extraction is involved in these cases.

(8) a. Who, did you say, *t* really likes Kate?
b. Who *t* really likes Kate (, did you say)?

In (4d), by contrast, the complementizer *that* provides an unambiguous signal that complementation is involved, meaning that *did you say* cannot be parsed out as a parenthetical, nor can it be extraposed as a tag, either with *that* (9b), or without it (9b'):

(9) a. *Who, did you say, that *t* really likes Kate?
b. *Who *t* really likes Kate, did you say that?
b.' *Who that *t* really likes Kate, did you say?

A similar contrast obtains in the sentences in (10) *vs.* (11): in (10), it is the absence of inversion, rather than the presence of a complementizer, which prevents the second clause in the string from being interpreted as a main clause; conversely, in (11) the presence of inversion in the second clause allows this construal, once the medial AUX-NP-V string is parsed out as a parenthetical:

(10) a. *Where would you happen to know (that) I can find the toy department?
b. *What do you know (that) her maiden name was?
c.' ?Who can you tell me brought what?
c.'' *What can you tell me (that) who brought?

[1] This is true of object relatives in some contexts; see below.
[2] In the case of relative clauses, the trace positions are generally not thought to be linked to the head-noun directly, but to an abstract operator (OP) in the root specifier position of the relative clause. Nevertheless, the grammatical configurations are identical in the two cases, so that without further stipulation, the judgment pattern for (4) and (5) should also be the same.

(11) a. ?Where, would you happen to know, can I find the toy department?
a'. Where can I find the toy department, would you happen to know?'
b. ?What, do you know, was her maiden name??'
b.' What was her maiden name, do you know?'

In short, the English examples in (9)-(11) demonstrate that obtaining a matrix scope reading for an apparently embedded *wh*-expression does not necessarily imply the existence of any long-distance dependency: on occasion, the containing clause may simply be parsed out of the analysis. The examples also show that whether or not the first clause can be parsed out in this way depends on the conventional pragmatic uses of the particular matrix construction involved, as well as on the presence of markers of root status—in this case, SUBJ-AUX inversion—in the second clause.

The meaning of the expression 'parsing out', as used here, is quite intuitive. Nevertheless, for the sake of explicitness, it can be informally defined as the disregarding of elements in a utterance string having low informational content in a given context—such as tags, parentheticals, hesitations, conversational particles—whose omission allows for an efficient and grammatical syntactic analysis of that string.[3] In the case of 'long-distance' subject questions, parsing out will favour treating the first verb complex in the string as a parenthetical (e.g., *did you say, do you think*,) *unless* a complementizer is present (as in (4d): hence, the **that*-trace effect.

It turns out that this Parsing Out approach extends beyond direct questions to the other *that*-trace contexts identified in [13]: in virtually every case, the resulting sentence is acceptable where parsing out applies to a string without the complementizer element, and remains unacceptable where *that* appears:

(12) a. This is the person who I thought (*that) met Sue. [relativisation]
i. This is the person who ~~I thought~~ met Sue (, I thought).
ii. *This is the person who ~~I thought~~ that met Sue.
b. Mary, we think (*that) met Sue. [topicalization]
i. Mary, ~~we think~~ met Sue (, we think).
ii. *Mary, ~~we think~~ that met Sue.
c. It is Mary that we think (*that) met Sue. [cleft]
i. It is Mary that ~~we think~~ met Sue (, we think).
ii. *It is Mary that ~~we think~~ that met Sue.
d. More people like Mahler than we think (*that) like Bruckner. [comparative]
i. More people like Mahler than ~~we think~~ like Bruckner (, we think)
ii. ??More people like Mahler than ~~we think~~ that like Bruckner.
e. ?Bill will be easy for us to say (*that) met Sue. ['tough'-movement]
i. ?Bill will be easy for us to say met Sue.
ii. ??Bill will be easy for us to say that met Sue.

[3] As used here, the term 'parsing out' here bears close resemblance to [28]'s notion of 'annotative usage', in explaining factive islands.

The exception is the 'tough'-movement construction in (12e). Notice, however, that even without a complementizer (12e1), the sentence is less than perfect. Just as important, there is no 'first clause' string that can readily be parsed out to derive a simpler main clause, without losing significant propositional content.

III. Parsing out in Vietnamese: Illusory Islands

By applying the same combined strategy of attending to non-syntactic factors and/or parsing out what would otherwise be analyzed as main clauses, it is possible to explain apparent 'neutralization of island effects' in Vietnamese. The relevant data, which are supposed to illustrate (covert) extraction from Complex Noun Phrases, Sentential Subjects and Adjunct Clauses, respectively, were first presented in [14], and are reproduced in (13)-(15) below.

(13) a. *Tân sẽ chụp hình [NP con hổ [CP đã dọa ai]]?
Tan ASP catch picture CL tiger ASP scare who
'Tan will take a photo of the tiger that scared who?'/'*Who will John take a picture of the tiger that scared?'
b. Tân vừa chụp hình [NP con hổ [CP đã dọa ai]] *thế*?
Tan ASP catch photo CL tiger ASP scare who PRT
'Tan took a photo of the tiger that scared who?'

(14) a. *[CP Ai sẽ bỏ đi] làm mọi người bối rối?
who ASP leave make everyone embarrass
'That who will leave will make everyone embarrassed?'/'*Who that will leave will make everyone embarrassed?'
b. [CP Ai vừa bỏ đi] làm mọi người bối rối *thế*??
who ASP leave make everyone embarrass PRT
'That who left made everyone embarrassed?'

(15) a. *Tân sẽ thua cuộc [CP vì [TP ai làm hư xe của anh ta]] ?
Tan ASP lose event because ai make damage car POSS PRN
'Tan will lose the race because who will damage his car?'/'*Who will Tân lose the race because who will damage his car?'
b. Tân thua cuộc [CP vì ai làm hư xe của anh ta] *thế*?
Tan lose event because who make damage car POSS PRN PRT
'Tan lost the race because who damaged his car?'

According to [14]—henceforth *BT*—the near minimal contrasts between the (a) and (b) sentences in (13)-(15) above show that a direct *wh*-interpretation for the *wh*-indefinite element *ai* ('who') is unavailable if it is contained within a 'syntactic island'—the (a) examples—*unless* the utterance also contains the final discourse marker *thế*, as in the (b) examples. In other words, the addition of final *thế* suspends island effects.

That there is a contrast between the (a) and (b) examples is not in dispute here: native speakers that I have consulted are agreed that the addition of *thế* makes a direct question interpretation more available, *to the extent such an interpretation is available at all.* The question is why.

Before sketching a 'dumb alternative' to the formal account proposed by BT, it is necessary to control properly for other confounds in the data. The most obvious of these is

the fact that the unacceptable (a) examples above contain the pre-verbal future morpheme *sẽ*, whereas the acceptable (b) examples contain *either* the adverbial aspectual morpheme *vừa*—in (13a) and (14a), or else no preverbal element at all (15a). This difference would be innocuous if these elements all had the same syntactic status: BT assign the same gloss (ASP) in each case. They do not, however, for although it is conceivable that *vừa* is an aspectual morpheme, on a par with (perfect) *đã* and (progressive) *đang*, there is nothing aspectual about future *sẽ,* either in its distribution or in its interpretation […. Furthermore, for a variety of reasons, *sẽ* is excluded from a range of other contexts, including (future) *Yes-No* questions (16a), declarative 'future-perfect' constructions (16c), and emphatic contexts containing the assertion marker *có* (16c). None of these restrictions is observed with true aspectuals or with temporal adverbs [29-31].

(16) a. Năm sau vợ anh (*sẽ) (có) làm việc ở Paris không?
year next wife PRN FUT Q work be.loc paris NEG
'Will your wife work in Paris next year?'
b. Đến cuối năm nay, tôi (*sẽ) đã ra.trường.
arrive end year DEM PRN FUT ANT go.out.school
'I shall have graduated by the end of the year.'
c. Năm sau vợ anh sẽ (*có) làm.việc ở Paris.
year next wife PRN FUT ASR work be.LOC paris
'Your wife will work in Paris next year!'

The reasons for the unacceptability of *sẽ* in these latter examples are not of immediate concern. What *is* relevant is the fact that *sẽ* is independently problematic. So, it turns out that once *sẽ* is deleted from the canonical examples in (13)-(15) above, the contrast in acceptability between the (a) and (b) examples is greatly reduced, such that addition of *thế* in (b) only alters the availability of a direct question interpretation slightly: from strongly to weakly dispreferred, that is (?? *vs*. ?), as opposed to (* *vs*. ?).

A second difficulty is that, once again, the supposed base-line example in (17) below is not perfectly acceptable with the intended (long-distance) reading, *cf.* Fig 1 above: specifically, the direct question reading (ii) is less accessible than it should be if other factors were not at play, even though no possible island is involved in this sentence:

(17) Tân biết ai đi New York.
Tan know who go New York
(i) 'Tan knows for which person x, x went to NY.'
(ii) 'For which person x, Tan knows x went to NY?'

Quite aside from the basic issue of acceptability—for many native-speakers consulted, the sentence has only one acceptable interpretation under normal intonation and stress, namely, (i), as a declarative (indirect question)—it is likely that for those that allow it the direct *wh*-interpretation is *not* the result of covert movement: it could as readily arise from parsing out the main clause in the right conversational context; compare (10) *vs*. (11) above. On the other hand, the availability of a direct reading may well be diagnostic of an alternative parse: namely, a relative clause analysis. Notice that in (17) there is an ambiguity within the 'direct question reading,' between a 'short-distance' *direct object* interpretation 'who is the person x, Tan know x (& x went to NY)?' and a long-distance *subject* interpretation 'Who is the person x, such that Tan knows that x went to New York.' In both languages, this ambiguity arises due to the optionality of relativisers/complementisers (in English *that*, in Vietnamese *rằng*, *mà*, respectively).[4] Compare the English examples in (18):

(18) a. Who do you know [NP *t* [CP *(that/who) *t* went to New York last year]]? CP = Relative Clause
b. Who did you say [CP (*that/??who) *t* went to NY last year]? CP = Embedded Clause
c. Who did you know [NP *t* [CP that Amy met *t* in NY last year]]? CP = Relative Clause
d. Who did you know [CP that Amy met *t* in NY last year]? CP = Embedded Clause

The acceptability of (18a) with *that*, as well as the substitutability of *that* with *who* in the subordinate clause, clearly indicates that the bracketed constituent is parsed as a relative clause here; conversely, the '**that*-trace effect' in (18b)—and the awkwardness of replacing the complementiser with a relative pronoun—suggests, equally forcefully, that this is a case of 'long-distance' movement. Observe that in examples (18c) and (18d), which involve *object* construal within the subordinate clause, the same string is compatible with either a relative clause or complement clause analysis. Crucially, in (18c) the relativiser *that* is not omissible, in spite of this being an object relative construction; *cf.* (4b) above.

By applying similar tests in Vietnamese—*i.e.*, inserting optional complementisers and relativisers, or using verbs that unambiguously select for clausal complements—it is possible to test whether *wh*-expressions can in fact be extracted from embedded domains, and also to exclude a relative clause interpretation. So, the examples in (19) below show that complement clauses in Vietnamese can be unambiguously marked by an overt complementiser: *rằng* in the case of declarative complements, *liệu* for indirect questions. In such contexts, the prediction based on BT is that long-distance *wh*-extraction should be possible in principle, and that this direct question reading should be facilitated by the addition of final *thế*. (The notation here is intended to mark that *thế* is obligatory if the string is used to express a question.)

(19) a. Ông ta nói rằng công.việc nào thích hợp với ông ta *(thế).
PRN say COMP work WH suitable with PRN PRT
The man said that for any job x, x was suitable for him./For which job x, the man said that x was suitable for him?
b. Người đàn ông tự hỏi liệu có ai ở lại với ông ấy không (thế).
man self ask COMP exist COMP AI BE stay with PRN DEM NEG PRT
The man wonders whether there is any person x, x would stay with him/For which person x, the man wonders whether x would stay with him.

The relevant judgments do not square with BT's predictions, however: while the addition of *thế* to embedded declaratives containing *wh*-indefinites does facilitates a direct question reading to some extent—in fact (19a) cannot be interpreted as a question without *thế*—its application to

[4] This difficulty is compounded by the semantic ambiguity of the verbs Eng. *know,* Viet. *biết*: compare Spanish *conocer* ('know' = be familiar with a person, place) vs. *saber* ('know' = understand a proposition).

embedded questions (19b) is only optional, and has no scope-altering effect: (19b) remains preferentially interpreted as an indirect question (*ai* = 'anyone').

The contrast sentences in (20), containing the disambiguating complementizer *rằng* and the relative marker *mà,* respectively—but which are otherwise identical to (18)—provide clinching evidence against BT's proposal:

(20) a. ?Tân biết *rằng* ai đi New York (**thế*).
Tan know that ai go New York
'Tan knows that someone went to New York.'
(No long distance *wh*-reading available)
b. ?Tân biết ai *mà* đi New York (??*thế*).
Tan know AI REL go New York
i. 'Tan knows (someone) who went to New York.'
ii. ?'Who does Tan know *t* (who went to New York)'
(No long-distance *wh*-reading available)

There are three points to observe. First, neither sentence is wholly acceptable: once the string [*ai đi New York*] is clearly marked as a subordinate clause in (20a)—alternatively, once *ai* is shown to be outside the relative clause, as in (20b)—speakers consulted judge both utterances as less than perfect, even on a declarative reading. This suggests that (18) is not equivalent to (20) with complementiser deletion; rather, it involves parsing out. The second point is that *no direct question interpretation is available in either case*, no matter that these are non-island contexts. Finally, as indicated by the diacritics within the parentheses, the addition of *thế* only makes things worse: in (20a), adding *thế* excludes any grammatically acceptable analysis of the string; in (20b), the relative clause reading is further degraded (? > ??).

The preceding discussion of the baseline data, taken in conjunction with the fact that *thế* is not restricted to *wh*-environments, casts real doubt on BT's analysis of the function of *thế*. Indeed, it suggests the more radical possibility that *wh*-constituents are barred from embedded contexts in Vietnamese entirely. In other words, there are no long-distance questions. But, if there are no long-distance dependencies, then there can be no islands; and if there are no islands, there can be no island effects, for *thế* to suspend.

With this in mind, let us revisit BT's paradigm, using the revised examples in (21a-c), with judgments for the direct question reading (after deleting the future morpheme *sẽ*):

(21) a. Tân vừa chụp hình con hổ đã dọa ai *thế*?
Tan ADV catch photo CL tiger ASP scare who PRT
String: 'Tan took a photo of the tiger that scared who.'
Interpretation = Who was the person x, the tiger scared x (Tân took a picture of the tiger)?
b. Ai (vừa) bỏ đi làm mọi người bối rối *thế*?
who ASP leave make everyone embarrass PRT
String: 'who left made everyone embarrassed?'
Interpretation = Which person x, x left & x made everyone embarrassed?

c. Tân thua cuộc vì ai làm hư xe của anh ta thế ?
Tan lose event b/c who damage car POSS PRN PRT
String: 'Tan lost the race because who damaged his car?'
Interpretation = Who was the person x, x damaged Tân's car (caused Tân to lose the race)?

Assuming now that no syntactic islands are involved in any of these examples, the question is how to derive a direct question interpretation from a different analysis of the utterance string in these three cases. In each case, we will assume that the discourse function of final *thế*—like SUBJ-AUX inversion in English, see above—is to signal a reanalysis of the preceding string.

With respect to the alleged Complex NP Island in (21a), an obvious alternative parse is that this utterance involves two consecutive main clauses linked by a common topic *Tân* with shared *pro* subjects, as in (22) (subscripts indicate co-reference):

(22) [TOP Tân] [TP ~~*pro* vừa chụp hình con hổ~~] [TP *pro* đã dọa ai] *thế*?
'Tan (he) took a picture of someone; who(m) did he frighten?'

Under this analysis—which comports with most speakers' intuitions, namely, that Tân rather than the tiger is the one frightening someone—the *wh*-phrase *ai* occupies a matrix object position; no embedding is involved.[5] Without *thế,* the most natural parse remains a relative clause analysis, treating *chụp hình* ('take picture') as the matrix predicate. However, as the revised string in (23) this is not a possible source of long-distance extraction, with or without *thế*: if the clause is unambiguously marked as a relative, with *mà* and a post-VP adverbial, no extraction is possible:

(23) *Tân vừa chụp hình con hổ mà đã dọa ai kieu gầm gừ (thế)?
Tan ADV catch picture CLF tiger REL ASP scare who by growling PRT
'Tan took a photo of the tiger that scared someone by growling.' (No direct question interpretation)

As for the alleged sentential subject in (21b), here there are two plausible alternative analyses—diagrammed in (24) below—both of which allow for the *wh*-phrase *ai* to be treated as a matrix argument. The first possibility is that the initial clause contains a relative, rather than a sentential subject; the second is that the utterance is analyzed as two consecutive main clauses, the rightmost being treated a conversational run-on. In either case, we derive the intended interpretation without long-distance extraction.

(24) a. ?Ai [(mà) (vừa) bỏ đi] làm mọi người bối rối thế?
who REL ADV leave make everyone embarrass PRT
'?Who, that just left, made everyone embarrassed?'
b. ?Ai$_i$ [(vừa) bỏ đi] *pro*$_i$ làm mọi người bối rối ?thế?
who ASP leave make everyone embarrass PRT
'?Who just left, who/he made everyone embarrassed?'

Finally, the data in (25) clearly speak against the possibility that extraction is possible from true adjunct

[5] Under this analysis, it is also possible to interpret the string such that the tiger as the subject of the second clause, with topic switch.

clauses: in cases where the adjunct clause is clearly marked as such by a subordinating conjunction (*bởi vì, mặc dù*), and fronted away from *thế,* a direct question interpretation for *ai* is completely excluded:

(25) a. *Bởi vì ai đã mang vài thu đặc biệt, Tân nấu bữa tối thế?
because ai ASP bring what special Tan cook dinner PRT
'*Who because brought something special did Tan cook dinner?'

b. **Mặc dù* ai làm hư xe, Tân thắng cuộc thế?
though ai damage car, Tan win race PRT
'*Who even though damaged the car did Tan win the race?

Hence, it seems very unlikely that (21c) involves an adjunct island configuration. However, an alternative option presents itself, which only requires treating the second clause as a main clause, similar to what is observed in the English example in (26) below. Once more, no long distance extraction is involved.

(26) '~~Tan lost the race because~~…*who* damaged his car?'

So, what does *thế* do, if it has no formal role, and if—as suggested here—there are no islands in Vietnamese, for the very good reason that there are no long-distance dependencies of any kind? The simple answer, I suggest, is 'not very much'. Or at least, *thế* does no more than tag questions do in English. What these two devices share is that they are typically restricted to root environments, such that where they appear *atypically*, they coerce a matrix reading from a subordinate clause. Finally, then, compare the examples in (27): on a normal first pass reading, the addition of a tag which is incompatible with the root subject yields an anomalous judgment, but—if accompanied by different intonation—the tag can trigger a re-analysis of the utterance, so that the first clause is parsed out (my judgments):

(27) a. Alexa thinks the girls will win, ??won't they/??will they?
a.' ~~Alexa thinks~~ the girls will win, won't they/will they?

IV. Conclusion

Acceptability judgments are as imprecise as most medical symptoms considered in isolation. A sore throat tells the doctor very little: it could indicate a bacterial or viral infection, or it could be due to environmental toxins, or too much shouting the night before. Or it could be throat cancer. Without further investigation, proper diagnosis is impossible; it would be unwise to prescribe antibiotics, much less chemotherapy, even if these treatments have proven effective in some other patients with similar symptoms. And where further tests allow us to exclude a particular underlying condition, it is irresponsible to continue treatment.

Returning to linguistics, it has been argued that, just as is the case for *that*-trace effects in English, BT's analysis of putative island contexts in Vietnamese is based upon questionable assumptions about the source of deviance in the unacceptable cases, and at the same time misinterprets the ameliorating function of final *thế* in the acceptable ones. Once non-syntactic factors are controlled for, and alternative parses are taken into account, the likelihood that syntax plays any significant role in explaining the initial contrast is markedly reduced.

Acknowledgment

I am grateful to Trang Phan for her advice on the current paper, and to Tue Trinh and previous reviewers for their comments on earlier drafts and presentations.

References

[1] Chomsky, N., Essays on form and interpretation. 1977, New York: North-Holland.

[2] Newmeyer, F., Grammatical Theory: its limits and possibilities. 1983, Chicago: University of Chicago Press.

[3] Schütze, C.T., The empirical base of linguistics. 1996, Chicago: University of Chicago Press.

[4] Duffield, N., Implications of Competent Gradience. Moderne Sprachen, 2004. 45, pp 95-117.

[5] Culicover, P., Syntactic Nuts: hard cases, syntactic theory and language acquisition. 1999, Oxford: Oxford University Press.

[6] Hofmeister, P., L.S. Casasanto, and I.A. Sag, Islands in the Grammar? Standards of Evidence, in Experimental Syntax and the Islands Debate, J. Sprouse and N.A. Hornstein, Editors. 2014, Cambridge University Press: Cambridge.

[7] Hofmeister, P., T. F. Jaeger, I. Arnon, I.A. Sag and N. Snider., The source ambiguity problem: distinguishing the effects of grammar and processing on acceptability judgments. Language and Cognitive Processes, 2013. 28(1/2): p. 48-87.

[8] Sag, I.A., et al., Processing Accounts for Superiority Effects. 2008, Stanford University: Stanford CA.

[9] Kluender, R., Deriving island constraints from principles of predication, in Island Constraints: Theory, Acquisition and Processing, H. Goodluck and M. Rochemont, Editors. 1992, Kluwer Academic Publishers: Dordrecht & Boston. pp. 195-222.

[10] Kluender, R. and M. Kutas, Subjacency as a Processing Phenomenon. Language and Cognitive Processes, 1993. 8, pp. 573-640.

[11] Phillips, C., M. Wagers, and E. Lau, Grammatical illusions and selective fallibility in real-time language comprehension, in Experiments at the Interfaces. Syntax & Semantics (Vol. 37) J. Runner, Editor. 2011, Emerald: Bingley, UK. pp. 153-186.

[12] Perlmutter, D., Deep and Surface Structure Constraints in Syntax. 1971, New York: Holt, Rinehart & Winston.

[13] Pesetsky, D., Complementizer-trace effects (encyclopedia article). http://ling.auf.net/lingbuzz/002385 (LingBuzz), 2015.

[14] Bruening, B. and T. Tran, Wh-questions in Vietnamese. Journal of East Asian Linguistics, 2006. 15, pp. 319-341.

[15] Duffield, N., Illusory Islands: On 'Wh-questions in Vietnamese'. submitted.

[16] Duffield, N., Reflections on Psycholinguistic Theories: Raiding the Inarticulate. 2018, Cambridge: Cambridge University Press.

[17] Chomsky, N. and H. Lasnik, Filters and Control. Linguistic Inquiry, 1977. 8, pp. 425-504.

[18] Bresnan, J., Theory of complementation in English syntax. 1972, MIT.

[19] Doherty, C., Clauses without 'that': The Case for Bare Sentential Complementation in English [originally published 2000 by Garland Publishing]. 2013, Abingdon, Oxon and New York: Routledge (Taylor and Francis).

[20] Henry, A., Dialect Variation and Parameter-setting: a study of Belfast English and Standard English. 1995: Oxford University Press.

[21] Rizzi, L., Relativized Minimality. Vol. 16. 1990, Cambridge, Massachusetts: MIT Press.

[22] Salzmann, M., Häusler, J., Bader, M., Bayer, J., That-trace effects without traces. An experimental investigation, in Proceedings of the 42nd Annual Meeting of the North East Linguistic Society, S. Keine, Editor. 2013, GLSA: Amherst, MA. pp. 149-162.

[23] Sobin, N., The variable status of Comp-trace phenomena. Natural Language & Linguistic Theory, 1987. 5, pp. 33-60.

[24] Sobin, N., The Comp-Trace Effect, the Adverb Effect and Minimal CP. Journal of Linguistics, 2002. 38, pp. 527-560.

[25] Culicover, P., Evidence against ECP accounts of the *that*-trace effect. Linguistic Inquiry, 1993. 24, pp. 557-561.

[26] Douglas, J., Unifying the that-trace and anti-that-trace effects. Glossa: a journal of general linguistics, 2017. 2(1) 60, pp. 1–28.

[27] Aoun, J., Hornstein, N., Lightfoot, D., and Weinberg, A., Two Types of Locality. Linguistic Inquiry, 1987. 18, pp. 537-578.

[28] Oshima, D.Y., On Factive Islands: Pragmatic Anomaly vs. Pragmatic Infelicity., in JSAI, T. Washio, et al., Editors. 2006, Springer. pp. 147-161.

[29] Cao, X.H., Tiếng Việt. Mấy Vấn đề Ngữ Âm, Ngữ Pháp, Ngữ Nghĩa [Vietnamese: some issues in phonetics, grammar, semantics]. 2003, Đà Nẵng: Nhà xuất bản Giáo dục.

[30] Phan, T., Syntax of Vietnamese Aspect, 2013, University of Sheffield: Sheffield.

[31] Duffield, N., On what projects in Vietnamese. Journal of East Asian Linguistics, 2017. 26, pp. 351-387.