Structural difference and difficulty in translation and interpretation

Draft PhD thesis

Selim Earls
University of Louvain
2024

Annex I

Semantic parsing

Document version

Easy-to-read version available online at:
https://structural-difficulty-in-translation-and-interpretation.com/annexes/semantic-parsing/

Annex I

Semantic parsing

This annex describes a method for parsing complex sentences into component propositions and indicating the functional relations between those propositions. That method is then applied to each sentence in the corpus, to record values for the variables fed into the statistical analysis. The semantic parsing method is described briefly in section 3.1. The details of the method are presented in the following sections.

Contents

1. Propositions

1.1 Predicate-argument structure

Lyons (1995) sees propositions expressed in different syntactic forms as equivalent if they meet the same truth conditions. This study applies that notion of semantic equivalence to propositions establishing similar logical relations among similar entities as expressed in different languages. Propositions are displayed here in their surface syntactic forms, rather than being analyzed in terms of their truth value. The method used mirrors the traditional three-way syntactic classification of a subordinate clause as a relative, complement or adverbial clause – that is, as an argument, modifier or adjunct in relation to its parent. But it applies that classification to the semantic relations between propositions. Accordingly, each functionally subordinate proposition is labelled as a semantic argument, modifier or adjunct in relation to its parent. There's also a fourth label used, for a proposition expressing reported speech or thought.

This study differs from many analyses of propositions in a few respects. First, it operates on a larger scale, examining the relations between propositions rather than their internal logical content. Second, it defines "proposition" in a broader sense, including closed propositions (expressed in a finite clause, with the tense and all arguments specified or implied), as well as open propositions (expressed in a non-finite clause or nominal structure, with the tense or at least one argument missing). Third, the term "proposition" is used here to refer not only to an underlying semantic function, but also to any predicate-argument structure giving syntactic expression to a proposition. In these respects, this study uses the term in the same way as Larson (1984), who uses the proposition as a unit of analysis to examine the preservation of meaning in translation.

Let's take a look at the various syntactic forms which can be used to express propositions, and at the semantic relations between them. Syntactically, an event or situation can be described in a finite clause, a non-finite clause or a clause-like nominal structure.

A **finite clause** is a clause where the tense and all the arguments are known, as in (25).

(25) I love you.

A **non-finite clause** is a clause where the tense or at least one argument is unknown, as in (26) and (27).

(26) my loving you *(no tense)*

(27) loving you *(no tense, one argument missing)*

A **clause-like nominal structure** is a noun phrase derived from an underlying clause and expressing the same semantic relations between a predicate and its arguments, as in (28).

(28) my love for you

Each description of an event or situation is built around a semantic **predicate**, as highlighted in (29).

(29) I **love** you.

A predicate is the linguistic expression of a logical function. ("Predicate" is used here in this modern linguistic sense, not as in traditional grammar, meaning the part of a clause which is left when the subject is removed.) A predicate takes one or more entities as input values and yields an actual or potential truth value as an output. (The perspective and criteria for assessing truth values aren't relevant here.) Such a logical function can be represented as in (30).

(30) LOVE (I, you) = {true} or {false}

The input values of a logical function – the entities involved in the event or situation it describes – are semantic **arguments**, as highlighted in (31).

(31) **I** love **you**.

Any modifiers of the predicate are **adjuncts**, as highlighted in (32).

(32) I **truly** love you.

A semantic predicate can appear syntactically as various parts of speech, as shown in (33)-(36).

(33) I **know**. ***(verb)***

(34) She's an **expert**. ***(noun)***

(35) I'm **interested**. ***(adjective)***

(36) We're **inside**. (***preposition***)

The word "inside" in (36) would be called an "adverbial particle" in traditional grammar. But it can also be seen as an intransitive preposition – that is, a prepositional predicate without a complement.

The idea of some prepositional phrases having internal argument structure, with the preposition functioning as a predicate, is well established in modern syntactic theory. Jackendoff (1973: 347), Emonds (1985: 253), Huddleston and Pullum (2002: 600) and others discuss the distinction between a transitive preposition, which takes a complement, and an intransitive preposition, which has no complement. Grimshaw and Williams (1993) distinguish between a "semantic" preposition, which acts as a predicate, and a "grammatical" preposition, which doesn't. Merlo and Ferrer (2006: 347) use various tests to determine when a preposition predicates a separate property of its head. Such descriptions of the internal argument structure of some prepositional phrases can also be applied to some postpositional phrases (in languages with postpositions) and to some case phrases (in languages with case).

In some languages, like English, a finite clause needs a verb. So a predicate other than a verb in a finite clause is linked to its subject by a semantically empty verb, like "be." The various forms of "be" in (34)-(36) are examples.

The predicates highlighted in (33)-(36) are intransitive, taking no complement. All these types of predicate can also be transitive, taking a complement in addition to the subject, as in (37)-(40).

(37) I know **Japanese**.

(38) She's an expert **at show-jumping**.

(39) I'm interested **in linguistics**.

(40) We're inside **the haunted house**.

Many of these relations can also be expressed in nominal structures, as in (41)-(43).

(41) my knowledge of Japanese

(42) her expertise at show-jumping

(43) my interest in linguistics

Propositions like (41)-(43) are referred to in this study as "clause-like nominal structures." A finite clause, a non-finite clause and a clause-like nominal structure can all express the same semantic relations between a predicate and its arguments, even though their syntactic structures are different, as illustrated in figure 61.

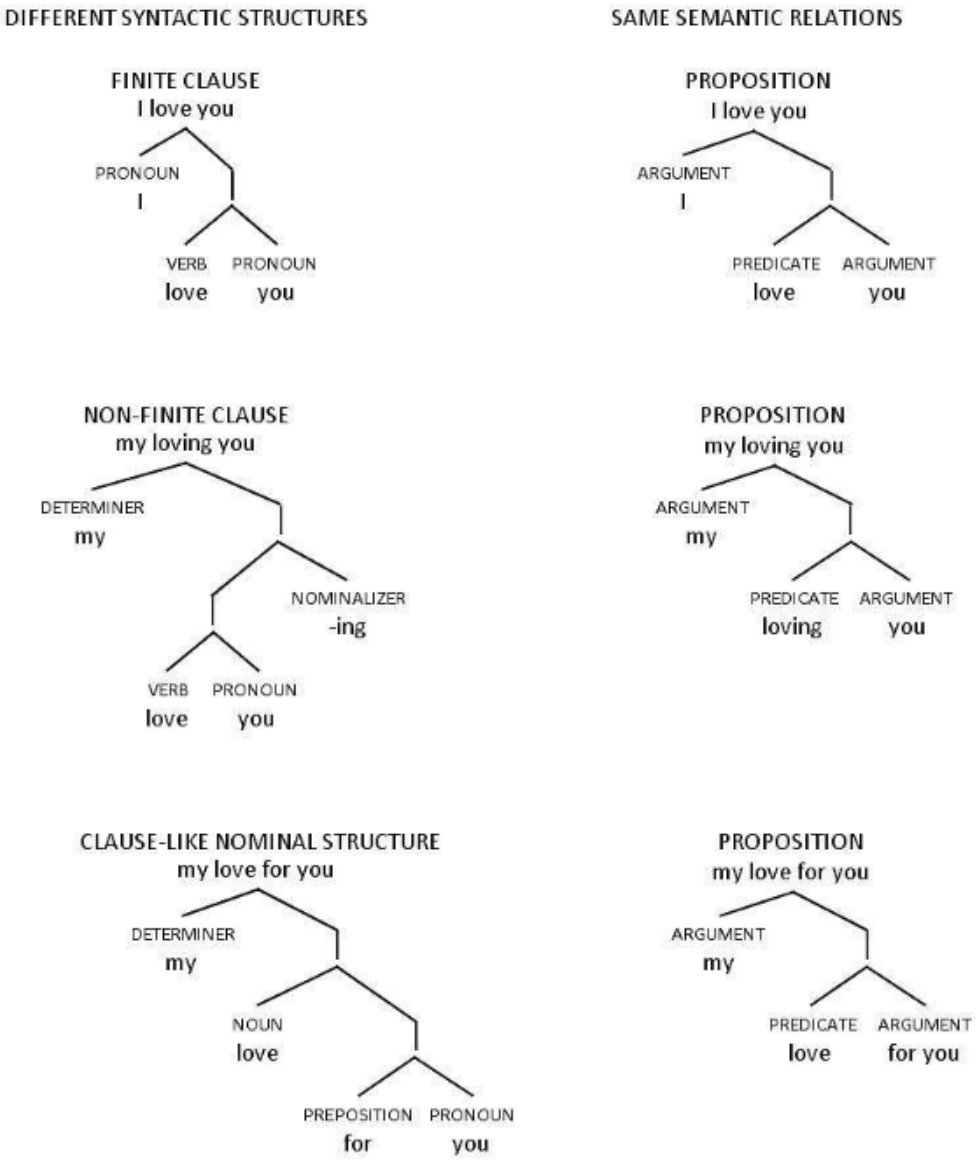

Figure 61
Different syntactic structures expressing the same semantic relations

A syntactic construction which describes an event or situation by establishing logical relations between a predicate and its arguments is a **predicate-argument structure**. This study uses the term "**proposition**" to refer to any such construction.

1.2 Relations between propositions

[This sub-section is included in the summary of the semantic parsing method in chapter 3, section 3.1.2. The figures are renumbered here.]

One proposition can be subordinate to another, parent proposition, as illustrated in figure 62.

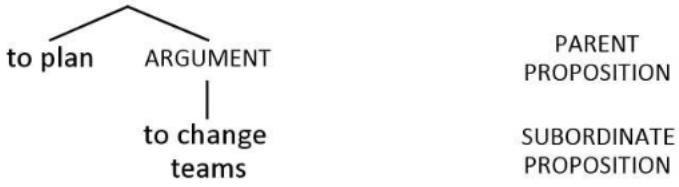

Figure 62
Parent and subordinate propositions

Each of the propositions in figure 62 can be expressed as a clause or as a clause-like nominal structure. The relations of semantic hierarchy between predicates and arguments in figure 63 are the same.

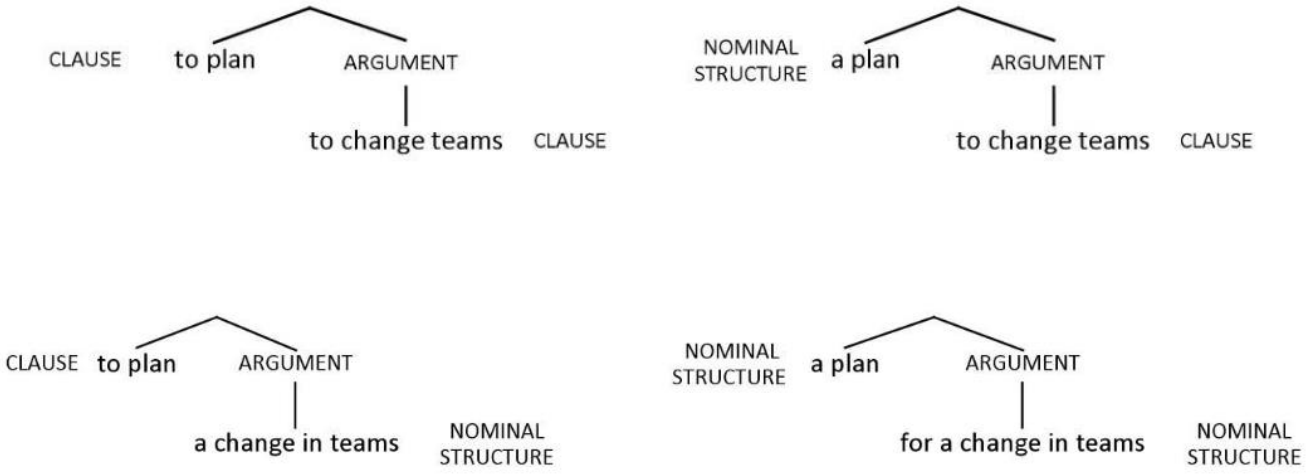

Figure 63
Parent and subordinate propositions with the same hierarchical relations

For simplicity and clarity, the parse trees used here include an overt link between propositions on the same branch as one of the propositions. If one proposition is subordinate to another, an overt link between them – like "for" in the last tree in figure 63 – is included on the same branch as the subordinate proposition.

The relations between a parent and a subordinate proposition can sometimes be recast with minimal differences in meaning from one syntactic form to another, as illustrated in figure 64.

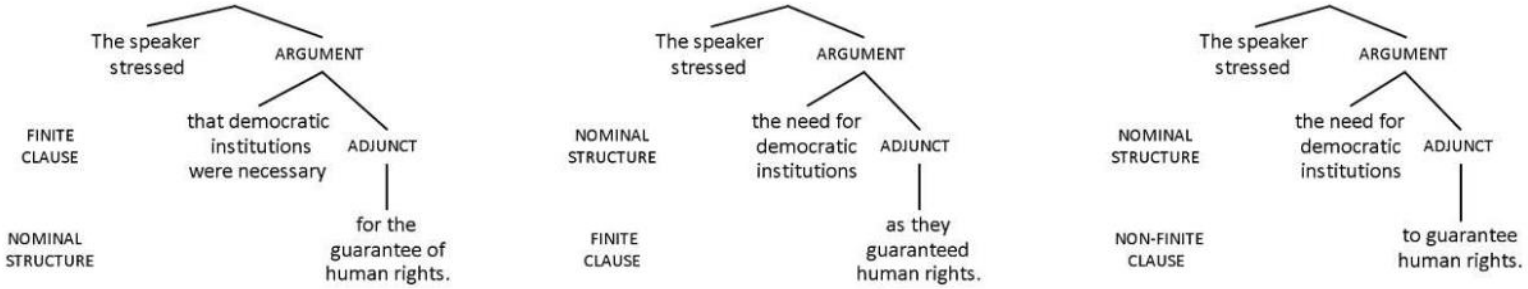

Figure 64
Propositions with similar hierarchical relations

This kind of switch is possible within a language, as in the examples in figures 63 and 64. It's also possible in translating or interpreting a message from one language to another. A method of analysis based on propositions, where clause-like nominal structures are treated in the same way as finite and non-finite clauses, reflects the semantic similarities between the different syntactic forms which various languages can use to describe an event or situation. It also helps highlight whether the hierarchical relations between propositions, whatever their syntactic form, are preserved in translation or interpretation.

Figure 65 shows a parse tree of the hierarchical relations between propositions in a complex sentence. The labels on the nodes indicate the semantic role of each subordinate proposition in relation to its parent. In this sentence, each subordinate proposition is a semantic argument of its parent.

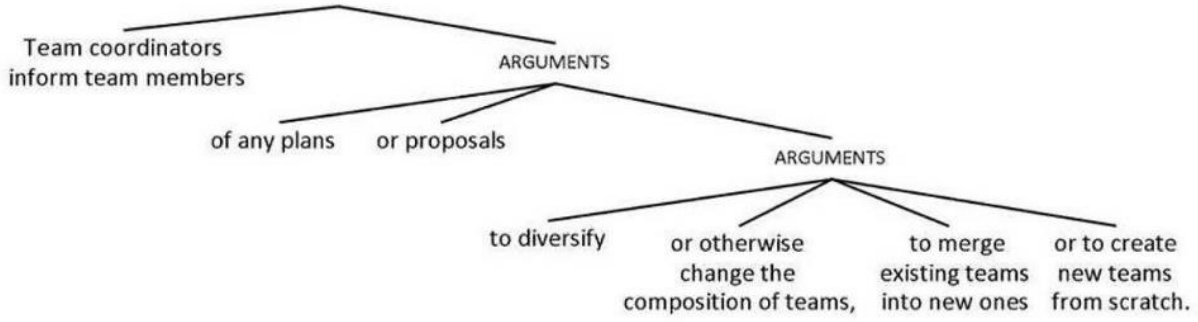

Figure 65
Semantic parse tree of complex sentence

In the parse tree in figure 65, the propositions in the bottom row function as arguments of the underlying verbs “plan” and “propose,” from which the nominal predicates in the second row are derived. “Plans” and “proposals” are treated here as process nominals – that is, predicates with argument structure. In contrast, “composition” in the second proposition in the bottom row is treated as a result nominal with no argument structure – just as if it said “names” or “numbers” instead of “composition.” The distinction between result nominals and process nominals is sometimes fuzzy. This distinction is discussed below, in section 1.7 of this annex.

The sort of semantic parse tree used here is similar in appearance to syntactic parse trees in the tradition of generative grammar. The main difference is that each leaf on the trees used here shows the syntactic expression of a semantic proposition. This method of segmentation and display allows for one-to-one comparison of corresponding propositions in translation and interpretation, however they’re expressed in different languages. It also helps illustrate problems in transferring complex sentence structure from one language to another.

1.3 Closed and open propositions

A clause-like nominal structure is unspecified for tense, as in (44).

(44) children’s respect for teachers	NOMINAL STRUCTURE tense unspecified

A corresponding finite clause would require the tense to be specified or implied, as in (45).

(45) Children respect teachers.	FINITE CLAUSE tense specified

A clause-like nominal structure can also have undefined arguments, while a corresponding finite clause would require all arguments to be defined. For example, the predicate [RESPECT] is a binary function. It takes two arguments as inputs, and returns a truth value – true or false – as an output, as shown in (46).

(46) Children respect teachers.	RESPECT (children, teachers) = {T, F}

Here both arguments are defined by known values – "children" and "teachers." A proposition whose arguments are defined by known values and where tense is specified or implied is a **complete** or **closed proposition**. Such a proposition returns an actual truth value – true or false. (Again, the perspective and criteria for determining that value are irrelevant here.)

The predicate [RESPECT], expressed as a noun, shares the argument structure of the verb it's derived from. But a proposition expressed with the nominal form of the predicate may have an undefined argument, as in (47).

(47) respect for teachers — NOMINAL STRUCTURE: one argument undefined

A corresponding finite clause would require all arguments to be defined, as shown by the incompleteness of (48).

(48) *respect teachers — FINITE CLAUSE: needs another argument

A predicate with one or more undefined arguments or unspecified for tense is an **incomplete proposition**. Such a proposition returns a potential truth value, rather than an actual one. This type of proposition can be called a "propositional function," in the tradition of Russell (1903) and Lewis (1918). Cresswell (1973) and others use the term "**open proposition**," by analogy to what Quine (1986) and others call an "open sentence." This study uses the term "proposition" for any predicate with at least one argument or adjunct, including open propositions as defined here.

A proposition with the same nominal predicate as in (47), but with a different undefined argument than in (48), is shown in (49).

(49) children's respect — NOMINAL STRUCTURE: one argument undefined

A corresponding clause would again need all arguments to be specified, as shown by the incompleteness of (50):

(50) *children respect — FINITE CLAUSE: needs another argument

There's no black-and-white rule for what constitutes the syntactic expression of a proposition. Rather, there's a continuum of predicate-argument structures in various languages expressing propositions, as defined broadly here. Ranging from fully fledged propositions to trivial ones, these include:

- finite clauses, as in figure 66

Wendy gave her secretary a present.

Figure 66
Finite clause

- tenseless clauses with full argument structure, as in figure 67

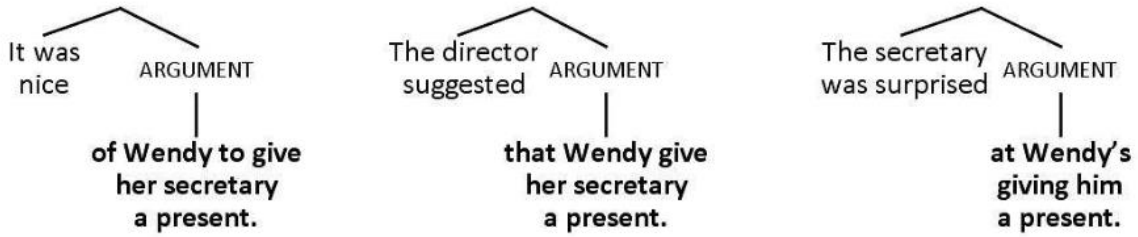

Figure 67
Tenseless clauses with full argument structure

- gerund and infinitive clauses with one or more undefined arguments, as in figure 68

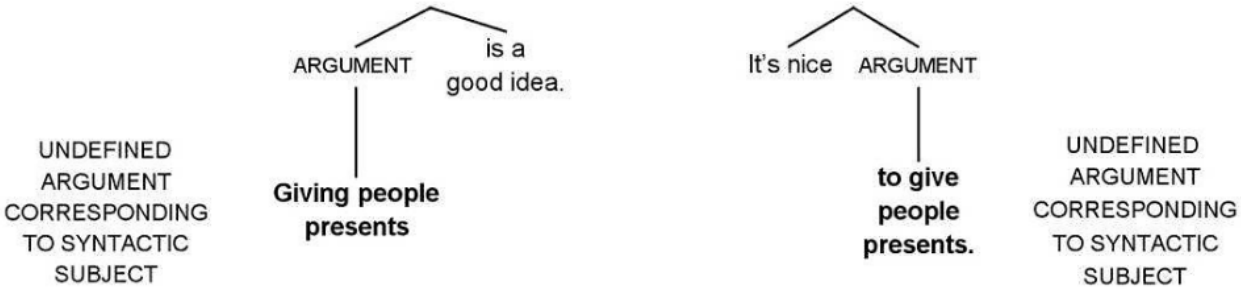

Figure 68
Gerund and infinitive clauses with undefined arguments

• clause-like nominal structures with one or more undefined or implied arguments, as in figure 69

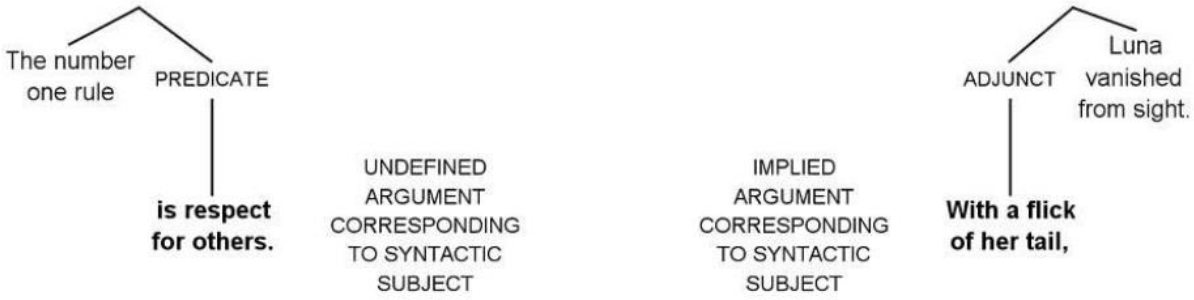

Figure 69
Clause-like nominal structures with undefined or implied arguments

• gerunds and infinitives with adjuncts but no arguments, as in figure 70

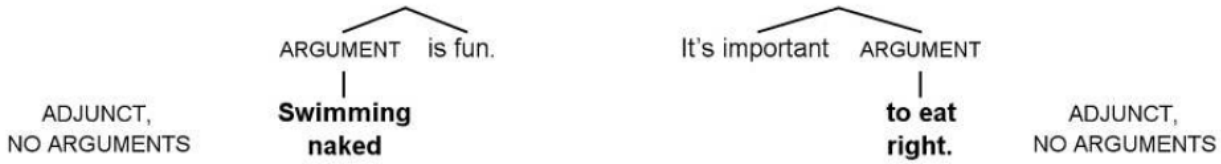

Figure 70
Gerunds and infinitives with adjuncts but no arguments

• Finally, a nominal may have empty argument structure, as in figure 71. Such nominals aren't treated as separate propositions in this study.

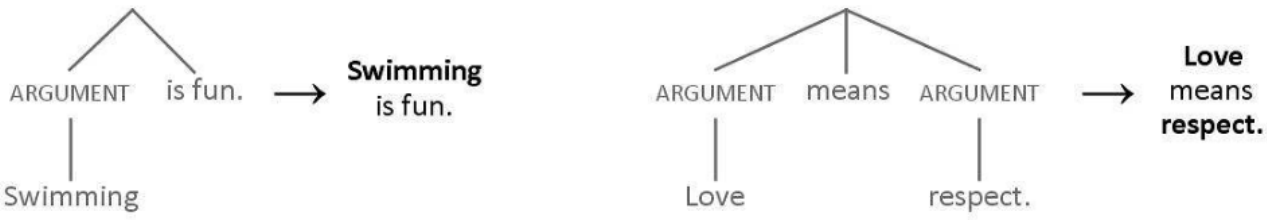

Figure 71
Nominals with empty argument structure – not treated here as separate propositions

1.4 Adjuncts to predicates

Besides a predicate and its arguments, a proposition can also include other elements which modify the predicate. Those elements aren't required inputs for the logical function to return an actual or potential truth value, though they can change the truth value returned. Such additional elements are **adjuncts**.

An adjunct can modify a clausal predicate, as in figure 72.

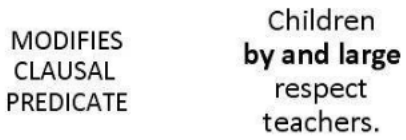

Figure 72
Adjunct to clausal predicate

Or an adjunct can modify a nominal predicate, as in figure 73.

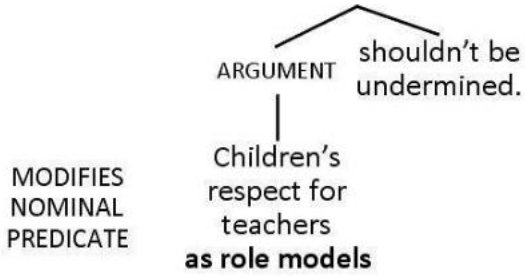

Figure 73
Adjunct to nominal predicate

Or an adjunct can have scope over an entire clause, as in figure 74.

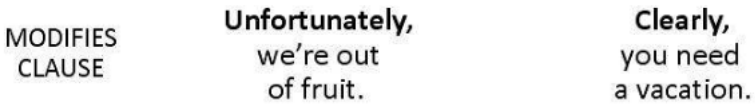

Figure 74
Adjuncts to entire clauses

An adjunct to a clause, like the ones highlighted in figure 74, is also called an “extra-clausal constituent,” and is considered part of the “macro-syntax” (as opposed to the “micro-syntax”) of a sentence (Kaltenböck et al. 2016). Blanche-Benveniste (1983) calls such clauses “sentence complements” (as opposed to “verb complements”). Quirk et al. (1989) call them “disjuncts” and “conjuncts” (as opposed to “adjuncts” – a term they reserve for adjuncts to predicates). Adjuncts to clauses are treated as operators separate from discourse units in some segmentation methods, such as the linguistic discourse model (Polanyi et al. 2004).

But it can be hard to tell if an adjunct has scope over an entire clause or if it modifies the predicate of that clause. Sometimes the place of an adjunct can help determine its scope. For example, in some languages, an adjunct at the beginning of a clause can be seen as having scope over the entire clause, whereas an adjunct placed later can be seen as modifying the predicate of that clause, as in figure 75.

Figure 75
Adjuncts to clauses vs adjuncts to predicates

But this criterion may not always work. Also, the distinction is too language-specific to be used reliably in comparing different language versions of a sentence. In any case, an adjunct to a clause generally remains attached to the same clause in the translated or interpreted version of a sentence as in the original version. For all those reasons, the parsing method used here doesn’t distinguish between adjuncts to clauses and adjuncts to predicates.

1.5 Subordinate propositions

Modifying propositions

A subordinate proposition can modify a noun in a parent proposition. In figure 76, the modifier is a shortened relative clause with a prepositional predicate. (Recall that a preposition can function as a predicate, with or without a complement, as discussed in section 1.1 above.) The modified noun on the upper leaf is underlined, to show what the subordinate proposition modifies.

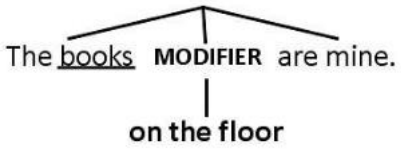

Figure 76
Subordinate proposition modifying a noun in its parent

Sometimes a subordinate proposition syntactically splits the predicate of a higher-level proposition from one or more of its arguments. In figure 76, the predicate and argument of the main proposition, "The books are mine," are syntactically split by the modifying proposition, "on the floor."

Modifying propositions include restrictive relative clauses, as in figure 77.

Figure 77
Restrictive relative clauses

Sometimes an adjective can be rephrased as a relative clause, as in figure 78.

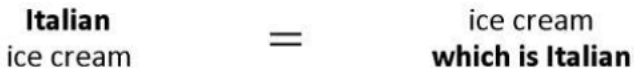

Figure 78
Adjective rephrased as relative clause

But this study groups a modifier without arguments or adjuncts in the same proposition as the noun it modifies, as illustrated in figure 79.

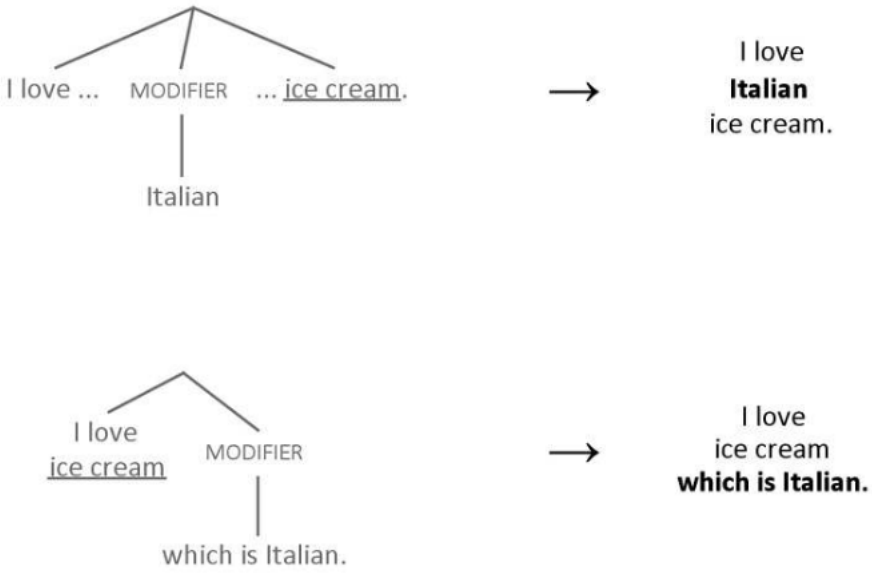

Figure 79
Modifiers with no arguments or adjuncts
grouped in same propositions as modified nouns

Argument propositions

A semantic argument is the semantic equivalent of a syntactic argument (subject or complement) of a predicate. A subordinate proposition can be a semantic argument of the predicate in a parent proposition, as in figure 80.

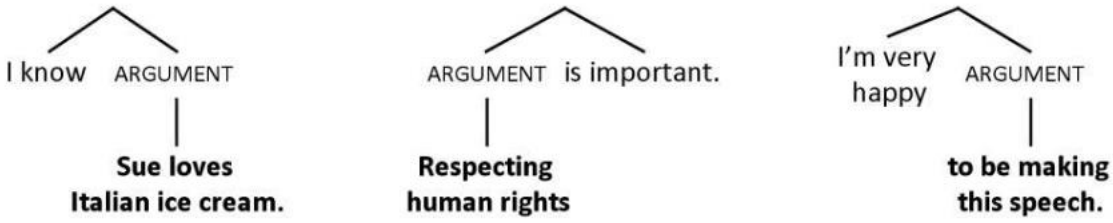

Figure 80
Subordinate propositions as arguments of predicates in parents

Adjunct propositions

A subordinate proposition can be a semantic adjunct to the predicate in a parent proposition, as in figure 81.

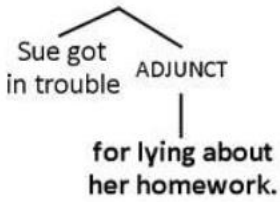

Figure 81
Subordinate proposition as adjunct to predicate in parent

Some non-clausal adjuncts don't appear to have predicate-argument structure, as they don't have an obvious predicate. But they still act like propositions in their own right. Schauer (2000) proposes treating a phrase as a separate discourse unit if it: (a) provides information other than semantically typical information like means, place or time, and (b) can be rephrased as a clause. Larson (1984: 219-223) considers indication of agent, causer, affected entity, beneficiary, accompaniment, result, instrument, location, goal, time, manner or measure to be semantically typical information in a proposition. This study adopts a combination of those two sets of criteria for treating a non-clausal adjunct as a proposition. Specifically, a non-clausal adjunct is treated here as a proposition if it: (a) provides information other than that in Larson's list of semantically typical information in a proposition or describes an event, and (b) can be rephrased as a clause. An example can be seen in figure 82.

Figure 82
Non-clausal adjunct treated as a proposition

A non-clausal adjunct like the one highlighted on the left in figure 82: (a) provides semantically non-typical information, as defined (information other than agent, causer, affected entity, beneficiary, accompaniment, result, instrument, location, goal, time, manner or measure), and (b) can be rephrased as a clause. So such an adjunct is treated here as a separate

proposition. In contrast, the adjunct in figure 83 provides semantically typical information (the instrument) and is not treated here as a separate proposition.

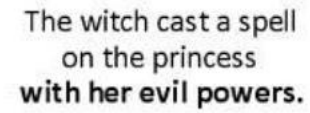

Figure 83
Adjunct not treated as a proposition

Of course, none of these tests are absolute, as the propositional status of phrases is on a continuum. That's why it's important for this study that any decision as to whether to assign propositional status to a given phrase in the original version of a sentence, based on criteria such as described above, should be applied consistently to all other language versions of that sentence when the meaning is judged to be the same.

Translating or interpreting a proposition from one language to another can involve changing its syntactic form from a finite clause to a non-finite clause, or from a clause to a clause-like nominal structure or a non-clausal adjunct proposition. For optimal cross-linguistic comparison, all phrases with propositional function as described above are treated in this study as separate propositions.

1.6 Reported speech or thought

A construction expressing reported speech or thought is functionally neither independent nor subordinate in the same way as other constructions. Consider the highlighted propositions in (51)-(54).

(51) She was like, **"Cool."**

(52) The teacher said: "**Please arrive early tomorrow,** since the buses will be leaving at 8:15."

(53) The sign says **that items may be returned for a refund within 30 days** if accompanied by the receipt.

(54) Commercials are so annoying. **"Buy me now!"** Like I really need a programmable vacuum cleaner.

Syntactically, such constructions may look like they're in a relation of subordination. But they're different from typical subordinate constructions in several ways. These include: (a) a shift of perspective away from the speaker or writer, (b) salience of information, (c) a wide range of complexity, (d) distinctive prosody and (e) the fact that they can appear without a syntactic parent, as in (54). Because of these cross-linguistic features, Spronck and Nikitina (2019) argue that reported speech or thought "involves a number of specific/characteristic phenomena that cannot be derived from the involvement of other syntactic structures in reported speech, such as subordination." In addition to modifying, argument and adjunct propositions, propositions like those highlighted in (51)-(54) are treated here as having a fourth type of semantic relation to an explicit or implied parent: reported speech or thought.

On the other hand, a statement or view which the speaker or writer identifies with isn't treated here as reported speech or thought. Consider the highlighted propositions in (55)-(57).

(55) We all know **we're living in challenging times**.

(56) The report proves **that temperatures are rising.**

(57) Passengers are informed **that the departure gate for flight AB 105 to Seoul has changed**.

Such propositions don't shift perspective away from the speaker or writer. So they're treated here as functionally independent propositions. Their effect is similar to the highlighted propositions in (58)-(60).

(58) We're living in challenging times, and we know it.

(59) **Temperatures are rising**, and the report proves it.

(60) **The departure gate for flight AB 105 to Seoul has changed**, and passengers are hereby informed of that fact.

1.7 Nominals

An area where parsing decisions are sometimes borderline is how to treat a construction headed by a nominal – an abstract noun derived from an underlying verb or adjective.

The question is whether to treat such an abstract noun as a "process nominal" or a "result nominal."

A nominal is characterized here as a "process nominal" if it describes a process, event or situation. Like its underlying verb or adjective, a process nominal can have argument structure. Examples can be seen in (61)-(66).

(61) my love for you

(62) my interest in linguistics

(63) their need for help

(64) reporting on trends

(65) the importance of change

(66) food security

A process nominal can often be rephrased as a gerund, as in (67) and (68).

(67) their efforts to help → their/them trying to help

(68) greenhouse gas emissions → the emitting of greenhouse gases

Or it can be rephrased as a clause with a complementizer (like "that" or "for … to"), as in (69) and (70).

(69) the importance of change → the fact that change is important

(70) food security → for food supplies to be secure

This study treats a process nominal as the predicate of a separate proposition if it has any arguments or adjuncts, as in (71) and (72).

(71) climate change (the changing of the climate)

(72) sustainable development (developing in a sustainable way)

Likewise, a participial modifier with any arguments or adjuncts is treated here as a separate proposition, similar to a relative clause, as in (73) and (74).

(73) a [country-driven] approach (an approach [driven by the features and needs of a country])

(74) the [rapidly changing] climate (the climate, [which is changing rapidly])

Of course, phrases like "food security," "climate change" and "sustainable development" are more or less established terms. As such, they may be less internally processed than other constructions and may have established equivalents in other languages. But they still have argument structure, consisting of an underlying predicate (SECURE, CHANGE or DEVELOP) with at least one argument or adjunct. Plus, there's no objective way to determine to what extent a nominal may be internally processed. "Climate change" is clearly an established term, so it probably wouldn't be processed much internally. But what about a less established phrase, like "weather change" or "rapid climate change"? To what extent would such phrases be internally processed? Probably more than "climate change," but less than "predicting the weather." And of course it would depend on the person doing the processing and how often they'd encountered the phrase. There's no way to measure it objectively. To avoid subjective guessing on such questions, this study treats all nominal constructions consistently in the same way. And of course, the same criterion is applied to all such constructions in each language version of a sentence. So a different decision as to whether or not to treat a given nominal as a predicate would have no effect on the comparative results between different language pairs as measured in this study.

In contrast, a nominal or abstract noun is characterized here as a "result nominal" if it describes the result of an action or something created by an action. Like any noun, a result nominal can take modifiers. But a result nominal tends to be closed to arguments (even if its modifiers could be arguments of an underlying verb or adjective). Examples can be seen in (75)-(78).

(75) my linguistic interest(s)

(76) a report on trends

(77) the effects of change

(78) support activities

Such a nominal generally can't be rephrased as a gerund. It's not an "-ing," but a "thing," as in (79) and (80).

(79) an announcement (a thing that's announced)

(80) a report (a thing that's written, things that are reported)

This study treats a result nominal as a simple noun, not as the predicate of a separate proposition, even if it's modified. Examples can be seen in (81) and (82).

(81) common responsibilities

(82) respective capacities

Several researchers (Melloni 2011; Alexiadou 2010; Rozwadowska 2017; Grimshaw 1992) discuss the distinction between process and result nominals. The distinction isn't always clear-cut, as in (83) and (84).

(83) greenhouse gas emissions = the process of emitting greenhouse gases?
or the gases that are emitted?

(84) their progress towards the goal = the process of their progressing toward the goal?
or the results they've achieved?

Sometimes the form of a nominal in one language suggests a reading as a process nominal, while its form in another language suggests a reading as a result nominal. For example, the standard Turkish term for "climate change" (*iklim değişikliği*) suggests that it refers to a change that's been produced in the climate, whereas the term used in most other languages is likely to be seen as referring to the process of the climate changing.

To minimize the effect of such differences on our data, care has been taken here to consistently parse each nominal in a translated or interpreted version of a sentence in the same way as in the original English version, whenever the information content is judged to be the same.

We've seen how this study uses the proposition as a unit of analysis. Now let's take a closer look at the parsing method used here to segment sentences into propositions and to indicate the hierarchical relations between them.

2. A functional approach

The method of parsing complex sentences used in this study is described as “semantic” and “functional.” That’s because it parses complex sentences into semantic propositions, as was explained in section 1 of this annex, and is based on underlying semantic function rather than surface syntactic form.

We’ve seen how a semantic parse tree, similar to a syntactic parse tree in the tradition of generative grammar, can be used to illustrate the propositional structure of a complex sentence. Here’s more detail on how the parse trees in this study are laid out: Though segmented semantically, propositions are displayed as leaves on the trees in their surface syntactic forms. Some features recorded for our statistical analysis involve the linear order of propositions in a complex sentence. So the parse trees are spread out horizontally to show both the linear order of propositions and the hierarchical relations between them.

Syntactically contiguous parts of a proposition are shown together on the same leaf of a tree. A functionally independent proposition – one which makes an assertion – is shown at the top level of the tree. A functionally subordinate proposition, including one which is part of a parent proposition, is shown at a lower level. Parts of a proposition which are syntactically split by another proposition are shown on separate leaves, at the same level and under the same node. An overt coordinating link between propositions is included on the same leaf as one of the linked propositions. An overt subordinating link is included on the same leaf as the subordinate proposition.

Each subordinate proposition is labeled arg (argument), mod (modifier) or adj (adjunct) in relation to its parent. A proposition containing reported speech or thought is labeled rep. The label for a subordinate or reported proposition is shown at the same level and under the same node as its parent. (For a reported proposition, the parent is the proposition where the perspective of the reported one is established.) A noun modified by a subordinate proposition is underlined. An example semantic parse tree of a complex sentence is shown in figure 84.

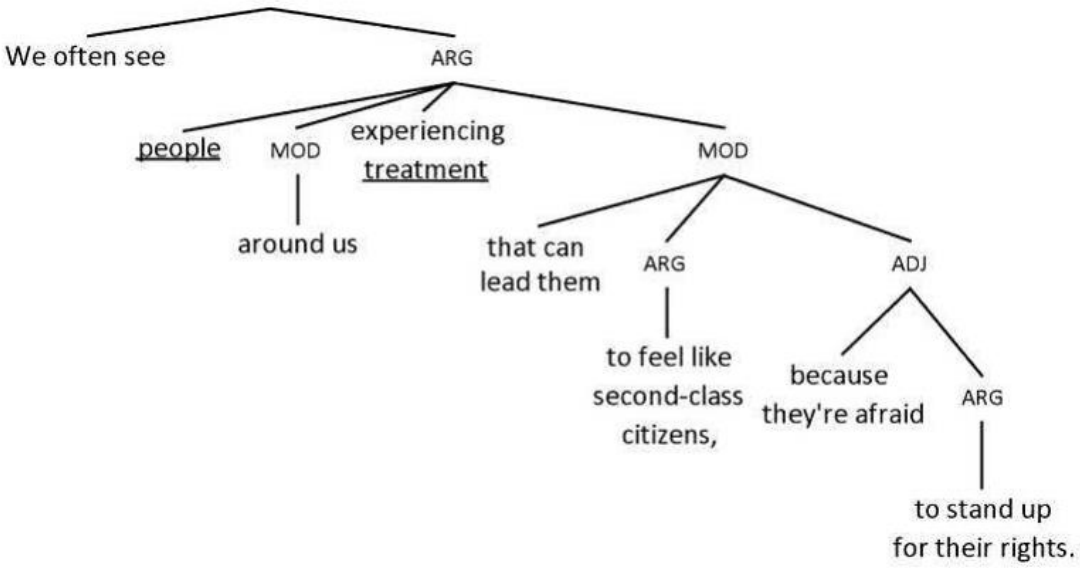

Figure 84
Semantic parse tree of a complex sentence

2.1 Not all propositions are clauses

The semantic, functional approach to complex sentence analysis used here is similar in many ways to Cristofaro's (2003) analysis of subordination, in that it focuses on underlying meaning rather than syntactic form. Instead of the traditional unit of analysis for complex sentences, the syntactic clause, Cristofaro uses the state of affairs – the description of an event or situation. That's a lot like the unit of analysis in this study – the proposition.

Cristofaro's (2003: 2) functional approach defines a pair of linked clauses as being in an asymmetric relation if one clause describes an event or situation which "lacks an autonomous profile, and is construed in the perspective of the other event (which will be called the main event)." In a sentence with two asymmetrically linked clauses, there's generally a main event or situation description, which is an assertion, and a subordinate description, which provides assumed or background information.

Cristofaro proposes three types of test – sentence negation, sentence questioning and tag questions – to determine whether a clause in a complex sentence is an assertion. Accordingly, she classifies that clause as functionally independent or subordinate. English tag questions (like "has she?" or "wouldn't they?") are handy for this purpose, as they reflect the person and tense of the verb in the clause being questioned. Cristofaro uses English tag questions – on original English sentences and on translations of sentences from other languages – to test

which of two linked clauses is functionally independent and which is subordinate. For example, consider (85).

(85) He thinks it will rain.

If we add a tag question to (85), we get (86).

(86) He thinks it will rain, doesn't he? (= Isn't it true that he thinks that?)

We don't get (87).

(87) *He thinks it will rain, won't it? (≠ Isn't it true that it will rain?)

As shown in (86), the main clause in (85), "He thinks," can be questioned with a tag question, indicating that it's an assertion. As shown in (87), the subordinate clause in (85), "it will rain," can't be questioned with a tag question, indicating that it's not an assertion. That second clause is functionally subordinate to the main clause, "He thinks."

Cristofaro gets close to a definition based on semantic propositions as used here, saying that her analysis is based on "functional relations" between states of affairs, "rather than any formal feature of specific clauses" (2003: 51). But the phrase types she applies her assertion tests to and characterizes as independent or subordinate are all finite or non-finite clauses. Her states of affairs don't seem to cover other syntactic structures that can express propositions as perceived in this study.

Such structures include clause-like nominal structures, as in (88).

(88) I must confess *my love for you*.

They include shortened relative clauses, as in (89).

(89) The papers *on the desk* are mine.

They also include other phrases with propositional function, as in (90).

(90) We'll get there in time *thanks to your help*.

Cristofaro gives a few examples of states of affairs expressed in clauses with nominalized verb forms, as in the Finnish sentence in (91).

(91) Huomaan *pojan osanneen suomea*.
literally: "I realized *the boy's knowing Finnish*."
= I realized *that the boy knew Finnish*.

But her analysis of subordination doesn't seem to include events described in functionally similar clause-like nominal structures, as in (92).

(92) I admired *the boy's knowledge of Finnish*.

Like Larson (1984), the method used here treats all phrases with predicate-argument structure as propositions, including ones like those highlighted in (88)-(90) and (92).

Using the proposition as a unit of analysis can lead to more unit divisions than using the clause, with a phrase like "thanks to your help" being treated as a functional proposition. As explained in section 1.5 of this annex, a non-clausal adjunct is treated here as a proposition if it (a) provides information other than agent, causer, affected entity, beneficiary, accompaniment, result, instrument, location, goal, time, manner or measure or describes an event, and (b) can be paraphrased as a clause. Some more examples are illustrated in figure 85.

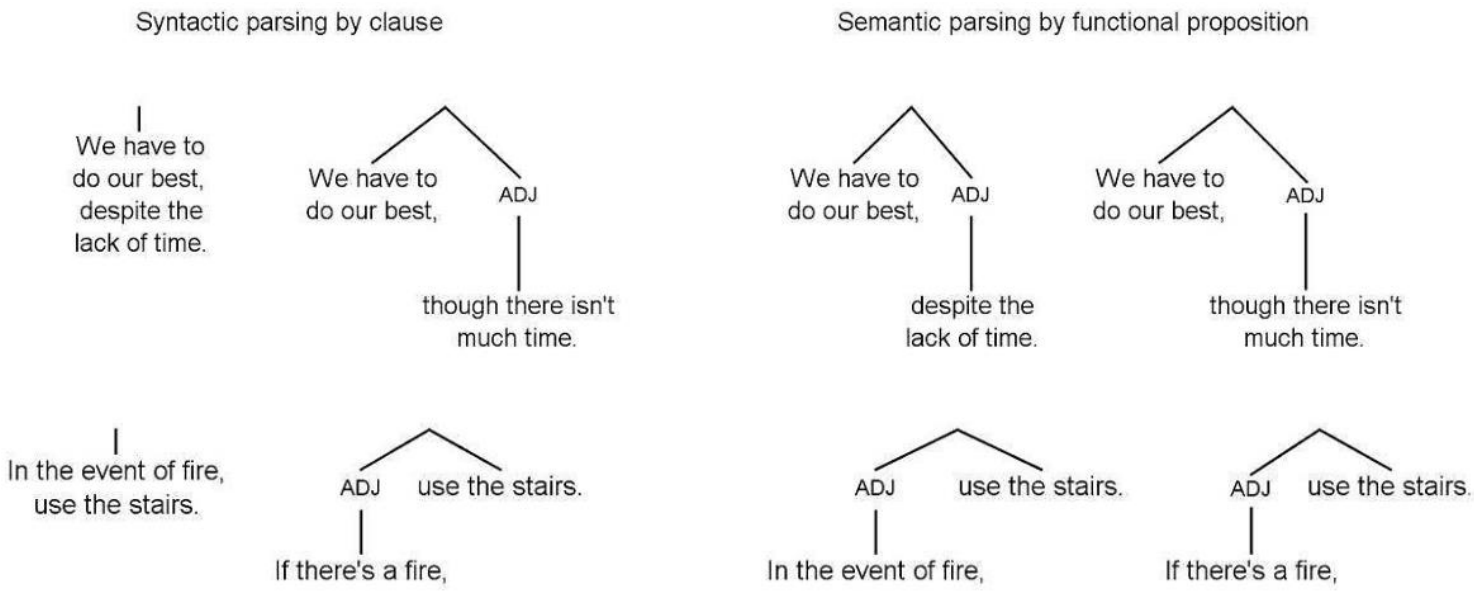

Figure 85
Syntactic vs semantic parsing of non-clausal propositions

Cosme (2008: 91) observes: "What is expressed by a phrase in one language may be expressed by a clause in another language…. Translating from English into German frequently triggers … a phenomenon whereby a clause in the source language is turned into a phrase in the target language." Doherty (1999: 115) gives examples of this phenomenon in German-to-English translation, as in (93).

(93) Auch werden *bei einem verlangsamten Zellwachstum* nur geringe Mengen an natürlichem Interferon gefunden.
(literally: "Also *with slowed-down cell growth*, only small amounts of naturally produced interferon are found.")
→ Also *when cell growth is slowing down,* only small amounts of naturally produced interferon are found.

This functional equivalence of different syntactic structures used to express propositions is key to the way proposed here for assessing complex sentence transfer between languages. Like Larson (1984), this study sees reproduction of the hierarchical relations between all propositions, regardless of syntactic form, as central to preserving meaning in a successful translation.

2.2 Not all clauses are propositions

We've seen how a semantic proposition isn't always expressed in a syntactic clause. So dividing a sentence into propositions can lead to more units than dividing it into clauses. But the reverse is true too: not every clause expresses a proposition. So dividing a sentence into propositions can also lead to fewer units than dividing it into clauses. This can happen when a clause doesn't appear to express a proposition at all. For example, consider the two clauses in (94).

(94) I think it will rain.

We could try Cristofaro's tag question test to help identify which of the two clauses in (94) is an assertion, as in (95).

(95) I think it will rain, won't it? (= Isn't it true that it will rain?)

Syntactically, the sentence in (94) is a complex sentence, with "I think" as its main clause. But the tag question test in (95) shows that the subordinate clause, "it will rain," can be taken

as an assertion. Understood that way, the main clause is “transparent” to the tag question, which “sees through” it, referring to the subordinate clause.

However, the main clause in (94), “I think,” can also be seen as an assertion. This is shown by the sentence negation test in (96).

(96) It’s not true that I think it will rain. (= I don’t think that.) (≠ It won’t rain.)

Both interpretations are possible, depending on emphasis. The statement can be seen as being mainly about the likelihood that it will rain, as in (97) and (98).

(97) I think it will rain, won’t it?

(98) I think it will *rain*, but it might *snow*.

Or the statement can be seen as being mainly about the speaker’s thought, as in (99) and (100).

(99) I *think* it will rain, but I’m not *sure*.

(100) *I* think it will rain, don’t *you*?

Main clause structures like “I think” are often spoken or intended to be read without emphasis, as in (97) and (98). Used that way, they function not as an assertion, but as a formulaic expression of the speaker or writer’s attitude to an assertion made in a subordinate clause. Brinton (2008) refers to such semantically weak main clauses as “comment clauses.”

A comment clause like “I think” can be replaced, with similar effect, by an impersonal adjunct like “probably”, as in (101).

(101) I think it will rain. ≈ It will probably rain.

Other comment clauses can be similarly replaced by impersonal elements, as in (102) and (103).

(102) They say it will rain. ≈ Supposedly, it will rain. ≈ It’s supposed to rain.

(103) It seems it will rain. ≈ Apparently, it will rain.

In contrast, a main clause that makes an assertion, like "he thinks," can't be replaced with similar effect by an impersonal adjunct, as in (104).

(104) He thinks it will rain. ≠ It will [probably / apparently / supposedly] rain.

Brinton (2008) also refers to comment clauses as "clausal pragmatic markers." Such grammaticalized markers can appear at various places in a sentence, as in (105)-(107).

(105) I think it will rain.	It will, I think, rain.	It will rain, I think.
(106) They say it will rain.	It will, they say, rain.	It will rain, they say.
(107) It seems it will rain.	It will, it seems, rain.	It will rain, it seems.

The semantic parsing method used in this study doesn't treat such pragmatic markers as separate propositions.

Comment clauses aren't the only type of syntactic clause that doesn't always seem to express a separate proposition. Consider the highlighted portions of (108)-(110), taken from President Obama's 2015 speech to the UN General Assembly, which is in the corpus for this study.

(108) *We see* an erosion of the democratic principles and human rights that are fundamental to this institution's mission.

(109) *There are* certain ideas and principles that are universal.

(110) *That's what* those who shaped the United Nations 70 years ago understood.

The highlighted portions of (108)-(110) are there for narrative cohesion and emphasis, rather than semantic content. Larson (1984, 457) notes: "The greatest amount of mismatch between languages probably comes in the area of devices which signal cohesion and prominence." Her translation method recommends first stripping away such semantically light elements from the original message, so that its syntax mirrors the semantic structure of the underlying propositions, before reformulating those propositions in the target language. Accordingly, we can restate (108)-(110) so that their syntax more closely mirrors their semantic structure, as in (111)-(113).

(111) The democratic principles and human rights that are fundamental to this institution's mission are eroding.

(112) Certain ideas and principles are universal.

(113) Those who shaped the United Nations 70 years ago understood that.

In cases like (108)-(110), where the main clause has a semantically light predicate, this study parses the short main clause along with the more informative subordinate clause as a single functional proposition, equivalent to (111)-(113).

This approach is useful for cross-linguistic comparison. One problem is that it's not always easy to decide if a short main clause which sets the framework for a more informative subordinate clause should be treated as a comment clause or as an assertion. The more semantically strong or emphatically spoken the predicate in the main clause, the more that clause can feel like an assertion in its own right. For example, consider the main clause in (114), taken from the same speech.

(114) *I recognize* that democracy is going to take different forms in different parts of the world.

The main clause in (114), "I recognize," behaves differently from the comment clauses or pragmatic markers we saw above, in several respects: First, it feels more emphatic than "I think," which can be spoken without emphasis. Second, it can't be replaced with similar effect by an impersonal adjunct, as in (115).

(115) ≠ Democracy is probably/definitely going to take different forms in different parts of the world.

Third, it's harder to move "I realize" to different parts of the sentence, as in (116) or (117).

(116) ?Democracy is, I recognize, going to take different forms in different parts of the world.

(117) ??Democracy is going to take different forms in different parts of the world, I recognize.

Fourth, (114) can be restated as two independent clauses, as in (118) and unlike (119).

(118): Democracy is going to take different forms in different parts of the world, and I recognize that.

(119): *It's going to rain, and I think that.

The above evidence supports the feeling that (114) is both an assertion of the speaker's recognition and an assertion about democracy. A case can be made for not attributing functionally independent status to the main clause in a sentence like (114): Verhagen (2001: 349) describes such a main clause as "conceptually dependent" on the subordinate clause, which "provides the most important information." Still, based on evidence like (115)-(118), this study treats both the main clause and the complement clause in a sentence like (114) as functionally independent propositions, for ease of cross-linguistic comparison. We'll come back to this in section 3 of this annex.

In contrast, the less semantically strong or emphatically spoken the predicate in the main clause, the more that clause can feel and behave like a comment clause. Consider the main clause in (120), taken from the same speech.

(120) *I believe* that capitalism has been the greatest creator of wealth and opportunity that the world has ever known.

The propositional status of the main clause is harder to pin down in (120) than in (114). Is (120) just asserting that capitalism has a good side, with the main clause, "I believe," functioning as a pragmatic marker, like unstressed "I think"? Or is the sentence also making an assertion about the speaker's belief that that's true? "I believe" in (120) seems to occupy a middle ground in terms of semantic strength – stronger than unstressed "I think," but weaker than "I recognize" in (114). The case for the comment clause interpretation of (120) is strengthened by the fact that "I believe" can be moved to a different place in the sentence with similar effect, as in (121).

(121) Capitalism has, *I believe,* been the greatest creator of wealth and opportunity that the world has ever known.

The comment clause interpretation of (120) is also supported by the fact that the sentence doesn't work very well if we try to restate it as two independent clauses, as in (122).

(122) ?Capitalism has been the greatest creator of wealth and opportunity that the world has ever known, and I believe that.

Another test, since the sentences in question are taken from a speech, is how emphatically the main clause is spoken. All the above tests give results which are on a continuum and open to subjective judgment. So the distinction between assertive force and pragmatic function for the main clause in a sentence like (120) is less than clear-cut.

In such borderline cases, this study makes a parsing decision based on tests such as those illustrated above. If the evidence favors seeing the main clause as an assertion, both main and subordinate clause are treated here as functionally independent propositions. If the evidence favors seeing the main clause as a pragmatic marker, the two clauses are parsed together as a single proposition. Whatever parsing decision is made for the original English version of a sentence, the same decision is applied as consistently as possible to all translated or interpreted versions of that sentence. This ensures that different language versions of a clause with borderline propositional status aren't recorded as involving changes in semantic relations because of different parsing decisions applied to the original and various translated or interpreted versions of that clause.

2.3 Subordination and coordination

We've seen how syntactic relations of subordination between clauses don't always correspond to semantic relations of subordination between propositions. In fact, the syntactic and semantic hierarchy is sometimes reversed, with a main clause providing background information and a subordinate clause making a functionally independent assertion. An example can be seen in (123).

(123) No sooner had the little pig arrived at the house of bricks than the wolf appeared.

Syntactically, the main clause in (123) is "No sooner had the little pig arrived at the house of bricks." But the word order and tense of that clause indicate that it's functionally subordinate to the other clause. It's clear that the pig arriving at the house of bricks is background, assumed information and that the salient, asserted information is that the wolf appeared. The tag question test confirms this, as in (124) and (125).

(124) No sooner had the little pig arrived at the house of bricks than the wolf appeared, didn't he?

(125) *No sooner had the little pig arrived at the house of bricks than the wolf appeared, hadn't he?

Functionally, “no sooner” acts here as a subordinating link, like “as soon as” in (126).

≈ (126) As soon as the little pig arrived at the house of bricks, the wolf appeared.

Following our semantic, functional approach, the independent assertion in (123) should be treated as the main proposition, even though it’s expressed syntactically in a subordinate clause. The contrast is illustrated in figure 86.

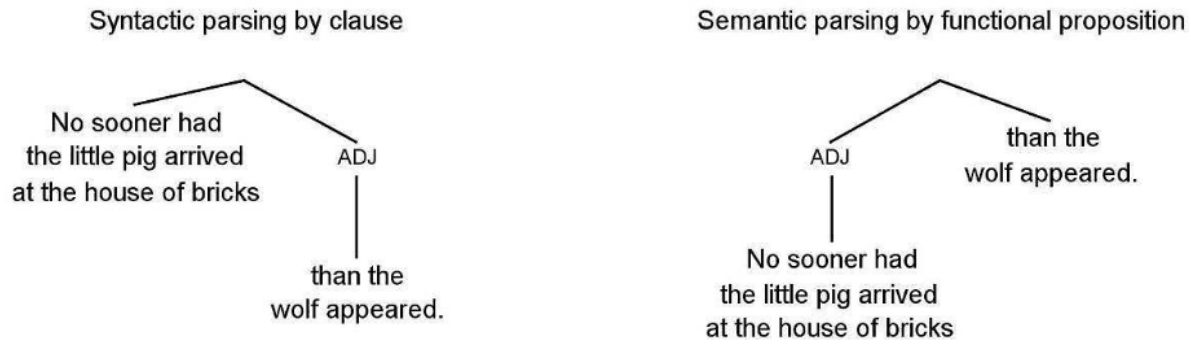

Figure 86
Contrast between syntactic and semantic analysis of subordination

The sentence illustrated in figure 86 is a clear case of mismatch between syntactic and functional subordination, with the proposition expressed in the syntactic main clause functionally subordinate to the one expressed in the subordinate clause. But it’s not always easy to tell which of two asymmetrically linked propositions is functionally subordinate to the other, or even if their link is asymmetrical at all. Consider a sentence like (127).

(127) My wife’s learning Japanese, because she loves Japan.

Is the link here one of functional subordination or coordination? And if it’s subordination, which proposition is subordinate to the other? Traditional grammar generally classifies “because” as a subordinating conjunction. But functionally, its subordinating function isn’t always so clear. The sentence in (127) can be seen as being as much about the speaker’s wife’s love for Japan as about the fact that she’s learning Japanese. The contrast is illustrated in figure 87.

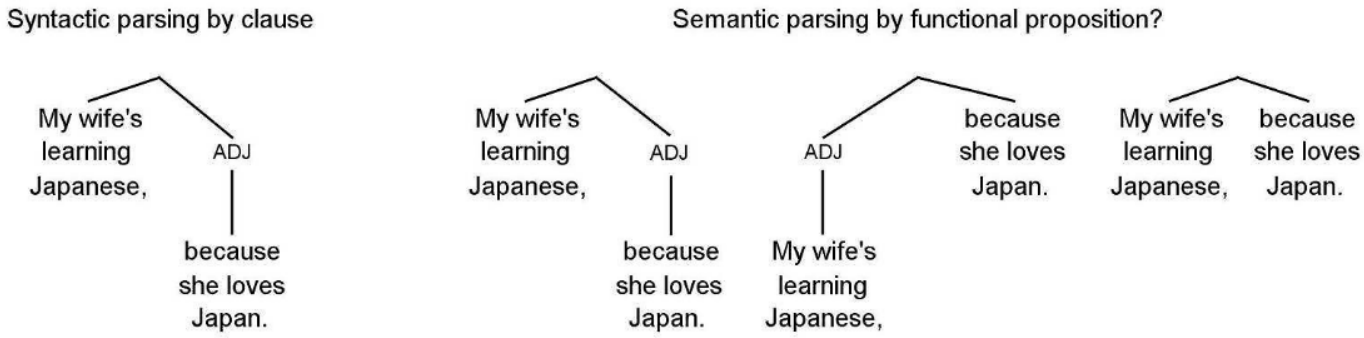

Figure 87
Borderline relations of semantic subordination

Or we can reverse the propositions from (127), as in (128).

(128) My wife loves Japan, so she's learning Japanese.

Traditional grammar generally classifies "so" as a coordinating conjunction. But, like before, this sentence may be seen as being as much about the fact that the speaker's wife is learning Japanese as about her love for Japan. Again, the choice is unclear, as illustrated in figure 88.

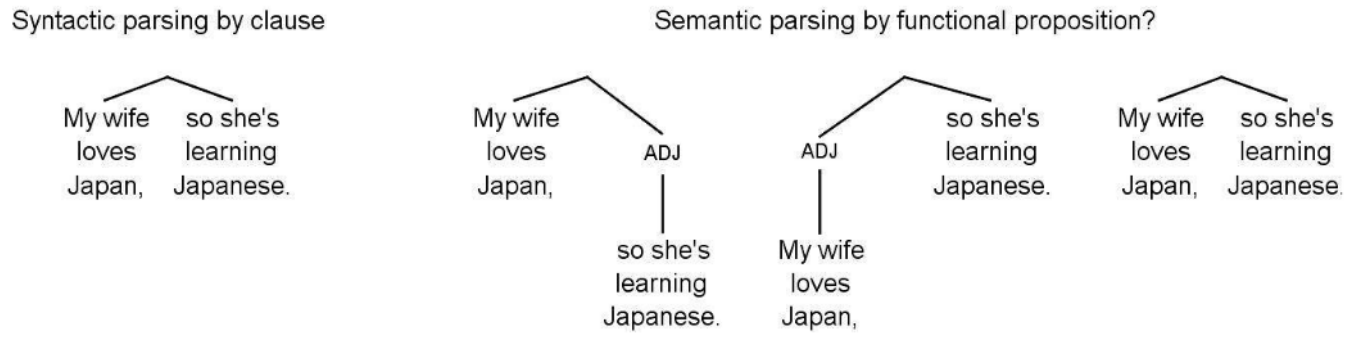

Figure 88
Borderline relations of semantic subordination

This functional ambiguity is characteristic of various clause links in different languages. Cristofaro (2003: 2) cites the example in (129) from Mandarin (gloss added), with a suggested English translation.

(129) 她 喝了 酒, 就 睡著了.
Tā hēle jiǔ, jiù shuìzháole.
She drink-perf wine, then go-to-sleep-perf.
After she drank the wine, she went to sleep.

From the suggested translation, Cristofaro concludes that the functionally subordinate event in the original Mandarin sentence is the woman drinking the wine, and that the main event is her going to sleep. But the sentence could just as well be translated as in (130), (131) or (132).

(130) She drank the wine before going to sleep.

(131) She drank the wine, then went to sleep.

(132) She drank the wine and went to sleep.

In the Mandarin sentence in (129), the only overt link between the two event descriptions is the word "就 (jiù)" (≈ "then") at the beginning of the second clause. Cristofaro's suggested translation sees this link as subordinating the previous clause. But depending on the function attributed to the link, the main event described in the sentence can be seen as the person drinking the wine, her going to sleep, or both, as reflected in the variety of possible translations. Paul (2016: 185) sees clause links in Mandarin as "a challenge for the traditional analysis of complex sentences into a 'subordinate' and a 'main' clause."

Culicover and Jackendoff (1997) and Yuasa and Sadock (2002) discuss clauses which appear as syntactically coordinate but are functionally subordinate – a phenomenon they refer to as "pseudo-subordination." Ross (2016) provides a cross-linguistic analysis of this phenomenon, which he calls "pseudocoordination." For example, the prototypical English coordinator "and" can function as a subordinator of the following clause, as in (133).

(133) Try *and* catch me! (≈ Try *to* catch me!)

It can also function as a subordinator of the previous clause, as in (134).

(134) One wrong step *and* we're done for! (≈ *If* we take one wrong step, we're done for!)

The role of "and" as functional coordinator or subordinator is harder to pigeonhole in a sentence like (135).

(135) You should lie down *and* have a nap. (≈ You should lie down *and* you should have a nap.)
(≈ You should lie down *to* have a nap.)

The same is true of the null link in juxtaposed clauses like (136).

(136) Go get your toothbrush. (≈ Go *and* get your toothbrush.)
(≈ Go *to* get your toothbrush.)

Such covert, functionally ambiguous clause links are typical of a Sinitic language like Mandarin, as shown in (137) and (138).

(137) 我 來 幫助 您.
Wǒ lái bāngzhu nín.
I come help you.
I'll come **and** help you. / I'll come **to** help you.

(138) 我 有 個 哥哥, 住 在 紐約.
Wǒ yǒu ge gēge, zhù zài Niǔyuē.
I have a brother, live in New York.
I have a brother **who** lives in New York. / I have a brother **and** he lives in New York.

Transferring such sentences from a language with covert links, like Mandarin, into a language requiring overt links can force a choice as to which functional box to put those links into. Cristofaro (2003: 41) relies on translation choices to classify clause links as coordinating or subordinating, saying that we can "assume that the translation used preserves the conceptual organization" of the original. But a Mandarin speaker listening to or reading a sentence like (137) or (138) may feel no obligation to make such a binary choice. The links between the clauses can be seen as both coordinating and subordinating, or somewhere in the gray area between the two.

Again, to minimize the effect of such borderline decisions on our data, care has been taken in this study to consistently parse all language versions of a sentence the same way, whenever the information content is judged to be the same. This rule has been applied even in cases where a proposition may feel more functionally subordinate in one language version than in another, because of features like punctuation, nesting or grammatical deranking. This last issue will be discussed in section 3 of this annex.

Now we're familiar with the semantic parsing method used in this study and how it differs from syntactic parsing, in being guided by whether clauses and other phrases are functionally independent assertions. And we've seen some of the problems and uncertainties that can arise in applying that method. Finally, let's see how the focus of attention on assertions can shift as a complex sentence moves along.

3. Shifting focus

3.1 A dynamic view of complex sentences

The relations between clauses in a complex sentence are often modeled as fixed relations between different hierarchical levels. But this study sees those relations as shifting. As Langacker (2014: 69) puts it, "grammar proceeds through time…. The static formulas and diagrams employed in describing it should not obscure the fact that grammar is something that *happens*" (emphasis in original). His dynamic view of cognitive grammar posits a "primary window" of attention, which opens successively onto each clause in a chain. At any stage in the chain, that window only needs to be big enough to contain one clause, plus a "mental space" for the next clause, establishing a relation of equal or unequal salience between them.

This study adopts the idea of successive windows of attention and adapts it to the analysis of relations between semantic propositions, whether expressed in clauses or any other syntactic form. Rather than trying to shoehorn each proposition into a fixed status of providing background or salient information, we can see the focus of attention as shifting as a sentence moves forward. This is a key feature of the semantic parsing method used here. An example can be seen in (139).

(139) This is a great book, which all new parents should read.

The second clause in (139) is a non-restrictive relative clause. Ramm (2008) characterizes such clauses as functionally coordinate rather than subordinate. As (139) moves forward, the focus of attention can be seen as shifting, as illustrated in figure 89.

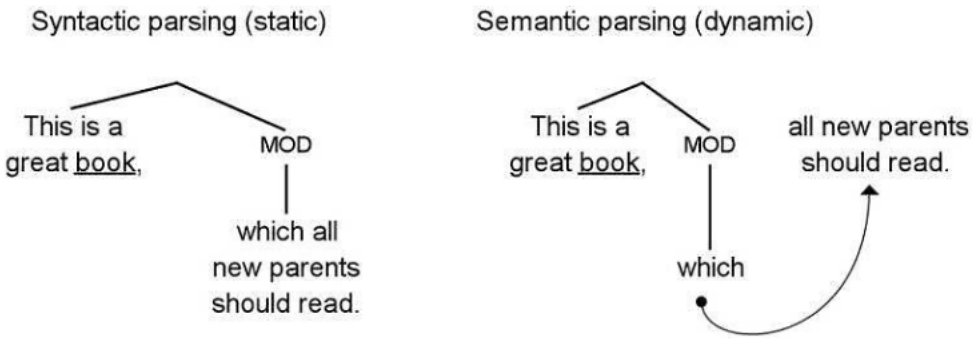

Figure 89
Syntactic vs semantic parsing of subordinate clause which makes an assertion

On the parse tree on the right in figure 89, the link “which” is shown on a separate leaf at a lower level, because it introduces a subordinate clause. The rest of the clause it belongs to is shown as moving up to the top level of the tree, to indicate that it expresses a functionally independent proposition.

According to Langacker (2014: 14), such clauses show “prosodic seriality,” suggesting a “flat structure with no internal grouping.” He sees a sequence of such clauses in a sentence as requiring “only local connections, with no more than two clausal processes appearing in any single window” (pp. 26-27). The window for each proposition holds all the elements needed to process that proposition. In the sentence in (139), a first processing window opens onto the assertion that the book being referred to is great. This is shown in (140) and illustrated in figure 90.

(140) This is a great book, (isn’t it?) …

Figure 90
Window opens

The next processing window opens onto the assertion that all new parents should read the book, as shown in (141) and illustrated in figure 91.

(141) This is a great book, which all new parents should read, (shouldn’t they?)

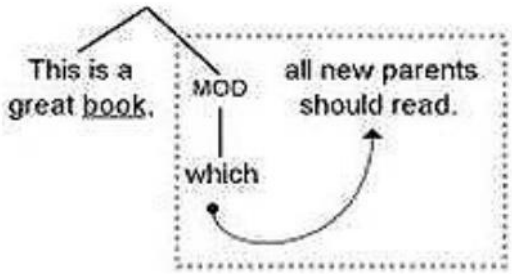

Figure 91
First window closes, next one opens

In a sentence like (139), as illustrated in figures 90 and 91, a processing window never needs to be bigger than a single proposition plus a link. That's because the first proposition with its predicate and arguments can already be processed and closed before the link. After that, the next proposition with its predicate and arguments can also be processed and closed. Each proposition in the sentence can be seen in turn as an assertion, and thus as functionally independent, even though the second proposition is expressed in a subordinate clause.

Non-restrictive relative clauses aren't the only type of subordinate clause which can make an assertion. In European languages, various links which are syntactically subordinating but introduce propositions making assertions are common in everyday speaking and writing. They can be used to create a chain of assertions in a sentence, with a processing window opening onto each proposition in turn, then closing, as a new window opens onto the next one.

To illustrate this, consider a hypothetical sentence like (142), which unfolds progressively as the speaker thinks of what to say next. (The same sentence was given as an example in the section on logical order in the epilogue.)

> (142) My wife's learning Japanese, because she loves Japan, which has always been her favorite place to visit, although she's always felt she could get more out of it if she spoke the language, which, as she's finding out, is a never-ending journey of discovery into a culture that holds endless fascination and startling beauty for the learner.

Is (142) mainly about the speaker's wife learning Japanese? Or about her loving Japan? Or Japan being her favorite place to visit? Or her wanting to speak the language? Or her finding out what that's like? Or the language being a constant discovery? Or the culture being fascinating and beautiful? The tag question test confirms that the sentence can be about each of these things in turn, as shown in (143).

> (143) My wife's learning Japanese (isn't she?), because she loves Japan (doesn't she?), which has always been her favorite place to visit (hasn't it?), although she's always felt she could get more out of it if she spoke the language (hasn't she?), which, as she's finding out (isn't she?), is a never-ending journey of discovery (isn't it?) into a culture that holds endless fascination and startling beauty for the learner (doesn't it?).

A syntactic analysis of the relations between the various parts of (142) would suggest that the sentence is mainly about the first assertion, that the speaker's wife is learning Japanese, as illustrated in figure 92.

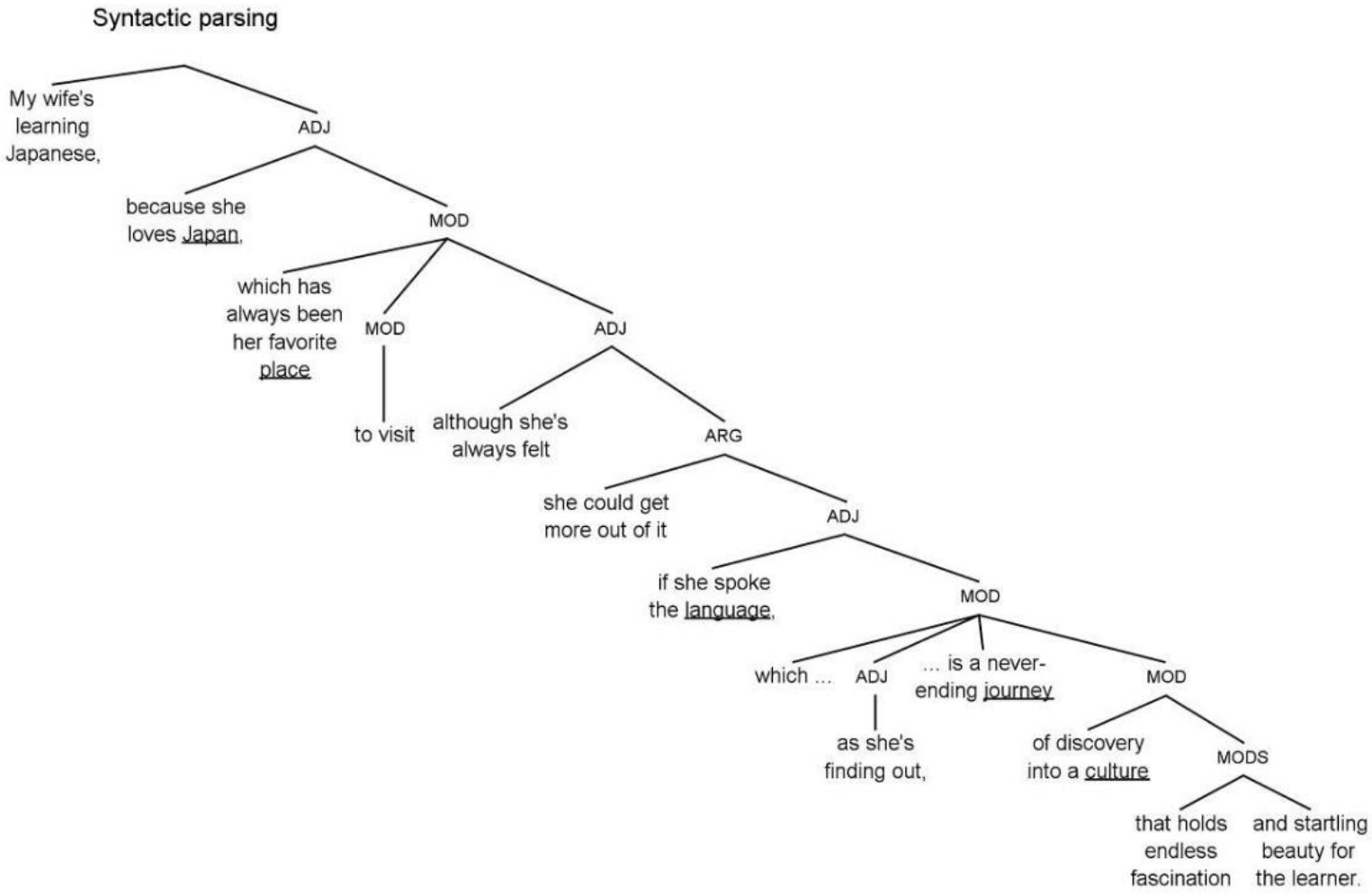

Figure 92
Syntactic parsing of complex sentence, suggesting prominence of main clause

But this study sees the semantic relations between propositions as more accurately reflected in a dynamic view, as illustrated in figure 93. Again, each link to a subordinate clause is shown on a separate leaf at a lower level of the tree. Most clauses are then shown as moving up to the top level of the tree, to indicate that they're assertions and can be seen as functionally independent propositions.

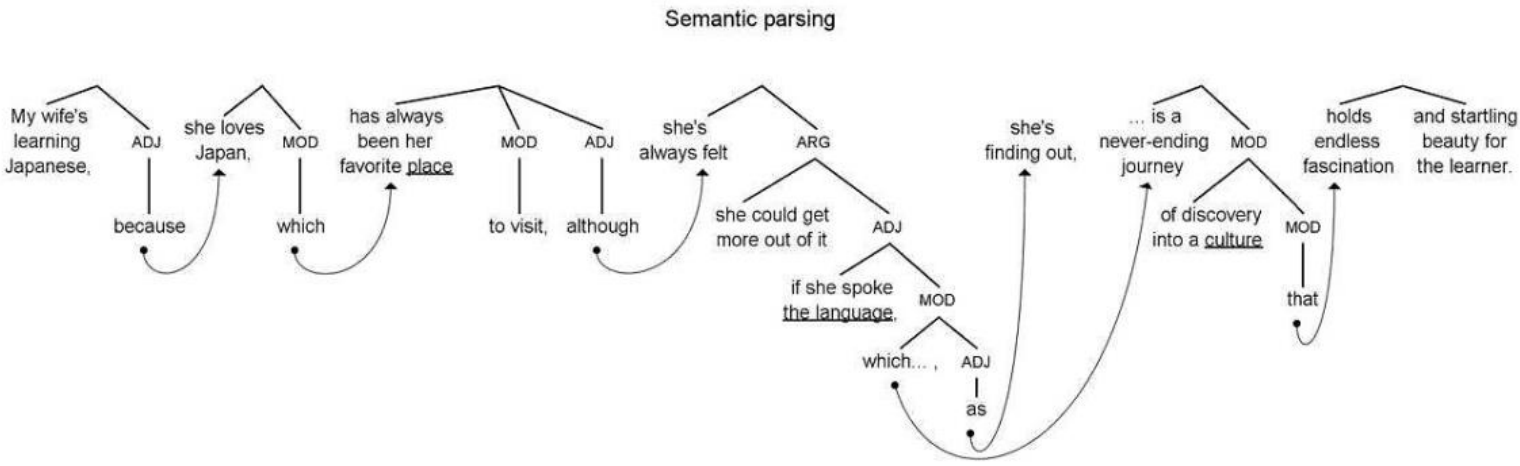

Figure 93
Semantic parsing of complex sentence showing shifting focus of attention

3.2 Examples from the corpus

Let's see some examples of shifting focus of attention, with clauses that are introduced as syntactically subordinate and go on to be functionally independent, in sentences from the corpus for this study. The following examples are again taken from President Obama's 2015 speech to the UN General Assembly.

One type of sentence with shifting focus involves an assertion in an complement clause. In section 2 of this annex we saw an example of such a sentence, reproduced in (144).

(144) I recognize that democracy is going to take different forms in different parts of the world.

Based on evidence from tests of assertive force, we concluded before that each of the two clauses in (144) could be taken as an assertion in its own right. This means seeing the focus of attention as shifting as the sentence moves forward. This analysis is illustrated in figure 94.

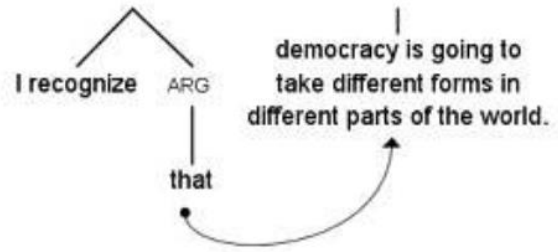

Figure 94
Focus shifting onto assertion in a complement clause

Another type of sentence with shifting focus involves an assertion in an adjunct clause, as illustrated in figure 95. The adjunct clause can be finite or non-finite.

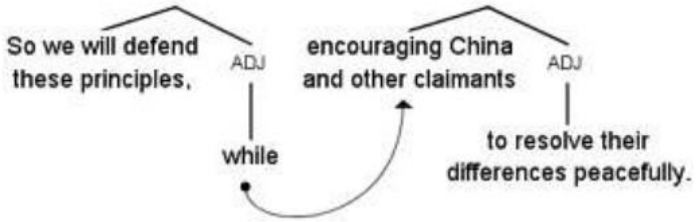

Figure 95
Focus shifting onto assertion in an adjunct clause

The second clause in figure 95 is a non-finite clause. But it can still be taken as a functionally independent assertion about what the US administration will encourage other countries to do.

The focus of attention in a sentence can also shift onto an assertion in a relative clause, as illustrated in figure 96.

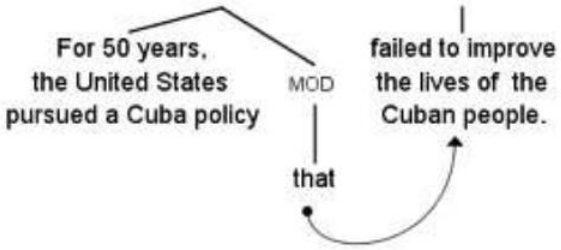

Figure 96
Focus shifting onto assertion in a relative clause

The second clause in figure 96 is a subordinate clause introduced by the relative pronoun "that." But it can still be taken as a functionally independent assertion about the effect of the US Cuba policy.

More than one type of shifting focus can occur in the same sentence, as illustrated in figure 97.

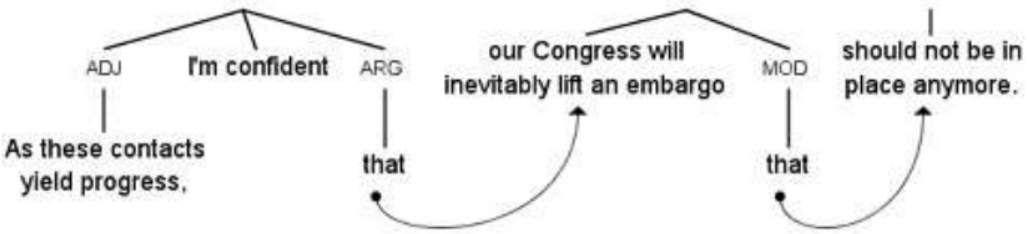

Figure 97
Multiple shifting focus

Sometimes the main clause of a sentence gives little information, but sets the context for an assertion in a subordinate clause, as illustrated in figure 98.

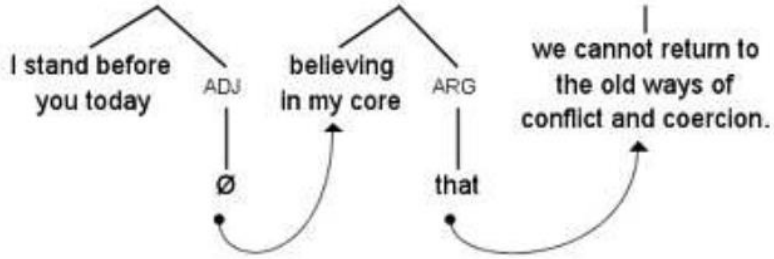

Figure 98
Main clause giving little information

The sentence in figure 98 is an assertion of the President's core belief and of the impossibility of returning to the old ways, much more than of where and when he's standing. In such cases, this study still treats the first subordinate clause as an assertion in its own right, despite its non-finite form.

A sentence can be structured as a main clause syntactically subordinating several clauses, with the focus of attention shifting progressively from one of those subordinate clauses to the next, as illustrated in figure 99.

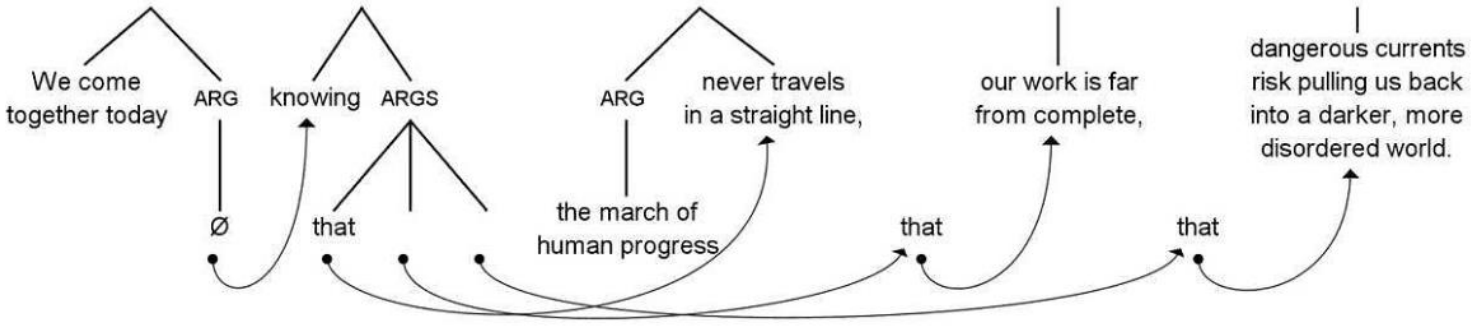

Figure 99
Focus of attention shifting from one subordinate clause to the next

The examples in figures 94-99 show how a clause introduced as syntactically subordinate can go on to be functionally independent. But the reverse can happen too. A functionally independent clause can be syntactically subordinated to a following clause, as illustrated in figure 100.

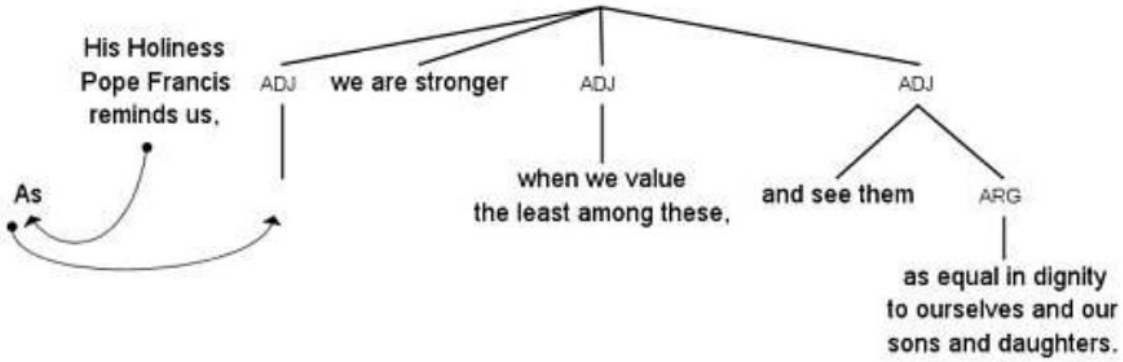

Figure 100
Functionally independent clause syntactically subordinated to the following clause

The first clause in figure 100 can be taken as an assertion in its own right of what the Pope says. This interpretation is supported by the fact that that clause can be restated as syntactically independent, as in (145) or in (146).

(145) His Holiness Pope Francis reminds us that we are stronger […].

(146) We are stronger […], and His Holiness Pope Francis reminds us of that.

Figure 100 illustrates a functionally independent clause syntactically subordinated to a following clause. In European languages, shifts of focus in this direction are largely restricted to cases where the first clause is an adjunct to its parent. That's because of the branching direction of complex sentence structure in European languages, where an adverbial clause can precede its parent (as in figure 100) or follow it. In contrast, a relative clause or a complement clause typically follows its parent.

One reason for this greater flexibility in position for adverbial clauses than for other types of subordinate clause may have to do with the way they attach to their parents syntactically. A complement clause attaches syntactically to a predicate in its parent, and a relative clause attaches to an argument of that predicate. In contrast, an adverbial clause attaches at a higher level, to a constituent consisting of the head of its parent clause plus any complement(s), in what's known as an "X-bar" relation. In any event, the most common direction of shifting

focus in a European language is for a clause introduced as syntactically subordinate to go on to be functionally independent.

In Sinitic languages like Mandarin, the branching direction of complex sentence structure is more mixed, with an adverbial clause or a relative clause typically preceding its parent, whereas a complement clause typically follows its parent. And in a largely left-branching language like Japanese, Korean or Turkish, adverbial, relative and complement clauses all typically precede their parents. In such languages, the most common direction of shifting focus is for a functionally independent clause to be syntactically subordinated to a following clause. However, this phenomenon may be more restricted in left-branching languages than the corresponding one in right-branching languages, because of factors like syntactic nesting and grammatical deranking (discussed below).

The semantic parsing model used in this study can see a syntactically subordinate clause as going on to be functionally independent, or a functionally independent clause as being syntactically subordinated to a following clause. That's useful for our analysis of functional equivalence between the original version of a sentence and a translated or interpreted version. Ideally, the status of a proposition as functionally independent or subordinate should be preserved in translation or interpretation, even if that status is syntactically established in different ways in each version. Translation or interpretation of a complex sentence based on a syntactic analysis of its structure can lead to a loss of semantic coherence, by failing to reproduce the shifts in focus from one proposition to the next.

3.3 Deranking

European languages can be flexible in allowing subordinate clauses to be functionally independent, as described above. One reason for this flexibility is that many subordinate clauses in European languages – whether relative, complement or adjectival – retain key features of a main clause, such as a finite verb form and tense. The only thing that distinguishes such a subordinate clause from a main clause is that it may be headed by a complementizer (like "that," "for … to" or "whether"), or that an argument may be replaced by a relative pronoun moved to the beginning of the clause. This similarity in form between a main and a subordinate clause makes it easy for a clause introduced as syntactically subordinate to go on to be functionally independent. Examples are shown in (147)-(149), along with tag questions which confirm that the subordinate clauses can be taken as functionally independent assertions.

(147) Your teacher says you've been skiving off again. [haven't you?]

(148) He's put all his stories in a book, which he's planning to have published. [isn't he?]

(149) My wife's learning Japanese, because she loves Japan. [doesn't she?]

If a subordinate clause can turn out to be functionally independent in a European language, that's even more the case in a Sinitic language like Mandarin. That's because clause links (like many grammatical relations) in such languages are often covert, so they can be seen ambiguously as syntactic coordinators or subordinators. An example is shown in (150).

(150) 他 把 所有的 故事 都 寫 在 書 裏 了, 打算 出版。
Tā bǎ suǒyǒude gùshi dōu xiě zài shū lǐ le, dǎsuàn chūbǎn.
He obj all story all write in book inside perf, plan publish.
He's put all his stories in a book, and is planning to have it/them published. [isn't he?]
or: He's put all his stories in a book, which he's planning to have published. [isn't he?]

This syntactic ambiguity is illustrated in figure 101.

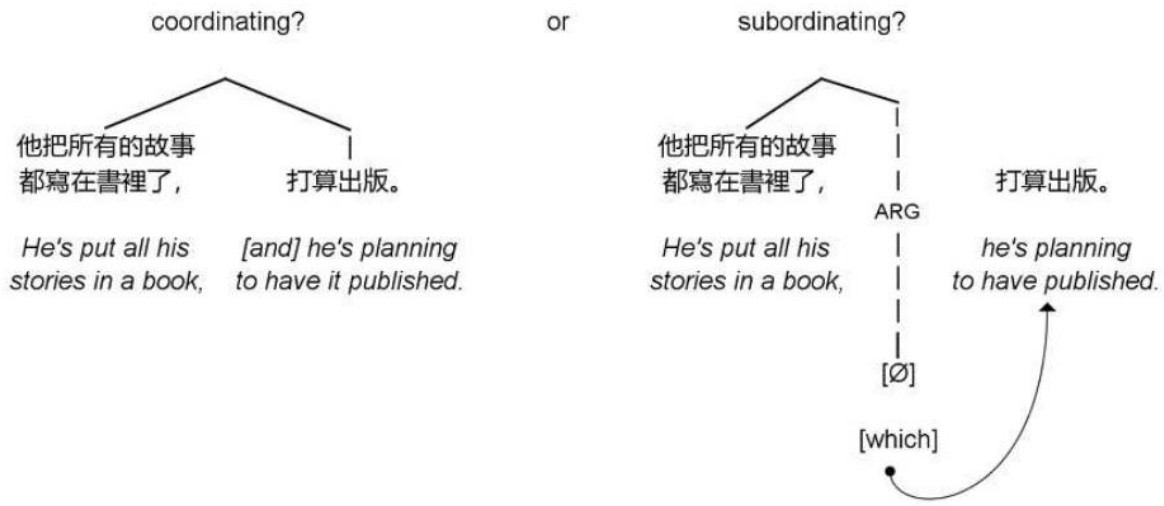

Figure 101
Coordinate clause or subordinate clause which goes on to be functionally independent in Mandarin

Similar ambiguity of syntactically coordinating or subordinating function can also be seen in Japanese clause chains. An example is shown in (151).

(151) 妻は　　日本が　　大好きな　　ので、　　日本語を　　勉強しています。
Tsuma-wa Nihon-ga daisuki-na no-de, Nihongo-o benkyou-shiteimasu.
Wife-top Japan-subj please greatly and/so/since Japanese-obj is learning.
My wife loves Japan, and/so she's learning Japanese.
or: Since my wife loves Japan, she's learning Japanese.

This syntactic ambiguity is illustrated in figure 102. The numbers on the second tree show the order the branches need to be read in to make sense in English.

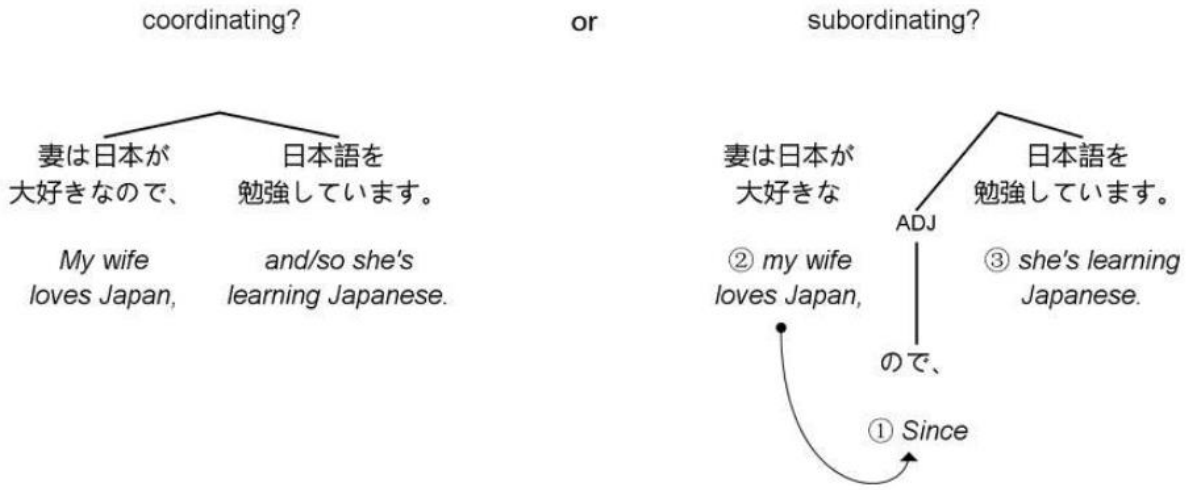

Figure 102
Coordinate clause or functionally independent clause which goes on to be syntactically subordinate in Japanese

The difference between the Japanese example in figure 102 and the English and Mandarin ones in figures 94-99 and 101 is that the Japanese link, if seen as a syntactic subordinator, works the opposite way. It takes a clause that was functionally independent – "my wife loves Japan" – and makes it syntactically subordinate to the following clause – "she's learning Japanese." In the Japanese sentence in figure 102, the first clause can be seen as either main or subordinate. That's because it retains the key features of a main clause, with the exception of being headed by a final (coordinating or subordinating) conjunction.

In some languages, such shift of focus is harder to achieve, as subordinate clauses tend to be grammatically deranked. That is, they're often formed in various ways to reflect their subordinate, background, non-assertive function. Bril (2010) discusses the phenomenon of deranked subordinate clauses in various languages, including the use of participles and other non-finite verb forms, as well as nominalized verb forms with adpositions and case markers. Dependent events can sometimes be expressed in non-finite or nominalized forms in European languages too. But in other languages, deranking for subordinate clauses is the rule.

This is true to some extent in Japanese, where verb forms in subordinate clauses may be deranked for politeness and topic markers may be absent. It's even more the case in Turkish, where most verb forms in subordinate clauses are deranked for tense and converted into nominal or adjectival forms, with subjects expressed as possessors. Examples are shown in (152)-(154), including literal English translations and proper English equivalents for comparison.

(152) Öğretmenin, *yine okulu astığını* söylüyor.
Your teacher says [lit.] "*your skiving off again*." [possessive nominal]
cf. Your teacher says *you've been skiving off again*. [finite verb]

(153) Bütün hikâyelerini, *yayımlatmayı düşündüğü* bir kitapta topladı.
He's put all his stories in a book [lit.] "*his planning to have published*." [poss. participle]
cf. He's put all his stories in a book, *which he's planning to have published*. [finite verb]

(154) Karım *Japonya'yı çok sevdiği için* Japonca öğreniyor.
My wife's learning Japanese [lit.] "*because of her loving Japan*." [possessive nominal]
cf. My wife's learning Japanese, *because she loves Japan*. [finite verb]

It may be possible to perceive a grammatically deranked clause in a language like Japanese or Turkish as an assertion. But that can be a lot harder than with a verb form that could appear in a main clause. That's because a deranked clause doesn't contain a string that can be directly taken as an assertion, negated or questioned. So it requires some mental reconstruction, including presumption of features such as tense and arguments, for the listener or reader to figure out what's being asserted.

The subordinate clauses in the Turkish sentences in (152)-(154) are deranked not only grammatically, but syntactically too. Each of those clauses is nested, syntactically splitting the predicate and arguments of its parent. Subordinate clauses can often be both grammatically deranked and syntactically nested in a language like Japanese or Turkish. Both these features can make such clauses harder to perceive as functionally independent than would be the case in a European language. So a syntactically accurate written translation of a complex sentence from a European language into a language like Japanese or Turkish can fail to reproduce the multiple assertions the original was perceived as making, resulting in a sentence where there appears to be only one main assertion.

To illustrate this, let's revisit the hypothetical run-on English sentence we saw above about the speaker's wife learning Japanese. That sentence is reproduced in (155).

(155) My wife's learning Japanese, because she loves Japan, which has always been her favorite place to visit, although she's always felt she could get more out of it if she spoke the language, which, as she's finding out, is a never-ending journey of discovery into a culture that holds endless fascination and startling beauty for the learner.

If we try to translate the sentence in (155) into a language like Japanese or Turkish by copying the syntactic structure of the English original, as a single sentence with one main clause and several subordinate clauses, the subordinate clauses are likely to be grammatically deranked. To show the problems this can create, a parse tree with a syntactically accurate Turkish translation of (155) can be seen in figure 103. The numbers show the order the branches need to be read in to make sense in English.

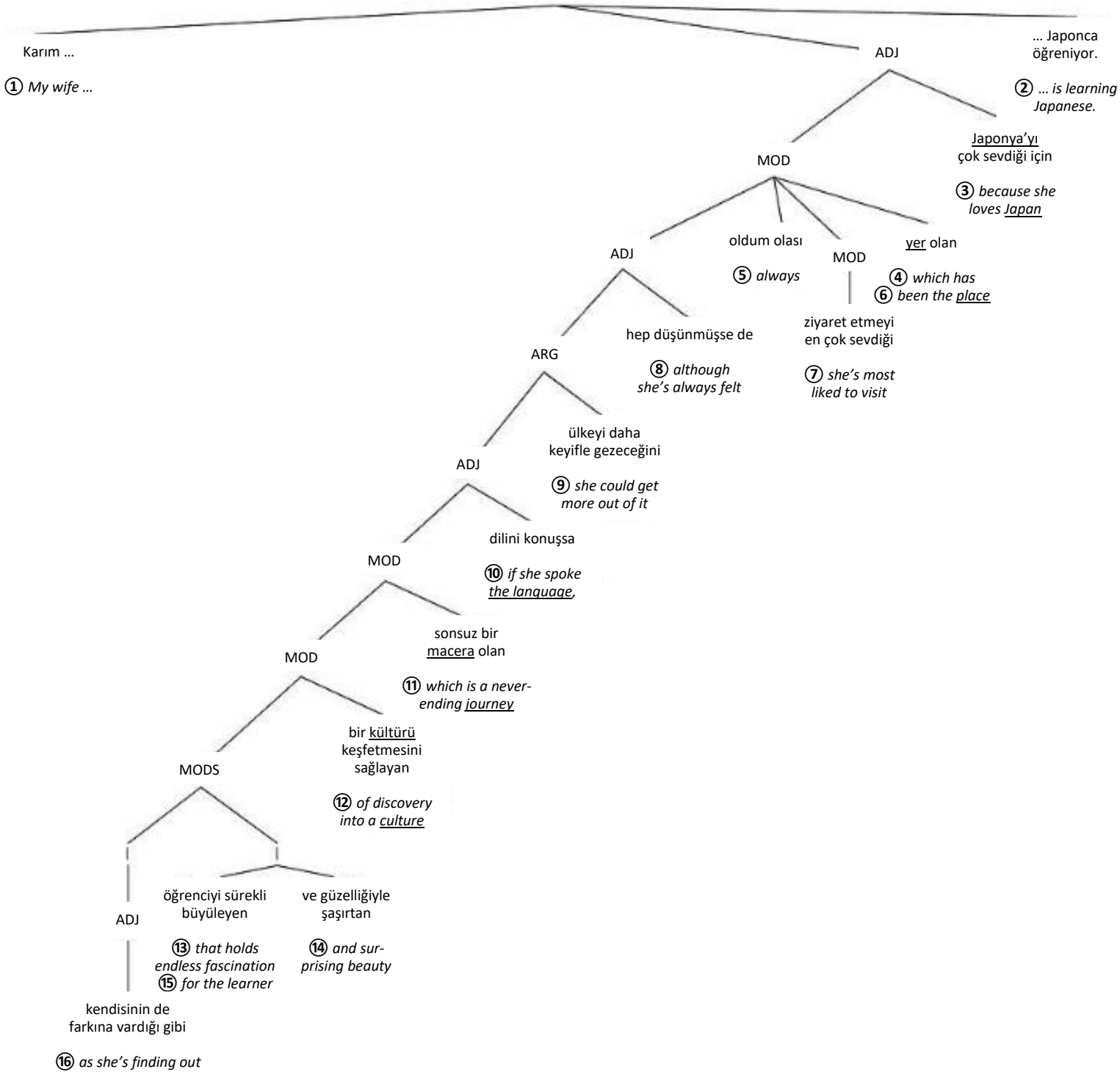

Figure 103
Syntactically accurate Turkish translation of complex sentence

Each of the clauses in the syntactically accurate translation in figure 103 is rendered nicely and idiomatically in Turkish. But the overall effect may be disturbing, for a number of reasons. The first reason is the big gap at the top level of the tree between the subject and predicate of the single syntactic main clause. Such dependency gaps can often occur in translation between languages with very different structure, as we saw in section 3.3.3 on nested structures. A second problem with the Turkish translation in figure 103 is that it doesn't recreate all the semantically flexible links which appeared in the original English version in (155). Instead, the translation reflects an analysis of the syntactically subordinate clauses as being functionally subordinate as well, because they have deranked verb forms and are nested within the single syntactic main clause.

Like subordinate status, assertive status is relative and can be hard to pin down. As explained above, the fact that the subordinate clauses in figure 103 are grammatically deranked makes it impossible to apply any empirical assertion test to them, like a tag question or a negation test. That suggests that their assertive force is weakened.

A third problem with the translation in figure 103 is the order in which the assertions – to the extent that they're seen as assertions in their weakened form – are presented. This problem is discussed in the section on logical order in the epilogue, which suggests an alternative translation to solve these problems.

We now have a semantic, functional method for analyzing complex sentence structure, which reflects the view that the focus of attention in a complex sentence can shift as the sentence moves forward. In a complex sentence with no nestings, each proposition can generally be processed before or as soon as the next one begins. Our model also sees many syntactically subordinate clauses as being functionally independent, rather than positing a complex hierarchical structure with multiple layers of subordination. That means that logical windows for processing propositions are generally seen here as successively opening and closing, without ever getting too big or too deep.

This section has illustrated various aspects of the semantic parsing model used in this study. The model doesn't ensure black-and-white parsing decisions for every sentence. But it's fit for its purpose – as a technical tool for comparing original and translated or interpreted versions of complex sentences in terms of the linear order of propositions and the hierarchical relations between them. The guidelines for using that tool are summarized in the next section.

4. Parsing guidelines

Below is an overview of the guidelines followed in parsing the original English version and the translated or interpreted versions of each sentence in the corpus for this study.

4.1 Segmenting the original English version of a sentence

1. Any syntactic phrase with a predicate and at least one argument or adjunct is segmented as a proposition. One proposition can be part of another. Each proposition is enclosed in square brackets, followed by a superscript number. That number indicates the linear position of the proposition in the sentence.

[We should always help others.][1]

[Helping others][1] [is a reward in itself.][2]

2. A coordinating link between propositions is included in the segment for the proposition that follows it. A subordinating link is included in the segment for the subordinate proposition.

[This tool slices][1] [*and* dices.][2]

[This tool is useful][1] [*for* peeling potatoes.][2]

3. Multiple arguments are included in the same proposition.

[We're having *turkey, stuffing and sweet potatoes*.][1]

4. Multiple predicates are segmented in separate propositions. Arguments, adjuncts and links shared by more than one proposition are included in the proposition they're syntactically contiguous with and implied in the other one.

[at the time of *signature*][1] [or *ratification* of the treaty][2]

5. A proposition that syntactically splits the predicate of another proposition from any part of its arguments counts as a nesting. A proposition that just splits an adjunct or a link from the rest of another proposition doesn't count as a nesting. A split proposition has both parts enclosed in brackets. The part with the predicate gets a normal black number. The number for the other part is gray and doesn't count in the numbering order.

[The cat][2] [*the dog was chasing*][1] [ran up the tree.][2]

A proposition that splits the predicate and arguments of another proposition also counts as a nesting if either of the surrounding parts is a proposition in its own right. In that case, each proposition is numbered separately.

[Respecting people's privacy][1] [*when handling personal data*][2] [is a key concern.][3]

6. A modifier with at least one argument or adjunct is segmented as a separate proposition.

[The][2] [rapidly changing][1] [climate is a huge challenge.][2]

[protection of data][1] [concerning *people's private lives*][2]

Modifiers with no arguments or adjuncts are included in the same proposition as the noun they modify.

[protection of *people's private* data][1]

7. A shortened relative clause is segmented as a separate proposition.

[The books][2] [*on the table*][1] [are mine.][2] (= [The books][2] [*which are on the table*][1] [are mine.][2])

but: [Leave the books *on the table*.][1]

8. A semantically weak predicate that can be omitted in rephrasing isn't segmented separately.

[if *there's* anything you need][1] (≈ [if you need anything][1])

[if such an act *occurs* as a result of *using* force][1] (≈ [if such an act results from force][1])

9. An adjunct providing typical information – beneficiary, accompaniment, resultant, instrument, location, goal, time (except for descriptions of events), manner and measure (Larson 1984: 219-223) – is included in the same proposition as its predicate.

[I wrote this poem *for you*.][1]

[I went to the movies *with my boyfriend*.][1]

[I painted the room *pink*.][1]

[We usually speak *on FaceTime*.][1]

[Let's meet *in the park*.][1]

[We did it *for fun*.][1]

[Let's meet *tomorrow afternoon*.][1]

[We can meet *in person*.][1]

[I'm *totally* convinced.][1]

An adjunct providing information other than typical information as defined above can generally be rephrased as a clause and is segmented as a separate proposition. In English and many other languages, such a proposition tends to be separated from its parent by a pause in speaking or a comma in writing.

[Because of these concerns,][1] [experts are advising caution.][2]

[Many are willing to try,][1] [despite these concerns.][2]

The same goes for an adjunct with the description of an event.

[*After the meeting*,][1] [we'll go home.][2]

(≈ [*After we meet / After the meeting finishes*,][1] [we'll go home.][2])

10. A phrase which is an adjunct to an entire proposition (not just its predicate) is segmented as a separate proposition if it has an argument.

[*In addition to these concerns*,]^1 [there's a mistrust of anything new.]^2

Otherwise, it's included in the rest of the proposition it's an adjunct to.

[*Clearly*, you need a vacation.]^1

11. A phrase headed by a noun in apposition with another noun is segmented as a separate modifying proposition.

[We're joined by Ms Nakamura,]^1 [Deputy Director for Marketing.]^2

The same goes for a phrase with a head like "including" or "such as" which modifies a noun.

[Members discussed various issues,]^1 [*including rules, fees and penalties.*]^2

11. A comment clause (a main clause used as a formulaic expression of the speaker or writer's attitude to an assertion made in a subordinate clause) isn't segmented as a separate proposition.

[*I think* it's going to rain.]^1 (≈ [It'll probably rain.]^1)

12. A process nominal (a nominal describing a process, event or situation) is segmented as the predicate of a separate proposition if it has any arguments or adjuncts.

[Bavaria is known]^1 [for *car manufacturing*.]^2 (≈ [the manufacturing of cars]^2)

[We need to mitigate]^1 [*climate change*.]^2 (≈ [the changing of the climate]^2)

Though a phrase like "climate change" is a set term with an established equivalent in other languages, it has argument structure, so it's segmented here as a separate proposition, as explained before.

In contrast, a result nominal (a nominal describing the result of an action or something created by an action) is treated as a simple noun, even if it's modified.

[We have common *responsibilities*.]

The distinction between process nominals and result nominals was discussed in section 1.7 of this annex.

13. For two predicates in an asymmetric relation with shared arguments, there's a continuum from nearly total propositional integration to increasing autonomy. As a guide, if each predicate can be modified separately, they're segmented in separate propositions. Otherwise, they're included in the same proposition. There are some borderline cases.

According to this test, both predicates are included in the same proposition in these sentences:

[He *has* to *leave*.] (✓[He *always* has to leave.] *[He has] [*always* to leave.])

[He *lets* me *stay*.] (✓[He *always* lets me stay.] *[He lets me] [*always* stay.])

[She *wants* to *help*.] (✓[She *always* wants to help.] *[She wants] [*always* to help.])

[He *forced* me to *leave*.] (✓[He *always* forced me to leave.]
?[He forced me] [*always* to leave.])
[She *tries* to *help*.] (✓[She *always* tries to help.] ?[She tries] [*always* to help.])

And each predicate is segmented in a separate proposition in these sentences:

[He *convinced* me] [to *stay*.] (✓[He *always* convinced me] [to stay.]
✓[He convinced me] [*always* to stay.])
[I *promised*] [to *help* her.] (✓[I *always* promised] [to help her.]
✓[I promised] [*always* to help her.])
[I'll *help* you] [*look* great.] (✓[I'll *always* help you] [look great.]
✓[I'll help you] [*always* look great.])
[I *said*] [I'd *help* her.] (✓[I *always* said] [I'd help her.]
✓[I said] [I'd *always* help her.])
[We *reaffirm*] [our *faith*.] (✓[We *always* reaffirm] [our faith.]
✓[We reaffirm] [our *constant* faith.])

4.2 Segmenting a translated or interpreted version of a sentence

The purpose of this parsing method is cross-linguistic comparison. So each proposition in a translated or interpreted version of a sentence has the same number as the corresponding proposition in the original English version.

Also, all translated or interpreted versions of a sentence are parsed as similarly as possible to the original English version. This means:

1. A phrase in a translated or interpreted version of a sentence isn't segmented as a separate proposition if it doesn't correspond to a separate proposition in the original English version, even if it would be segmented separately if the parsing method was applied directly to the translated or interpreted version.

2. A phrase in a translated or interpreted version of a sentence is segmented as a separate proposition if it corresponds to a separate proposition in the original English version, even if it wouldn't be segmented separately if the parsing method was applied directly to the translated or interpreted version.

3. If a proposition in a translated or interpreted version of a sentence doesn't include or imply a predicate or an argument included in the corresponding proposition in the original English version, or if it includes a predicate or an argument not included or implied in the original

version, the segment number of the proposition in the translated or interpreted version is followed by a "Δ" ("delta" for "change").

English: [Such an act is a violation of this Treaty.][1]

→ other language: [Such an act is punishable.][1Δ]

4. If a proposition in a translated or interpreted version of a sentence doesn't correspond to any part of the original English version, it's marked by an "X" instead of a segment number.

English: [Such an act is punishable.][1]

→ other language: [Such an act is punishable][1] [as determined by the Court.][X]

4.3 Number lines

1. Each proposition segmented and numbered as above is represented by its number in a line below the sentence.

2. The number for each subordinate proposition is labeled "ARG," "MOD" or "ADJ," according to its semantic role (argument, modifier or adjunct) in relation to its parent, followed by the number of its parent.

ARG2		ARG2	ADJ3	MOD4
1	2	3	4	5

The number line above means that propositions 1 and 3 are arguments of proposition 2, proposition 4 is an adjunct to proposition 3, and proposition 5 modifies a noun in proposition 4.

3. A proposition containing reported speech or thought is labeled "REP", followed by the number of the proposition (if any) where its perspective is established.

	REP1	ARG2	ADJ3	MOD4
1	2	3	4	5

The number line above means that proposition 2 contains reported speech or thought, and that its perspective is established in proposition 1.
A statement or thought which the speaker or writer identifies with isn't treated as reported speech or thought, as explained earlier in section 1.6 of this annex.

4. For a split proposition, where the part containing the predicate has a normal black number in the segmented text and the other part has a gray number, only the normal black number is copied onto the number line.

5. The number for a nested proposition (a proposition that syntactically splits the predicate of another proposition from any part of its arguments) is enclosed in curly brackets.

ARG2		ARG2	ADJ3	MOD4
1	2	3	{4}	5

The number line above indicates that proposition 4 is nested.

6. If the label over a number in the number line for a translated or interpreted version of a sentence doesn't match the label above the same number in the number line for the original version, that number is followed by a "Δ" ("delta" for "change") in the number line for the translated or interpreted version.

ARG2		ADJ2	ADJ3	MOD4
1	2	3Δ	4	5

The number line above indicates that proposition 3 is attached with a different relation or to a different parent in a translated or interpreted version of a sentence compared to the original version.

7. If a proposition is marked with a "Δ" or an "X" in the segmented text for a translated or interpreted version of a sentence (see points 3 and 4 of section 4.2 above), its number is followed by a "Δ" in the number line for that version.

ARG2		ADJ2	ADJ3	MOD4
1	2	3Δ	4	5

The number line above indicates that the predicate and arguments of proposition 3 of a translated or interpreted version of a sentence don't match those of the original version, or that proposition 3 of the translated or interpreted version doesn't correspond to any part of the original version.

8. If a proposition in the original version of a sentence has no corresponding proposition in a translated or interpreted version, its number appears at the end of the number line for that version, but in gray, followed by a "Δ".

ARG2		ADJ3	MOD4	
1	2	4	5	3Δ

The number line above indicates that proposition 3 of the original version of a sentence is missing in a translated or interpreted version.

4.4 Sentence division

Annex II to the online version of this study contains a display page with a data table for the original English version and the five translated or interpreted versions of each sentence in our corpus. The English version of each sentence is divided and punctuated as it appears in the original text or transcript. The translated versions of sentences in the legal texts and subtitled talks are divided and punctuated as published online. The interpreted versions of sentences in the speech are divided and punctuated as transcribed by expert interpreters and interpreter trainers for the Russian, Mandarin and Japanese versions, and by myself for the Hungarian and Turkish versions. Transcribers were guided only by the recordings of interpretation, without referring to the original English version of the speech. They followed the conventions of intelligent verbatim transcription, providing an accurate written representation of each recording, while editing out fillers and corrections.

The translations and transcriptions have some differences in sentence division from the original English texts and transcripts. Longer sentences in the original English version of a legal text or subtitled talk were sometimes split into two or more shorter sentences in translation. This may have been due to individual stylistic choice by the translator. In subtitle translation, it may sometimes have been due to the need for subtitles to match the video image or to space restrictions. Guidelines for translating TED talks instruct translators to keep subtitles to at most two lines of 42 characters each, and reading speed to at most 21 characters per second.

Similarly, longer sentences in the original English version of the interpreted speech as transcribed and published were sometimes split into two or more shorter sentences in a transcription of recorded interpretation. This may have been due to individual stylistic choice by the transcriber. It may also reflect syntactic and prosodic breaks made by the interpreter to reduce cognitive load. Sometimes two or more sentences in the original English transcription of the speech were combined into one longer sentence in a transcription of recorded interpretation. And there were borderline cases, where a subjective decision may have been made as to whether to transcribe a certain passage as one or more written sentences.

In general, such changes in sentence division don't affect the values of variables recorded here for statistical analysis. This is the case when an extra sentence division in one language appears between propositions which are functionally independent in the other language too, as shown in figure 104.

English: [I'm here]^{1} [because I want]^{2} [to help.]^{3}

1 2 ^{ARG2}3

Other language: [I'm here.]^{1} [I want]^{2} [to help.]^{3}

1 2 ^{ARG2}3

Figure 104
Change in sentence division – no effect on variables recorded

The English version in figure 104 is written as a single sentence with an overt link between propositions 1 and 2. The other language version is written as two separate sentences with no overt link between those propositions. But in both versions, propositions 1 and 2 are functionally independent, and proposition 3 is an argument of proposition 2. Whether functionally independent propositions are written in the same sentence or in different sentences has no impact on the method used here for showing linear and hierarchical relations between propositions and recording counts for our dependent variables. Neither does the presence or absence of an overt link between propositions.

Sometimes a change in sentence division results from the appearance in translation or interpretation of a proposition which doesn't correspond to any part of the original version. This happens when information not contained in the original is added, or when information implied in the original is made explicit. Sometimes the opposite happens: a change in sentence division results from the absence in translation or interpretation of a proposition contained in the original version. This happens when information contained in the original is omitted, or when information explicitly stated in the original is made implicit. Any such change is reflected in the number lines for the original English version and the other language version. And it's recorded as a change in semantic relations, as shown in figures 105 and 106.

English: [I'm here]^{1} [to help.]^{2}

1 ^{ADJ1}2

Other language: [I'm here]^{1} [to help]^{2} [tidy up.]^{X}

1 ^{ADJ1}2 ^{ARG2}XΔ

Figure 105
Addition of proposition – recorded as change in semantic relations

English: [I'm here]^{1} [to help]^{2} [tidy up.]^{3}

1 ADJ1 2 ARG2 3

Other language: [I'm here]^{1} [to help.]^{2}

1 ADJ1 2 3Δ

Figure 106

Removal of proposition – recorded as change in semantic relations

4.5 Segment size

The method for parsing sentences into propositions described in sections 1-3 of this annex is applied to the original English version of each sentence in the corpus. For ease of comparison, the other language versions of that sentence are divided into segments corresponding in information content to the propositions in the original English version. That means the segments in a translated or interpreted version of a sentence may not correspond exactly to propositions as they would be identified if the parsing method was applied directly to that language version.

The reason for doing this, rather than applying the parsing method to each language version of a sentence directly, is to avoid changing the number of segments when the translation or interpretation contains the same information as the original. An example is shown in (156).

(156) English: inalienable rights Other language: rights which can never be taken away

In the English version of the phrase in (156), the modifier "inalienable" has no arguments or adjuncts. So our parsing method wouldn't segment it as a separate proposition. Accordingly, the relative clause "which can never be taken away" in the other language version wouldn't be segmented here as a separate proposition either, even though a relative clause would be treated as a separate proposition if the parsing method was applied directly to that version. Care has been taken to ensure that the segments in each translated or interpreted version of a sentence match the segments in the original English version when the information content is judged to be the same, as in (156). This minimizes the impact of minor phrasing differences between languages on the values for variables in our data.

Sometimes a segmenting mismatch between the original English version and another language version of a sentence does affect the number of segments recorded. Examples are shown in (157) and (158).

(157) English: my older brother Other language: my brother, [who's older than me]

(158) English: Paris, [which is in France] Other language: Paris, France

In the English version in (157), the modifier "older" has no arguments or adjuncts. In the other language version in (158), the same is true of the modifier "France." So our parsing method wouldn't treat the single words "older" or "France" in those versions as separate propositions. But in (157), "older" corresponds in the other language version to a functionally independent proposition which makes an assertion. And in (158), "France" corresponds in the English version to a functionally independent proposition which makes an assertion. Because of this change in semantic status, the relative clauses in those versions would be segmented here as separate propositions.

Sometimes segmenting decisions are borderline, as in (159).

(159) English: [There are many people(] [)who love jazz.]
Other language: [Many people love jazz.]

The English version in (159) could be treated either as one or as two propositions. The choice would depend on whether the semantically weak predicate "there are" is taken as the predicate of a functionally separate proposition. In a case like this, if the same information content was expressed in a single proposition in most other language versions, the method used in this study would treat the English version as a single proposition too.

Our analysis doesn't measure the length or the content of propositions, just the linear and hierarchical relations between them. Some text chunks segmented as separate propositions in English can be quite short, as in (160).

(160) [climate change] goals

The treatment of nominal predicates, like "change" in "climate change," was discussed in section 1.7 of this annex.

The equivalent of a phrase like (160) in another language may be longer, and may literally mean something like (161).

(161) goals [relating to dealing with the effects of climate change]

If the parsing method used in this study were applied directly to the longer phrase in (161), that phrase could be segmented as several separate propositions, as in (162).

(162) goals [relating] [to dealing with the effects] [of climate change]

Again, the parsing method used in this study is applied to the original English version of each sentence. If it were applied directly to the other language versions, one consequence could be higher counts for changes in semantic relations in some languages, because some of the other languages considered can be wordier than English. Also, the Turkish, Mandarin and Japanese versions of sentences in our corpus already show much higher rates of nesting than the Russian and Hungarian versions. Applying our parsing method directly to the translated or interpreted versions of sentences could well lead to even higher nesting rates for languages like Turkish, Mandarin and Japanese, which often have structures with multiple nestings.

A short English phrase like (160) may correspond to longer phrases in other languages. The components of those phrases may be sequentially ordered as shown in (163) and (164).

(163) countries' [relating] [to dealing with [climate change] effects] goals (Mandarin)

(164) countries' [climate change] [effects-with-dealing-to] [related] goals (Turkish, Japanese)

Despite such differences, if the original English version of a sentence is segmented as one proposition – or two or three – and if the equivalents of those propositions in other language versions are judged to have the same information content and functional status as in English, those versions are divided into the same number of segments as the English version. This makes the various language versions of each sentence easier to compare, minimizing the impact of minor phrasing differences between languages on the values of variables in our data. This represents a deliberately conservative choice to refrain from recording some information, in order to avoid any suggestion that the counts for indicators of difficulty in structurally different language pairs – which are already comparatively high – have been inflated by the inclusion of irrelevant data.