The present paper is concerned primarily with translational correspondencies between sentences on the one hand and sentence sequences on the other hand, within the general theoretical setting of Discourse Representation Theory (Kamp/Reyle 1993, Asher 1993) but without technicalities. To some extent correspondencies between clausal and non-clausal (‘phrasal’) sentence constituents will be considered, too. Both are instantiations of translation relations between pairs of texts (text extracts) that differ with respect to what I have elsewhere (Fabricius-Hansen 1996, 1998) called informational density. It will be shown, however, that they illustrate different aspects or parameters of informational density – *incrementality* of discourse information vs. *explicitness* (redundancy) of local information (section 2). It is argued (section 3) that a target sentence sequence will normally contain a designated member, the *Principal Counterpart* of the source sentence. That allows us to view sentence splitting as *forward* and *backward information extraction*, depending on the position of the Principal Counterpart; variants of these two sentence splitting techniques are discussed in section 4. Section 5 discusses the relevance of the (S)OV - (S)VO parameter for sentence splitting (and clausal expansion) (section 5.1) and the effects of extensive sentence splitting on global discourse structure (5.2). Section 6 contains a brief summary and some suggestions for further research.

1. Introduction: types of translational correspondences

Depending on the degree of relatedness between source and target language and the type or genre of the text to be translated, (machine) translation demands different modes or algorithms of increasing complexity. Correspondingly, comparing source and target texts, we may distinguish 3–4 different types of translational correspondences (Dyvik 1995, Thunes 1998; cf. also Hauenschild 1986).

Thus, languages that are very intimately related – dialect variants – may often be translated literally word by word, adjusting morphology but without consulting semantic representation; given her/his/its general knowledge of how the two language systems are related to each other lexically and syntactically, the translator or translation system only needs to know the syntactic structure of the string to get the translation right. This simple type of translational correspondence – Type 1, illustrated in (1) – will be of no further interest here.

(1) a Mennen hente straks boka. (Norw.)
Manden hente de straks bogen. (Dän.)

Man DEF fetched at-once book DEF

“The man at once fetched the book.”

For a pair of languages showing a high degree of lexical and syntactic correspondence but differing in some syntactic respects like, for instance, their basic word order, the translation algorithm cannot just proceed word by word; rather, it will have to change certain syntactic constructions in the source string when producing the target string, the result being a translational correspondence of Type 2 as illustrated in (2). (Translating simple German sentences into Norwegian may often work that way, in fact.)

(2) a [Der Mann] hat [das Buch geholt].

b [Mannen] har [hentet boka]

“The man has fetched the book.”

But mostly, translation in this syntactic mode will not suffice either. Lexical and/or structural discrepancies between the languages in question will force the translator or translation system to consult not only the syntactic structure but also the semantic representation of the source string as well as the syntactic rules of the target language in order to derive a semantically equivalent target string; and he/she/it may even need information that is not contained in the source string itself: information from the wider linguistic or extralinguistic context, general or domain-specific background knowledge etc. Translational correspondences resulting from such a semantic or pragmatic mode of translating – correspondences of Type 3 and 4 according to Thunes (op. cit.) – can be seen in the following extracts from Konrad Lorenz’ Das sogenannte Böse (KoLo) and its published Norwegian (KoLo-TN) and English (KoLo-TE) translations (3):¹ In KoLo-TE, for instance, the source adjective ortstreuer (italics in the text) is semantically ‘spelled out’ as more firmly attached to their territories, and the prepositional phrase auf verhältnismäßig beschränktem Raum (underlined in the text), which is rendered quite straightforwardly by in a comparatively limited space in KoLo-TE, has been expanded into i et akvarium med relativt begrenset rominnhold ‘in an aquarium with relatively limited space’ in KoLo-TN, undoubtedly motivated pragmatically by the assumption that the behavior of fish is normally studied in aquaria.²

¹ The tags in the top line identify the source sentence – sentence unit no 46 of chapter 2 – and the sentences it is aligned to in the respective translations; for details, see also Fabricius-Hansen (1998:Section 6).

² As a matter of fact, the reasoning in this case is not convincing and the Norwegian translation not necessarily adequate: being ‘ortstreu’, the fish in question don’t move around very much and consequently can be studied outside aquaria, too.
Comparing fictional and non-fictional translations between English and Norwegian, Thunes (1998) is able to show or at least make it plausible that the relative frequency of the various types of correspondences depends to a certain degree on the text type; and she offers a detailed and interesting semantic subclassification of Type 3 and 4 correspondences, which cover the majority of the correspondences registered in the texts.

Dyvik’s (1995) and Thunes’ (1998) typology of translational correspondences is designed for strings that do not comprise more than one sentence or finite clause, in accordance with the sentence-oriented Situation Semantic framework it is based on. Thus in (3), it would match the main clause of the source sentence KoLo <s46> to the sentences KoLo-TN <s55> and KoLo-TE <s40>, respectively, and the subordinative clause (SO DAß …) to <s56> and <s41>, respectively; but it cannot cope with the relationship between <s46> as a whole, on the one hand, and the sentence sequences KoLo-TN <s55–56> and KoLo-TE <s40–41>, on the other hand. That would demand a semantic framework that allows for translation equivalences between single sentences and sentence sequences, or more generally: between texts as wholes, – a kind of dynamic update semantics where the meaning of a sentence is its semantic contribution to its context and the semantic representation that is eventually assigned to a sentence sequence (text) may be identical or equivalent to the representation assigned to a single sentence in the same context. That is the kind of framework that so-called Discourse Representation Theory (DRT) (Kamp/Reyle 1993, Asher 1993) offers and which will be adopted or rather presupposed here.4

3 For an overview, see Thunes (1998).
2. Clausal expansion and sentence splitting: explicitness vs. incrementality

2.1 Clausal expansion and explicitation

Comparing German non-fictional texts with their Norwegian (or English) translations one often encounters translational correspondences of the type illustrated below where a phrasal (i.e., non-clausal) constituent, for instance a prepositional phrase or a premodifying adjectival/participial phrase, is translated by a subordinate clause, thus *ceteris paribus* increasing the number of clauses in the target text as opposed to the source text but respecting sentence boundaries.\(^5\)

(4) a Er forderte mich auf, ihm genau meinen Tagesablauf … zu schildern. (cit. Solfjeld 1998:145)\(^6\)  
“He asked me to describe *my day-course* exactly to him.”

b Han ba meg gi en nøyagtig beskrivelse av hvordan min dag forløp … . (ib.)  
“He asked me to give an exact account of *how my day went.*”

(5) a …, so daß man die zu untersuchenden Vorgänge … gut beobachten kann. (cf. (3) above)  
Lit.: “… the to-be-observed patterns …”

b De atferdsmønstrene man ønsker å iaktta, kan derfor lett studeres … (ib.).  
“The behavioral patterns *one wants to observe can therefore easily be studied …*”

(6) a … ein … Ungeheuer …, das sich erst beim nahen Herankommen in eine Unzahl freundlicher Purpurmäuler auflöste. (cf. (9) below)  
Lit.: “… a monster that only on close approaching …”

b Først da dette skremslet kom ganske nær meg, løste det seg opp i et utall av vennligsinnede gryntefisk. (ib.)  
“… *when this monster came very close to me …*”

c … and only as it drew very near did it become resolved into a crowd of friendly grunts and snappers. (ib.)

How this kind of clausal expansion is carried out depends to a large extent on the syntactic category and function of the constituent giving rise to the expansion; for details see

---

\(^5\) The converse phenomenon, i.e., (adverbial) clauses rendered as non-clausal phrases in German translations (from English) is studied by Doherty (1998).

\(^6\) The examples cited from Solfjeld (1998) all come from original German non-fictional prose texts and their authorized Norwegian translations.
Solfjeld (1996) and especially (1998). In any case, the target clause, in contradistinction to the non-clausal source constituent it translates, has to express time reference (tense) and modality and it will need a subject and possibly some obligatory verb complement(s), too. Often, the information that for grammatical reasons needs to be encoded in the target clause is supplied unambiguously by the source text, whether the source sentence itself structurally determines e.g., a ‘logical’ or ‘underlying’ subject or a tense, or more pragmatic principles of text coherence and the like leave no other plausible options, given the context the sentence in question is interpreted against. In that case, the subordinate clause in the target sentence and the phrasal constituent it translates can be considered equivalent in their given contexts. That might hold, for instance, for the prepositional phrase beim nahen Herankommen and its clausal translations in (6). But the target sentence, of course, expresses some pieces of information more explicitly than the source sentence, encoding it in overt linguistic material (signs), and so, other things being equal (which they normally are not), one could therefore call the target text informationally less dense than the source text (Fabricius-Hansen 1996).

It may, however, also remain open how a given phrasal source constituent should be interpreted with respect to e.g., temporal reference or the identity of participants involved in the described situation or relation; and in that case, a clausal translation may in fact convey more information than is, strictly speaking, licensed by the source sentence, specifying information that is left underdetermined in the latter and thus representing only one possible interpretation or ‘reading’ of the source text (so-called explicitation). For instance in (5), taken from (3) above, the premodifying participial phrase (die) zu untersuchenden (Vorgänge) (‘the behaviour’ – lit. ‘happenings’, ‘events’, ‘things’) ‘to be studied’ is rendered in the Norwegian target text by a relative clause (de atferdsmønstrene) man ønsker å studere (‘the behavioural patterns) one wants to study’, explicating the type of modality beyond what is expressed in the source text.

The examples given above indicate that it may not always be easy to decide whether a target sentence containing a clausal translation of a phrasal source constituent is, in fact, informationally equivalent to or more specific than the source sentence in the given context. So let us conclude by stating that, other things being equal, translating non-clausal constituents by subordinate clauses without changing sentence boundaries will increase the amount of explicit information expressed by the target sentence as compared to the source sentence, whether this locally ‘added’ information (e.g., temporal reference, modality) is (implicitly) conveyed by the source sentence, too, given the context, or rather represents a specification of the information expressed by the source sentence relative to its context. Normally, however, this type of sententialization (rather: clausalization) will not significantly reduce the syntactic complexity (hierarchical depth) of the target sentence as compared to the source sentence, given that a normal subordinate clause either occupies the same structural positions in the syntactic tree as a non-clausal constituent or else may be considered hypotactically adjoined to its matrix clause.

---

7 The ‘source construction’ (Quellenkonstruktion) according to Solfjeld (1996, 1998).
8 The term ‘sententialization’ (Sententialisierung) is used by Solfjeld (1996, 1998) indiscriminately to cover the translational relation described here and cases where one source sentence is translated by a sequence of independent sentences (see below).
2.2 Sententialization

Now let us turn to the type of translational information splitting mentioned in section 1 and illustrated again below, i.e., sententialization in the stricter sense where a sequence of independent target sentences corresponds to a single sentence in the source text.


“When the 41 years old housewife Gerda W. was going (lit.: went) through a midlife crisis, wanting (lit.: wanted) to start a new life and go to Africa as a developmental worker, she had a typical car dream.”

b Gerda W., husmor og 41 år gammel, gjennomgikk en “middelalderskrise”. Hun ville begynne et nytt liv og reise til Afrika for å drive u-hjelp. Da hadde hun en typisk bildrøm. (Norw. transl., ib.)

“Gerda W., housewife and 41 years old, was going (lit.: went) through a midlife crisis. She wanted to start a new life and go to Africa to developmental work. Then/At that time she had a typical car dream.”

(8) a Durch die außergewöhnlich starke symbolische Ausdruckskraft des Autos benützt der Traum das Auto, um eine bestimmte Lebenssituation aufzuzeigen. (ib.)

“Due to (lit.: through) the exceptionally strong symbolic power of cars (lit.: the car), the dream uses a car (lit.: the car) to picture a specific situation in life.”

b Bilen har en uvanlig sterk symbolsk uttrykkskraft. Derfor benytter drømmen seg av bilen for å skissere en bestemt livsisjeron. (Norw. transl., ib.)

“Cars (lit: the car) have an exceptionally strong symbolic power. Therefore the dream uses a car (the car) to picture a specific situation in life.”

Likewise, in example (9) below (section 2.4), one complex sentence (s-unit⁹) in the German source text, is rendered by a sequence of three sentences (s-units) in the Norwegian translation, the last one corresponding to a relative clause in the source text. In the English target text, however, the same sentence is translated by two independent sentences divided by colon; the first one corresponds to the first sentence of the Norwegian translation

---

⁹ An s-unit in terms of the SGML mark-up language is a sentence delimited by full stop, exclamation mark, or question mark.
whereas the second one consists of two explicitly conjoined sentences corresponding to the second and third sentence of the Norwegian translation, respectively.

This of course raises the question of how to define so-called independent sentences and which role punctuation should play in that respect.

As pointed out by Givón (1984), among others, explicit clause coordination and full stop can be viewed as extremes on a scale expressing the degree to which neighbouring clauses or sentences are to be integrated into one larger (discourse) unit during processing, with verb-phrase coordination and paragraph intendment still further to the left and right, respectively; cf. fig. 1. But obviously, punctuation conventions may differ considerably from one language (society) to another.10

![Fig. 1](image-url)  
Fig. 1 Clause independency continuum

The traditional distinction between clause subordination or hypotaxis on the one hand and non-subordinating juxtaposition or parataxis on the other hand is not very clear-cut, either. In written language, we rather seem to have to do with a continuum or prototypically structured area stretching from clauses that are marked as ‘subordinate’ by way of word order or specific subjunctions (complementizers) and fully integrated (embedded) into the matrix clauses, syntactically as well as semantically, to syntactically independent self-contained sentences following one another and divided by full stop.11

I shall not try to solve the categorization problems outlined here. But in what follows, semicolon and colon are treated as sentence division marks along with full stop etc.; and so are commas before clauses that are not evidently subordinated or conjoined.

2.3 Principles of Incremental Discourse Organization (PIDO)

In structural terms, sentences typically are domains of e.g., scoping and binding phenomena. And it seems reasonable to assume that the sentences making up a written text or some smaller section like, for instance, a paragraph are, in principle, processed sequentially, i.e., linearly and one by one, so that a mental representation of the information contained in the text is built up gradually, incrementally, each sentence updating the representation (the Discourse Representation Structure, DRS, in terms of Discourse Representation Theory) construed that far – and the more definitely so, the more to the right on the division scale of fig. 1 the sentences are located. That is, I shall assume that independent sentences divided by full stop, at least, are processed separately, as a kind of updating atoms, without wanting

10 Cf. Santos (1998) who discusses the theoretical problems such contrasts cause for multilingual corpus analysis.

to exclude the possibility that coordinated clauses may, under certain conditions, function in a similar way. But on the whole, coordination is left out in the discussion below.

Normal complex sentences, on the other hand, i.e., sentences consisting of a main clause and one or more clearly subordinate clauses, may be taken to be processed ‘on line’ like simple sentences, meaning that the whole sentence will have to be processed before updating of the current mental representation (DRS) of the discourse can take place.\textsuperscript{12} Syntactic parsing and the parallel construal of a ‘local’ semantic representation must be finished before updating of the context can occur; but as with simple sentences, of course, the preceding context represented in DRS construed so far is relevant for anaphoric resolution (Garrod 1994), presupposition justification (Kamp/Roßdeutscher 1994) and the like.

Suppose, for instance, that the German text containing the complex sentence (a) in (7) – a text on women’s dreams (cf. Solfjeld 1998:111ff) – and its Norwegian translation are assigned equivalent DRSs *Ki and Kj up to the occurrence of (a) and (b), respectively.

Then (a) will update *Ki in one single step, yielding the DRS *Ki+1, whereas the Norwegian text will go through three steps of updating, corresponding to the three sentences in (b), the result Kj+3, however, being equivalent to *Ki+1; cf. (a) and (b) in fig. 2.\textsuperscript{13}

\begin{verbatim}
(a)  *Ki
     .... n
     ...
     ...
     ...

     ***

     *Ki+1
     .... n x y t ev1 ev2 ev3
     ...
     ...
     ...
     Gerda W. (x)
     housewife(x)
     41-years-old(x,t)
     ev1: go-through-midlife-crisis(x)
     Africa(y)
     ev2: want-to-start-new-life & go-to-y-as-developmental-worker(x)
     t < n
t included in ev1
t included in ev2
ev3: have-a-typical-car-dream(x)
ev3 included in t

Fig. 2a DRS updating for (7a)
\end{verbatim}

\textsuperscript{12} Explicit coordination without comma may also be taken to express that the ‘local’ DRS being construed is not yet finished.

\textsuperscript{13} These DRSs are very much simplified; n: speech-time variable, ev: variable over eventualities, i.e., events, states or activities; t: time variable.
That is, I take it that complex sentences are, in principle, processed like simple sentences containing phrasal constituents instead of subordinate clauses only that in the latter case, as illustrated for instance by (4a) and (6a), the context will normally play a more active role, specifying information that is underdetermined or not linguistically encoded in the phrasal constituent itself (cf. section 1). It should, however, be stressed that there are less prototypical variants of subordination where the subordinate clause follows its matrix clause and is not semantically in the scope of any predicate or operator contained in the latter; and in that case, it might perhaps be processed as an independent sentence.

But in general, sentence splitting in the stricter sense has the effect of making the target text more \textit{incremental} than the source text; the information conveyed by the text in its totality – the discourse information \cite{Fabricius-Hansen1996} – is portioned out in smaller chunks in accordance with principles of discourse organization that are especially well suited to the conditions of spoken language.\footnote{Cf. for instance Biber (1988).} These principles can be stated informally and somewhat sloppily as follows:
Principles of Incremental Discourse Organization (PIDO)

1. Minimize the number of new discourse referents introduced per sentence.
2. Minimize the number of conditions (the amount of information) stated for new or reactivated discourse referents in the sentences used to declare or reactivate them in the discourse.
3. Minimize the need for accommodation of discourse referents or conditions on referents, i.e., the use of presupposition triggering expressions whose presuppositions are not justified, explicitly or by inference, by the preceding context.

Viewed in isolation as stated above, these principles are, of course, too strong. “Keep low” might for instance be a more adequate wording than “minimize” in PIDO 1 and 2. But in any case, PIDO has to compete with other principles governing language production which I cannot go into here (cf. Levelt 1989). And given that information is encoded in lexical items and syntactic structure, the effect of PIDO will be constrained by the lexicon and the syntactic rules of the language, in addition to more general principles of minimal informativeness or relevance. Using, for instance, the finite verb form discovered to declare a new event referent, it may well be necessary to specify the temporal location of the event within the same sentence; and the sentence will have to contain constituents referring to the participants of the event. But then again, Principle 1 would demand that at least one of the participants should be identifiable with a discourse referent already introduced in the discourse universe, that is, it should be an anaphoric pronoun or a definite noun phrase with an identifiable antecedent (PIDO 3); and according to PIDO 2, if, say, the object is a noun phrase introducing a new discourse referent, it should not contain a series of adjuncts that are not needed in order to distinguish the referent in question from other referents already present in the universe or to be introduced in subsequent sentences; likewise, the description of the new event referent itself should not be unnecessarily specified by way of additional optional adverbials or the like.

2.4 PIDO at work

Let us look at another example – (9) – from Konrad Lorenz for illustration.

Here the German source sentence violates PIDO 1 by introducing two new event referents within one sentence, viz. the emergence of a monster heading towards the narrator (main clause event, bold face) and its subsequent dissolving into small fish (relative clause event); and it violates PIDO 2 both by a heavy premodifying adjunct (italics) giving a detailed additional description of the monster-referent (capitals) in the same sentence that declares that referent in the discourse universe, and by specifying in a participal adjunct (dotted) within the event-introducing clause itself the effect of the main-clause event, viz. that it was startling.
Both translations conform better to PIDO 1 in that they render the relative clause as an independent sentence, replacing the relative pronoun (DAS) by an anaphoric expression (DETTE SKREMSLET ‘this scary creature’ in KoLo-TN and IT in KoLo-TE) that takes the monster referent already present in the universe (cf. ET skrekkinnjagende UHYRE ‘a frightening monster’ in <s77> and A MONSTER in <s50><b>) as its antecedent.

The two target texts, however, cope in different ways with the violations of PIDO 2: KoLo-TN postpones the detailed description (italics) of the monster to a separate sentence (<s77>) whose subject is an anaphor (DET ‘it’) for the said monster, and gets rid of the event adjunct in the first sentence by ascribing the property of being frightening to the monster itself rather than to its appearance on the scene (which doesn’t make much of a difference). KoLo-TE, on the other hand, uses the first sentence (<s50><a>) to introduce the event referent as such (‘a … thing happened’, bold face), determining its temporal location by way of the subordinate temporal clause and its effect by way of the participal adjunct startling, but leaving the type of the event open, to be specified (bold face) in the following sentence <b> (see section 4.3). That is, KoLo-TE postpones not only the detailed description of the monster but also its introduction in the discourse universe along with the full specification of the main event. Thus, the English translation conforms better to PIDO 1, distributing the introduction of the main event referent and its participant over two sentences whereas in KoLo-TN, both are established in the first sentence, the second sentence (lit.: ‘This monster was several meters wide, almost equally high …’) having solely the function of further characterizing the monster introduced in the first. On the whole,
then, the sentence splitting strategy chosen in KoLo-TE may be said to ensure a more balanced informational load on the sentences resulting from the split than is the case for KoLo-TN.

The Norwegian translations in (7) and (8), repeated below, both illustrate information splitting that remedies violations of PIDO 3, using independent sentences to introduce explicitly discourse referents which are in the source sentences ‘taken for given’, i.e., referred to by definite descriptions without proper antecedents in the preceding context (cf. Solfjeld 1998:111ff), viz. the housewife Gerda W. and the fact that she had a midlife crisis, for (7), and the fact that cars have a strong symbolic power, for (8). In addition, (7b) conforms better to PIDO 2 than (7a) in that the description of Gerda W.’s intentions during her midlife crisis is separated from the introduction of the midlife crisis itself.


“When the 41 years old housewife Gerda W. was going (lit.: went) through a midlife crisis, wanting (lit.: wanted) to start a new life and go to Africa as a developmental worker, she had a typical car dream.”

b Gerda W., husmor og 41 år gammel, gjennomgikk en “middelalderskrise”. Hun ville begynne et nytt liv og reise til Afrika for å drive u-hjelp.

Da hadde hun en typisk bildrøm. (Norw. transl., ib.)

“Gerda W., housewife and 41 years old, was going (lit.: went) through a midlife crisis. She wanted to start a new life and go to Africa to do developmental work. Then she had a typical car dream.”

(8) a Durch die außergewöhnlich starke symbolische Ausdruckskraft des Autos benützt der Traum das Auto, um eine bestimmte Lebenssituation aufzuzeigen. (ib.)

“Due to (lit.: through) the exceptionally strong symbolic power of cars (lit.: the car), the dream uses a car (lit.: the car) to picture a specific situation in life.”

b Bilen har en uvanlig sterk symbolsk uttrykkskraft.

Derfor benytter drømmen seg av bilen for å skisse et bestemt livssituasjon. (Norw. transl., ib.)

“Cars (lit: the car) have an exceptionally strong symbolic power. Therefore the dream uses a car (the car) to picture a specific situation in life.”
2.5 Incrementality and global discourse structure

Not surprisingly, the Principles of Incremental Discourse Organization as stated above have parallels in the literature on discourse structure and discourse processing. Thus, PIDO 2 is comparable to the Principle of the Separation of Reference and Role stated by Lambrecht (1995:184) to account for so-called detachment constructions, viz. that the lexical representation of a topic referent takes place separately from the designation of the referent’s role in a proposition, which Lambrecht takes to be based on the following pragmatic principle: “Do not introduce a referent and talk about it in the same clause.” Lambrecht, however, is concerned with the organization of information on sentence or clause level whereas the Principles of Incremental Discourse Organization stated above have to with discourse information.

Taking the clause to be “the incremental unit of processing new textual information”, Givón (1984:358) formulates a “One-chunk-per-clause constraint” to the effect that “clauses in natural texts tend to have only one chunk (usually a word) of new information per clause”. Apparently, ‘clause’ in this context may be paraphrased as ‘one-clause sentence’, and in that case, Givón’s constraint can be viewed as a very radical variant of PIDO 1 and 2 above.

Finally, van Dijk/Kintsch (1983), outlining fundamental strategies of discourse production, describe a substrategy that corresponds to PIDO 3:

“In the description of whole facts, first specify some minimal description of discourse individuals and then the additional properties or relations of these individuals.

Follow the presuppositional structure of the discourse, that is, always first specify the facts (individuals and their properties and relations) presupposed by later facts described in the discourse.” (van Dijk/Kintsch 1983:275f.)

According to van Dijk/Kintsch (ib.), deviations from the ‘natural’ incremental order of saying things may occur for various reasons such as the need to obtain specific illocutionary or rhetorical effects (“pragmatically” and “rhetorically” motivated deviations). In addition, so-called factive verbs often do not follow the principle stated above: They tend to establish the presupposed fact, by way of a subordinate complement clause, along with the relation it enters into, rather than separating the two as incrementality would demand (a “cognitively” motivated deviation, according to van Dijk/Kintsch). On the whole, it is an interesting empirical question under which conditions deviations form PIDO 3 do occur regularly and to what extent they are felt to be (un)acceptable to readers/listeners.

In principle, of course, the same discourse information allows for different linearizations with more or less the same degree of incrementality. But the options will be severely restricted by other, more global principles of discourse organization. Thus, Klein/von Stutterheim (1991) point out that linearization or what they call the referential movement of a text or a text section depends very much on what is its ‘quaestio’, that is the question it sets out to answer; and depending on its communicative function, illocutionary strategies
may play an important role too, as suggested e.g., by van Dijk/Kintsch. These are factors that I cannot go into here. But following Klein/von Stutterheim (1991) and von Stutterheim (1997:15ff), I shall, in principle, distinguish between the ‘main structure’ which constitutes the answer to the text quaestio and other parts of the text (‘side structures’) which fulfill some other functions. And discourses that follow PIDO very strictly, can be expected to shift continuously between sentences that primarily have the function of introducing discourse referents which are to play a role in the development of the main structure, sentences that directly develop the main structure, and sentences that add information on main structure referents but do not contribute to the main structure itself – in addition, perhaps, to sentences that exclusively have to do with ‘side structure’ elements.

2.6 Summary

Let us take stock: We have seen that translating a non-clausal (phrasal) constituent by a subordinate clause is accompanied by explicitation in a broad sense without necessarily reducing syntactic complexity very much (depending on how one analyzes complex sentences). To the extent that the expansion only expresses what is already ‘there’ in the DRS assigned to the source text, this type of sententialization may be said to reduce informational density by using more linguistic material, more elaborate constructions, to convey the same information, i.e., by increasing local redundancy. But as often as not the increased explicitness demanded by a clausal as opposed to a phrasal construction will force the translator to choose between two or more equally possible – albeit perhaps not equally plausible – interpretations of the source sentence, thus, in fact, adding information by reducing underspecification or ambiguity.

Sententialization in the stricter sense, however, i.e., translating a source sentence by a sequence of independent target sentences, aims at reducing informational density of the target text as opposed to the source text by increasing incrementality: Other things being equal, the discourse information that is conveyed by \( n \) sentences in the source text will in the target text be distributed over \( m \) sentences, with \( m > n \), with the result that, on average, the sentences of the target text contain less discourse information than the sentences of the source text.

Now, what corresponds to a ‘new’ independent target sentence may be a phrasal constituent of the source sentence translated by a sentence sequence, thus combining sentence splitting with clausal expansion, as can be seen in (8). Then the target text, drawing on clause making strategies like those discussed above, will not only be more incremental but also more explicit than the source text.

---

18 See also Brandt/Koch/Rosengren/Viehweger (1982).

19 We may distinguish between the local information and the discourse information (new information) conveyed by a sentence in a given context:

Let \( S_{i+1} \) be a sentence and \( K_i \) the DRS construed for the text (= \( S_1 \ldots S_i \)) preceding \( S_{i+1} \). Then the local information conveyed by \( S_{i+1} \) relative to \( K_i \) is the semantic representation construed for \( S_{i+1} \) during syntactic parsing according to the rules of construal (Kamp/Reyle 1993); and the discourse information conveyed by \( S_{i+1} \) relative to \( K_i \) is the change \( S_{i+1} \) induces in the DRS for the discourse, i.e., what is added when \( K_i \) is updated by \( S_{i+1} \) to yield \( K_{i+1} \) (i.e., \( K_{i+1} = K_i + S_{i+1} \)).
But the source constituent that is sententialized in the target text may itself be a full clause: Sentence splitting may, for instance, be obtained simply by ‘converting’ a subordinate clause of a source sentence into an independent target sentence, as can be seen in (7). In that case, increased incrementality is not accompanied by increased explicitness – rather on the contrary, as we shall see below (section 5.2). And depending on the type or function of source clauses turned into independent sentences, it might even be discussed whether there is a significant change of incrementality at all (cf. the discussion in section 2.2). In any case, however, sentence splitting, by ‘removing’ or ‘extracting’ whole – phrasal or clausal – constituents from the sentence being split, obviously reduces syntactic complexity: Normally none of the sentences in the sequence corresponding to a single source sentence will reach the hierarchical depth of their ‘mother sentence’.

The translational relations discussed thus far are summarized in fig. 3.

<table>
<thead>
<tr>
<th>SOURCE TEXT</th>
<th>TARGET TEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-clausal constituent at VP or NP level</td>
<td>Subordinate clause</td>
</tr>
<tr>
<td>Subordinate clause</td>
<td>Independent sentence</td>
</tr>
</tbody>
</table>

(a) Clausal expansion without sentence splitting: Increased explicitness (redundancy) or specificity at the propositional level

(b) Clausal expansion combined with sentence splitting (strong sententialization): Increased explicitness (redundancy) or specificity at the propositional level and increased incrementality

(c) Sentence splitting without clausal expansion (weak sententialization): Increased incrementality

Fig. 3 Clausal expansion vs. weak and strong sententialization

Clause coordination, verb (phrase) coordination and other types of coordinate structures involving finite verbs may also give rise to sentence splitting as shown by Solfjeld (1998:159ff, 203ff). The incrementality effect, in this case, may seem less perspicuous than when embedded phrases or clauses are converted into independent sentences (cf. 2.2), that is we have to do with a weak kind of sententialization; cf. (10).

(10) a Das Verhalten des anderen wird imitiert, weil dieser als Vorbild angesehen wird, oder weil das gezeigte Verhalten belohnt wird, oder weil gar keine andere Verhaltensmöglichkeit angeboten ist. (cit. Solfjeld 1998:159)

“The other person’s behaviour is imitated because he/she is seen as a role model or because the demonstrated behaviour is rewarded or because there are no other behaviour options.”
b En annens atferd blir imitert fordi vedkommende blir betraktet som et forbilde, eller fordi denne atferden blir belønnet. Det kan også være at det ikke finnes andre atferdsmuligheter. (ib.)

“Someone else’s behaviour is imitated because he/she is seen as a role model or because that behaviour is rewarded. It may also be that there are no other behaviour options.”

And conversely, in addition to clausal expansion as described above and sentence splitting proper, information splitting may take the form shown in (11)–(12) where a single finite clause or verb phrase of the source text has been expanded into explicitly coordinated verb phrases or clauses. This, too, could be called a milder variant of incrementality increase; information that is contained in one source clause is distributed over two clauses or clause-like constructions in the target text that constitute a single sentence but none of which is subordinated to the other.

(11) a In einer zwanzigminütigen Pause setzen wir uns auf die Eisentreppe, wo etwas weniger Staub ist. (Cit. Solfjeld 1998:129)

“In a 20 minutes break we sit down on the iron staircase where there is somewhat less dust.”

b Vi har en pause på ca. 20 minutter og da setter vi oss på en jerntrapp, der det er mindre støv. (ib.)

“We have a 20 minutes break and so/in that break we sit down …”

(12) a Sie fühlen sich wohl in ihrem Element. (Zit. Solfjeld 1998:155)

“They feel well in their element.”

b … de er i sitt eget element og føler seg vel derved. (ib.)

“They are in their own element and feel well by that.”

In what follows, however, I shall concentrate on sentence splitting that neither arises from nor results in explicitly coordinated structures.

3. Sentence splitting, sentence collapsing and the Principal Counterpart

It follows from what was said above that a target text T will be more incrementally structured than the source text *T it translates iff at least one sentence *Si in *T is translated by a sequence of sentences S1/i … Sn/i in T and the opposite – sentence collapsing, i.e., a sequence of source text sentences corresponding to a single target sentence – does not hold or holds to a lesser degree. That is, T is more incremental than *T iff translational correspondences of type (a) in fig. 4 outnumbers those of type (b).
(13) illustrates sentence collapsing between the German source text and the English translation whereas the Norwegian translation – as usual (Fabricius-Hansen 1998) – is more incremental than its source text.
Of course, sentence splitting and sentence collapsing may be mixed, resulting in correspondencies of type (c) as shown for the German-English pair in (14). But in general, sentence boundaries in the target text T correspond to sentence boundaries in the source text *T.

Normally, when a translational relation holds between one sentence on the one hand and a sequence of sentences on the other hand, the sentence sequence will contain a designated member which I shall call the Principal Counterpart of the corresponding single sentence: the sentence that corresponds most closely to the main clause predicate of the latter in that it introduces into the DRS of the sentence sequence the abstract discourse referent (eventuality, proposition, …) which that predicate declares in the DRS of the single sentence. As far as sentence splitting is concerned, then, we can define this notion as follows:

**The Principal Counterpart of *S**

Given a source text *T and a target text T: For any sequence of sentences S1/i … Sn/i (i, n ≥ 1) in T translating a single sentence *Si (i ≥ 1) in *T, the Principal Counterpart of *Si in T, symbolized S*/i, is the sentence Sj/i (j ≥ 1) in T that introduces into T the abstract discourse referent which in *T is established by the main (clause) predicate of *S, embedding it into a context corresponding to the one it has in *T.

---

20 Allowing the possibility n = 1, the definition above subsumes the trivial (non-splitting) cases where a source sentence is translated by a single target sentence.
Dette fenomen er virksomt overalt, og kalles det naturlige utvalg eller den naturlige seleksjon.

Det naturlige utvalg er den ene av de to store konstruktører som står bak artenes omdannelse.

Den andre, som leverer materialet for det naturlige utvalg, er DE ENDRINGER I ARVEANLEGGENE SOM VI KALLER MUTASJONER.

Med genial forutseennhet postulerte Darwin at MUTASJONENE var en nødvendighet – på en tid da DERES eksistens ennå ikke var påvist.

Dieses allgegenwärtige Geschehen nennt man natürliche Zuchtwahl oder Selektion.

Die Selektion ist der eine von den beiden großen Konstrukteuren des Artenwandels;

……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………
In example (9) above (p.186), on the other hand, it is the first sentence of the target sentence sequence – i.e., <s77> and <s50a>, respectively – which functions as the Principal Counterpart of the German source sentence <s54>, introducing the temporally located event referent that the source text describes as the (startling) emergence and approaching of a monster. This is so not only in the Norwegian but also in the English translation; in the latter, however, the full description of that referent is distributed over two sentences (cf. section 2.4).

<table>
<thead>
<tr>
<th>(15)</th>
<th>(15)</th>
<th>(15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;s172&gt; Forskjellige pattedyr-kjennere har reist [NP DEN INNVENDING at slik duftmar-kering ikke kan ha noe å gjøre med territorialhevdelse, fordi den forekommer hos sosialt levende pattedyr som ikke forsvarer noe enkeltterritorium, såvel som hos arter som streifer vidt omkring].</td>
<td>&lt;s137&gt; [NP DER von verschiedenen Säugetierkunden digen erhobene EINWAND, daß derlei Geruchsmarken mit Revierbesitz nichts zu tun hätten, da sie sowohl bei sozial lebenden, keine Einzelreviere verteidigenden Säugern vorkommen als auch bei solchen, die weit umherziegen n].</td>
<td>&lt;s123&gt; &lt;a&gt; THE OBJECTION has been raised by some students of mammals that such scent marks cannot have anything to do with territorial ownership because they are found not only in socially living mammals which do not defend single territories, but also in animals that wander far and wide; &lt;b&gt; but THIS OPINION is only partly correct.</td>
</tr>
</tbody>
</table>

4. Directions of sentence splitting

4.1 Sentence splitting as information extraction

The fact that a sequence of sentences translating a single source sentence normally contains a designated member or kind of nucleus which the other sentences either precede or follow makes it possible to distinguish between backward and forward sentence splitting depending on the position of the Principal Counterpart relative to the other sentences in the sequence, as illustrated, for example, by (7)–(8) and (9) (pp. 185f), respectively.
**Backward sentence splitting**
A sequence of sentences $S_1/i \ldots S_n/i$ ($1 < n$) that translates (part of) a source sentence $S_i$ is related to $S_i$ by backward sentence splitting iff the last sentence in the sequence is the Principal Counterpart of $S_i$, i.e., iff $S*/i = S_n/i$.

**Forward sentence splitting**
A sequence of sentences $S_1/i \ldots S_n/i$ ($1 < n$) that translates (part of) a source sentence $S_i$ is related to $S_i$ by forward sentence splitting iff the first sentence in the sequence is the Principal Counterpart of $S_i$, i.e., iff $S*/i = S_1/i$.

Given the Principal Counterpart, we may now view sentence splitting as a kind of information extraction: Under forward sentence splitting, information is extracted to the right of what might have been a single sentence corresponding to the source sentence; and under backward sentence splitting, information is extracted to the left of its ‘mother sentence’, i.e., $S*$; cf. fig. 5(a) and (b).

![Diagram of information extraction](image)

**Centered information extraction**
A sequence of sentences $S_1/i \ldots S_n/i$ ($1 < n$) that translates (part of) a source sentence $S_i$ is related to $S_i$ by centered sentence splitting which one might call centered information extraction; cf. fig. 5(c).

Fig. 5 Directions of information extraction
The Principal Counterpart need not be identical to either the first or the last sentence in a sequence corresponding to a single source sentence; it may be that $S*/i = S_j/i$ with $1 < j < n$, as seen in the Norwegian target text of (16) below. In that case, the target sequence is related to the source sentence by a combination of backward and forward sentence splitting which one might call centered information extraction; cf. fig. 5(c).
Because of the enormous nutritional possibilities, every fish, whatever its speciality, requires only a few square yards of sea bottom for its support.

In the following sections, we shall have a closer look on backward and forward sentence splitting, their triggering conditions, and their effect on the discourse structure of the target text.

4.2 Backward information extraction

Backward sentence splitting primarily seems to be triggered by complex, presupposition carrying constituents in the source sentence *Si, in particular if the presupposition in question is not justified by the preceding context: Extracting presuppositions to the left of the sentence they are embedded into, thereby converting them into explicit assertions, is a natural way of reducing informational density caused by violations of PIDO 3 (section 2.3).

Typical triggering constructions, then, are definite descriptions containing pre- and/or postmodifying adjuncts on the one hand and fronted subordinate clauses on the other hand. In either case, the information – discourse referents together with their defining conditions – that is extracted from S*ni to be distributed over “new” sentences to the left of S*ni must...
be represented in S*n/i by some explicit or implicit anaphor to ensure equivalence between the target sentence sequence S1/i ... S*n/i and the source sentence *Si.

Definite descriptions giving rise to backward sentence splitting are seen e.g., in (15) and (8). Thus, in (15) the noun phrase DER von verschiedenen Säugetierkundigen erhobene EINWAND daß ... ‘the objection raised by some experts on mammals to the effect that …’ triggers the presupposition that such an objection has been raised – a presupposition that does not seem to be satisfied in the preceding context; and in both translations, the first sentence, being based on that NP, fulfils exactly the task of explicitly introducing into the discourse the event described as experts raising a certain objection along with the content of the objection, leaving a reduced definite description DENNE INNVENDING ‘this objection’ and THIS OPINION in the second, nuclear sentence as an anaphor for the discourse referent established in the preceding sentence. As for (8), the definite noun phrase (durch) die außergewöhnlich starke Ausdrucks Kraft des Autos ‘(through/because of) the exceptionally strong symbolic power of the car’ containing a postmodifying possessive genitive phrase gives rise to an independent target sentence with the possession verb Norw. ha ‘have’ – Bilen har en uvanlig sterk symbolsk uttrykskraft ‘the car has an exceptionally strong symbolic power’ – which introduces the fact that cars are powerful symbols into the discourse; in the following nuclear sentence, i.e., the Principal Counterpart of the single source sentence <s137>, this new discourse referent is ‘picked up’ anaphorically by the connective therefore which places it in a causal context roughly corresponding to the relation described by the preposition durch ‘through’ in the source sentence. In all such cases, of course, we have to do with strong sententialization: sentence splitting combined with clausal expansion (cf. section 2.6).

The Norwegian translations in (7) and (16) illustrate the other main variant of backward sentence splitting: independent target sentences arising (by weak sententialization) from fronted subordinate clauses having a more or less obvious presuppositional status; and again, an anaphor of a suitable type has taken the place of the extracted constituent in what is left as the Principal Counterpart (S*) of the source sentence. In (7), a temporal clause is converted into a sequence of two independent sentences and the corresponding anaphor is the Norw. temporal adverb da ‘then, at that time’ which takes as its antecedent the time interval occupied by the event described in the first of these two sentences. In (16), it is a causal clause with the German subjunctor da ‘since’ that is translated by an independent sentence in the Norwegian target text, and the abstract discourse referent it establishes is ‘picked up’ in the following nuclear sentence by the definite description resultatet (er at ...) ‘the result (is that …)’ which has to be interpreted as ‘the result of this (is that …)’. The English translation, too, renders the fronted subordinate clause by a main clause, but uses comma instead of full stop for a division sign and reduces the main clause of the source text so ergibt sich (daß ...) ‘so (it) follows (that ...)’ to the causal-consecutive connective so (…).

The fronted causal clause in the source text of (16) is not only in itself presuppositional, it also contains an embedded definite noun phrase as part of an adverbial (causal-circumstantial”) prepositional phrase bei dem ungeheuren Nahrungsangebot des Korallenriffs ‘with/in view of/because of the enormous nutritional possibilities of/offered by the coral reef’. This definite description is left as such in the English target

---

text, perhaps because the (familiarity) presupposition it gives rise to is in fact justified by the preceding text: At this point, the fact that coral reefs offer enormous nutritional possibilities is certainly not new to the reader. But obviously, incrementality of the target text would be further increased by identifying or ‘(re)activating’ the discourse referent in question outside the sentence that establishes its consequences. As shown in Fabricius-Hansen (1998), the natural way to do this is by backward information extraction applied to the noun phrase in question which, consequently, turns up as an independent sentence to the left of the target sentence that has already been split off. This is demonstrated (16’a); the example is also meant to suggest that in case a contextually justified presupposition is sententialized, a metatextual comment like as we have seen is needed as a reminder that the information conveyed by the independent sentence is not contextually new but should be identified with information already present in the context, i.e., in the DRS construed that far.

(16’)

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The coral reef offers enormous nutritional possibilities (as we have seen)</td>
<td>Every fish, whatever its speciality, requires only a few square yards of sea bottom for its support.</td>
<td>The coral reef offers enormous nutritional possibilities (as we have seen).</td>
<td>Because of the enormous nutritional possibilities, every fish, whatever its speciality, requires only a few square yards of sea bottom for its support.</td>
</tr>
<tr>
<td>Therefore, every fish, whatever its speciality, requires only a few square yards of sea bottom for its support.</td>
<td>For the coral reef offers enormous nutritional possibilities (as we have seen).</td>
<td>Because every fish (therefore/or for that reason), whatever its speciality, requires only a few square yards of sea bottom for its support,</td>
<td>……………………………</td>
</tr>
<tr>
<td>So in this small area there can be as many fish as there are ecological niches, …</td>
<td>So in this small area there can be as many fish as there are ecological niches, …</td>
<td>There can, in this small area, be as many fish as there are ecological niches, …</td>
<td>So in this small area there can be as many fish as there are ecological niches, …</td>
</tr>
</tbody>
</table>

The procedure that leads from the complex source sentence of (16) to the three-sentence target text in (16’a) can be described as recursive top-down backward presupposition extraction; it has the effect that, in the end, the ‘new’ sentences arising from extraction will be ordered in a way that mirrors the embedding hierarchy of the presupposition-triggering constructions, as illustrated in fig. 8 below (section 5.2): For every pair Sj, Sj+1 (with 1 ≤ j < n) in the target sequence S1 ... Sn, Sj ‘derives’ from (a source constituent corresponding to) Sj+1; thus, Sj+1 can be viewed as nuclear relative to Sj; and S1 will correspond to the presupposition most deeply embedded in the source sentence. This (re)ordering, of course, reflects the nature of presuppositions, and it is the one that has the
greatest incrementality effect (according to PIDO 3) as far as the extraction of unjustified presuppositions is concerned. Neither of the two other ‘logical’ alternatives demonstrated in (16’b–c) – sententializing the most deeply embedded presupposition without sententializing its mother construction as in (c) or extracting both but in the opposite order as in (b) – seems quite adequate; the (b)-version, in particular, is really bad, disturbing the anaphoric relation between so in the nucleus and the fact that every fish requires only a few square yards of sea bottom for its support. (For comparison, the authentic version presented in (16) is added as (d) in (16’)). – (11) and (12) above (p.190) illustrate backward information extraction at verb phrase level resulting in clause or verb phrase coordination.

In the examples discussed above, the presuppositions to be extracted occur in factual or extensional contexts; that is, they ‘float up’ to the main DRS, to the top level of the DRS eventually assigned to the source sentence; otherwise, they could not have been realized in translation as unmodalized declarative sentences. But presuppositions may also occur in intensional contexts that prevent them from reaching the main DRS and, by consequence, prevent presupposition extraction without further modifications (modalization); it would, however, take us too far to go into these matters here.23 It also remains to be seen how backward extraction works when one source constituent triggers two or more mutually independent presuppositions; obviously, the generalization made above concerning the order of extracted presuppositions needs some refinement in order to account for such cases; but this, too, is a topic I cannot expand on. Finally, it should be stressed that backward information extraction does not only apply to obvious presupposition triggers like definite noun phrases; but there is reason to believe that it is restricted to material that belongs to the topic domain rather than the focus domain – as understood e.g., by Klein/von Stutterheim (1991), von Stutterheim (1997:33ff) – of the source sentence.

4.3 Forward information extraction

Although forward sentence splitting may be applied to violations of PIDO 3, as an alternative to the strategy described above (Fabricius-Hansen 1998:208f.), its natural domain would be expected to be structures that violate PIDO 1 or 2, defining too many new discourse referents or stating unnecessarily many conditions on new or familiar discourse referents in one processing step. In practice, then, forward information extraction should tend to be triggered by adjuncts at different levels of the sentence structure: extensional adverbials at sentence or verb phrase level and pre- and postmodifying adjuncts at noun phrase level which contain information that is not necessary in order to define or identify the discourse referent in question, including appositions and non-topical subordinate clauses. Typical examples can be seen in (9) which was discussed in section 2.4. and which, in particular, illustrates two different forward splitting strategies: (i) head-anaphoric and (ii) head-specifying forward information extraction24.

24 Fabricius-Hansen (1998) uses the terms (i) forward adjunct distribution via head anaphorization and (ii) head specification (or head splitting).
(i) Under head-anaphoric forward sentence splitting conditions (information) encoded in free adjuncts within a verb phrase or an (indefinite) noun phrase are extracted from (the Principal Counterpart) of the source sentence and distributed over one or more sentences to the right of S*, each containing an anaphor of some sort for the discourse referent introduced by the relevant phrase in S*. This is the strategy applied by the Norwegian translation of the subject noun phrase EIN ... UNGEHEUER, DAS ... of the German main clause in (9): The premodifying participle phrase and the postmodifying relative clause have both been turned into independent sentences containing an anaphor (definite description, demonstrative pronoun) for the discourse referent introduced by the relevant phrase in S*. (13) and (14) also show examples of non-restrictive relative clauses translated by independent clauses (in both target texts). – (17a, b) and (17a, c), on the other hand, are (non-authentic) examples of this type of forward information extraction applied at verb phrase rather than noun phrase level (Fabricius-Hansen 1998).

(17) a Meine Kollegen *spielen*, um fit zu bleiben, zweimal in der Woche Tennis.
   “My colleagues play in order to keep fit twice in the week tennis.”

b My colleagues play tennis twice in the week.
   They do it/so in order to keep fit.

c My colleagues play tennis in order to keep fit.
   They play twice in the week.

In such a case, it is not only the eventuality referent itself (play tennis) that is represented by an anaphoric element (do so/it, play), ellipsis included; the same will happen to discourse referents involved in that eventuality as a side effect of event anaphorization, as illustrated by the referential ‘chain’ my colleagues – they vs. meine Kollegen in (17).

(ii) The sentence splitting strategy I have termed head specification and which might also be termed head splitting distributes criterial conditions on the discourse referent declared by a verb or noun phrase, i.e., information encoded in the nucleus (including complements) of that phrase over two sentences in such a way that the second sentence specifies information conveyed by the relevant constituent in the first sentence S*. This is what happens in the English version of (9) where the central event is first defined in an underspecified manner as a startling thing happening, to be specified in the following sentence as the emergence of a monster from the darkness etc. In cases like this, the second sentence does not contain an anaphoric expression which triggers the unification with the relevant discourse referent – its antecedent – by simple anaphoric resolution. The discourse referent defined in S2 by the more specific description – in the present case: the event characterized as the emergence of a monster etc. – has no antecedent in the proper sense of that word; its unification with a less specified discourse referent declared in the sentence before – in casu: an event that has only been defined with respect to its temporal (and physical) location and its effect on the narrator – is based on lexical meaning, discourse knowledge and general pragmatic reasoning – and on punctuation: A colon typically marks a relation of elaboration...

25 See Fabricius-Hansen (1998:220ff) for further details, including examples of forward extraction of verb phrase adjuncts.
or specification between the sentences it divides to the effect that the eventuality described in S2 should be subsumed under or is included in the eventuality described in the preceding sentence.  

In the English version of (9), where sentence splitting is applied to a verb phrase, S2 specifies the type of an event(uality) established in S1 (= S*, the Principal Counterpart) without corresponding anaphoric links between the two. When applied to object-denoting noun phrases, the head-specifying variant of forward information extraction typically results in sentence pairs like (18a, c), to be compared with (18a, b) which – like the Norwegian version of (9) – is a normal case of adjunct head-anaphoric forward information extraction (Fabricius-Hansen 1998).

(18)  

a Neulich hat JJ ein äußerst interessantes, aber schwer zugängliches Buch über Fokus im Deutschen publiziert.  

“JJ recently published a very interesting but rather demanding book on focus in German.”

b JJ har nettopp publisert en bok om fokus i tysk.  

“JJ recently published a book on focus in German.”

Boka/Den er svært interessant, men vanskelig tilgjengelig.  

“The book/It is very interesting but rather demanding.”

c JJ har nettopp publisert et arbeid som er svært interessant, men vanskelig tilgjengelig.  

“JJ recently published a work that is very interesting but rather demanding.”

Det er en bok om fokus i tysk.  

“It is a book on focus in German.”

In (18c), the indefinite object noun phrase of the first sentence (the Principal Counterpart of the source sentence) like (18b) declares a new discourse referent as a participant (result) of the publication event described by the main verb, ascribing to it the property of being very interesting but rather demanding. The second sentence opens with the personal pronoun det ‘it’. But in contradistinction to the corresponding non-neuter pronoun den as an alternative to the definite NP boka ‘the book’ in (18b), det “it” in the present construction (“it is” + NP) does not function as a normal anaphorical pronoun ‘referring to’ an identifiable noun phrase in the preceding context (in casu: et arbeid ‘a piece-of-work’). This can be seen from the fact that det, although neuter singular, does not in this construction put any restrictions on its putative antecedent as far as animacy, number and grammatical gender is concerned: et arbeid and det only accidentally correspond in gender and number; cf. (19).

---

Cf. Asher (1993:267ff), Behrens (1998). Asher (1993) uses the term ‘elaboration’ rather than ‘specification’ but in general, the literature on discourse relations does not make any clearcut distinctions here. One might, however, differentiate along the lines suggested above, using ‘specification’ when S2 specifies the type of the referent in question without a proper anaphoric link – that is, when the sentence sequence could have come about by head splitting, as shown above – and ‘elaboration’ in case S2 adds information on a discourse referent that is represented anaphorically in S2 itself, as shown under (i) above.

---

26 Cf. Asher (1993:267ff), Behrens (1998). Asher (1993) uses the term ‘elaboration’ rather than ‘specification’ but in general, the literature on discourse relations does not make any clearcut distinctions here. One might, however, differentiate along the lines suggested above, using ‘specification’ when S2 specifies the type of the referent in question without a proper anaphoric link – that is, when the sentence sequence could have come about by head splitting, as shown above – and ‘elaboration’ in case S2 adds information on a discourse referent that is represented anaphorically in S2 itself, as shown under (i) above.
(19)  a Jeg har kjøpt en ny computer. Det er en Macintosh.
    “I have bought a new computer-masc. It is a Macintosh.”
    
    b Jeg hørte noen barn leke i gården. Det var naboenes døtre.
    “I heard some children playing in the back yard. It was the neighbor’s daughters.”

In fact, the whole construction rather seems to be interpreted as an elliptic cleft, i.e., (18c) and (19a, b) are understood as (18c’) and (19a’, b’), respectively, and det ‘it’ might consequently be classified as a kind of topic anaphor. 27

(18)  c’ Det er en bok om fokus i tysk JJ netopp har publisert.
    “It is a book on focus in German JJ recently published.”

(19)  a’ Det er en Macintosh jeg har kjøpt.
    “It is a Macintosh I have bought.”

    b’ Det var naboenes døtre jeg hørte leke (i gården).
    “It was the neighbor’s daughters I heard playing in the back yard.”

In other words, it is as a participant (result) of the publishing-event described in the preceding sentence rather than as an entity of the type defined by the head (arbeid) of the object NP that the new discourse referent introduced with that NP – if at all – can be said to serve as an antecedent of the anaphor det in (18c). It is to a discourse referent thus identified that (18c) ascribes the property of being a book on focus in German, thereby specifying the classification that has already been made, viz. that the thing published by JJ recently is et arbeid – a ‘piece-of-work’. An alternative way of applying head splitting to an object-denoting noun phrase can be to repeat or copy the main clause predicate of S* combined with the more specific NP, as shown in (18d); the latter, then, will carry the sentence focus, the rest being topical.

(18)  d JJ har netopp publisert et arbeid som er svært interessant, men vanskelig tilgjengelig.
    “JJ recently published a piece-of-work that is very interesting but rather demanding.”
    
    Han har har publisert en bok om fokus i tysk.
    “He has published a book on focus in German.”

The head-anaphoric and the head-specifying strategies illustrated above may also be combined in non-standard ways that I cannot go into here (see Fabricius-Hansen 1998).

27 Instead of the copula construction det er ... one could have used the construction det dreier seg om ... in Norwegian, corresponding to the German construction es handelt sich um ... which cannot be translated analogously into English. Traditionally, det/es is called a formal subject in this construction, but obviously it is anaphoric at some level; the term ‘topic anaphor’ might, in fact, be even more appropriate in this case.
5. Causes and effects of sentence splitting

5.1 Why sentence splitting?

So far I have, at least implicitly, suggested that sentence splitting in translations from German into Norwegian or English is triggered by violations of one or the other Principle of Incremental Discourse Organization, as viewed from the perspective of the target language. That would imply that the source and the target language communities differ with respect to the incrementality standards they set for written texts of the types where information splitting can be observed, i.e., that they do not follow the same stylistic norms. This may very well be true, at least to some degree, and in particular as far as German vs. Norwegian is concerned (Solfjeld 1998). But it is not the whole story. German, in fact, has some structural peculiarities that allows or even favors hierarchical information packaging to a larger extent than is feasible in English and, especially, Norwegian. 28

Being inherently an (S)OV language, German, for once, permits a whole series of phrasal adverbials to occur in the so-called middle field, i.e., between the complementizer/finite V2 and the final V-position; 29 but (clausal) adverbials also occur to the right of the final V, in so-called extraposition. In English and Norwegian, on the other hand, adverbials mostly have to follow the verb and in case there are several adverbials, that may cause structural (attachment) ambiguity. This means that it is in a way easier or more natural for German to violate PIDO 2 by packing additional non-criterial information on an eventuality referent defined by the head verb into adjuncts within the same sentence or verb phrase. (9) is a good example of that: The German main clause contains two adverbial adjuncts erschreckend 'startling' and aus dem Dunkel unter dem Steg hervor 'from under the darkness of the stage' in addition to the main clause predicate nucleus kam ... auf mich zugeschossen 'came ... shooting against me’ and it is hard to imagine how both could be retained as such in a one-clause translation even without the complicating modification of the subject noun phrase.

At noun phrase level there are related but still more conspicuous differences between German on the one hand and English and Norwegian on the other hand as far as possibilities of information packaging is concerned: Neither English nor Norwegian allow extended (participal or adjectival) prenuclear modification of the type shown (italics) in (9) whereas German has the same or even more 30 options as the two other languages when it comes to postnuclear modification (prepositional phrases, relative clauses etc.). So again, by simply exploiting the structural possibility of expanding the noun phrase to the left and to the right at the same time, one may pack more information into a German noun phrase than could be handled, in a reasonably balanced and unambiguous way, within an English

29 According to Doherty (1998), German even, for ease of processing, prefers phrasal realization of adverbials to adverbial clauses.
30 In German, noun phrase adjuncts in the genitive case are normally postnuclear and range over a wider functional spectre than its (prepositional) counterparts in English and Norwegian.
or Norwegian noun phrase. In the typical cases, heavy premodifying adjectival or participal phrases are rendered as postmodifying relative clauses or, especially in English, participal phrases (reduced relative clauses); but in that case, if the source NP contains a postnuclear adjunct – e.g., a relative clause – in addition, that will have to make room for the adjunct coming from the left and, as a result, it may by a kind of chain reaction be converted into an independent sentence as witnessed e.g., in (9) and in (20) (see below).

German also has a wider range of word formation and nominalization structures than English and Norwegian, in particular as far as so-called synthetic compounding is concerned;\(^{31}\) and that again facilitates patterns of hierarchical information packaging which trigger information splitting under translation.

When all these factors – recursive compounding, repeated nominalization, heavy prenuclear and postnuclear noun phrase modification, and accumulation of adverbial adjuncts – come together, they may present an almost insurmountable obstacle to translation although the text may not be equally difficult to read or understand. But altogether, English seems better equipped to cope with such translation difficulties, partly because participal phrases are allowed more unrestrictedly as free adverbial adjuncts and as postnuclear noun phrase modifiers, thus presenting a genuine translation alternative to finite clauses, and partly because English conventionally makes more refined use of punctuation, including colon and semicolon, as a means of bundling sequencies of syntactically independent sentences into larger informational units (Fabricius-Hansen 1998).

5.2 Sentence splitting and global discourse structure

In section 2.5, I pointed out that highly incremental text passages will be characterized by repeated shifts between at least three functionally different types of sentences: (i) sentences whose primary task it is to introduce discourse referents to be used in the development of the so-called main structure, (ii) sentences that directly develop the main structure (by answering the ‘quaestio’) and which might therefore be called focal at text level, and (iii) sentences that neither establish relevant discourse referents nor develop the main structure itself although they may add information on discourse referents involved in the main structure. Now, given that in complex sentences it is normally the main clause rather than some subordinate clause that (if at all) contributes to answering the quaestio of the text (section), it should be clear that sentence splitting, while making processing easier at sentence level, may have its cost when it comes to discourse processing at the more global level (Fabricius-Hansen 1996). Let us first try to consider the effects of backward and forward information extraction separately.

A target text \( T = S_1 \ldots S_n \) which is related to its source text \(*T = *S_1 \ldots *S_m \) (with \( n \geq m \geq 1 \)) by consistent backward information (presupposition) extraction will have the right-focal virtual structure shown in fig. 6(a): Every target sentence \( S_j \) that is the Principal Counterpart of some source sentence \(*S_i\) and thus at least potentially focal (bold face) may be preceded by a sequence of one or more sentences that are preparatory with respect to \( S_j \), establishing discourse referents which are taken for granted in \( S_j \) and thus justifying the presuppositional content of \( S_j \) – that is, sentences of type (i) above; and neighbouring non-

Focal sentences may even be linked in a similar way, as suggested in section 4.2; but focal sentences will not directly enter into discourse relations with non-focal sentences to their right although, of course, there may be cohesive ties between them. Reading a text (section) that opens with a right-focal sentence sequence, one will not know what it is really about, i.e., one will not be able to (re)construct its quaestio or understand its contribution to a given quaestio until the end of the sequence32 whereas with a corresponding single sentence, the syntactic structure itself will guide the discourse interpretation; cf. (16’a) and its German source in (16). And if two or more right-focal sentence sequences follow each other, each corresponding to one source sentence as shown in fig. 6(a), it may be quite difficult to establish discourse relations between the (focal members of the) subsequences corresponding to the relations holding between the source sentences.

![Diagram](image)

Fig. 6 Target text focality after sentence splitting

Consistent forward sentence splitting, on the other hand, will make the target text left-focal as compared to the source text, at least when it is done the normal way, i.e., by head anaphorization; in the case of head-specifying forward splitting, the focus may rather seem to spread over two sentences, as demonstrated in the English translation of (9). At any rate, the target text will have the virtual structure shown in fig. 6(b): Every target sentence S*j that is the primary counterpart of some source sentence may be followed by one or more sentences that either (a) elaborate on the description of the (main) eventuality referent introduced by S*j, thereby increasing the granularity of the main structure or (b) convey additional information on discourse referents from S*j which does not contribute to the development of the main structure, i.e., sentences of type (iii) above; S*j should not, however, be discourse-connected directly with a non-focal sentence to its left. And again, it may in the end be more difficult to establish the main structure of the target text than is the case with the source text. (20) – the last example to be discussed – illustrates this point quite well; and it also demonstrates the degree to which English and Norwegian translations may differ in the way they solve translation problems caused by typical German information packaging.

32 That is, continuous Topic Revision will be needed, according to Asher (1993:330ff).
<table>
<thead>
<tr>
<th>&lt;KoLo-TN.3.s123&gt;</th>
<th>&lt;KoLo.3.s100–102&gt;</th>
<th>&lt;KoLo-TE.3.s90&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>&lt;s123&gt;</strong> Her kan <strong>EN</strong> FISKEART, utviklingshistorisk talt, “velge mellom de mest forskjellige yrker”</td>
<td><strong>&lt;s199&gt;</strong> EINE FISCHART kann hier, stammesgeschichtlich gesprochen, “die verschiedensten Berufe ergreifen”.</td>
<td><strong>&lt;s90&gt;</strong> Here FISH SPECIES can, in an evolutionary sense, <strong>take up very different professions</strong>:</td>
</tr>
<tr>
<td><strong>&lt;s124&gt;</strong> FISKEN kann slå seg meget godt igjennom som “ufaglørt arbeider”, med det [som enhver gjenomsnittsfisk er i stand til å prester].</td>
<td><strong>&lt;s101&gt;</strong> DER FISCH kann sich als “ungelehrter Arbeiter” sehr wohl mit dem durchbringen, [was ein Durchschnittsfisch sowieso kann],</td>
<td><strong>&lt;s199&gt;</strong> in other words <strong>hunting</strong> CREATURES [<strong>THAT</strong> are neither poisonous, nor armour-plated nor prickly]</td>
</tr>
<tr>
<td><strong>&lt;s125&gt;</strong> DEN kann drive jakt på <strong>MINDRE</strong> ORGANISMER [SOM ikke er giftige, ikke har panser eller pigger og ellers ikke er kampdyktige på noen måte.]</td>
<td><strong>&lt;indem ER Jagd auf KLEINERE, (nicht giftige, nicht gepanzerte, nicht stachelige oder sonstwie wehrhafte) LEBEWESEN macht,</strong></td>
<td><strong>&lt;s199&gt;</strong> hunting <strong>CREATURES</strong> [<strong>THAT</strong> are neither poisonous, nor armour-plated nor prickly]</td>
</tr>
<tr>
<td><strong>&lt;s126&gt;</strong> SLIKE SMÅORGANISMER kommer inn til revet i store masser fra det åpne havet,</td>
<td><strong>&lt;s101&gt;</strong> als “Plankton” passiv von Wind und Wellen getrieben,</td>
<td><strong>&lt;s199&gt;</strong> others as <strong>active swimmers</strong> “intending” to settle on the reef,</td>
</tr>
<tr>
<td><strong>dels</strong> drejer det seg om “plankton” [som driver passivt av gårde med vind og strøm],</td>
<td><strong>teils</strong> als “Plankton” passiv von Wind und Wellen getrieben,</td>
<td><strong>&lt;s199&gt;</strong> as millions of free-swimming larvae of all coral-dwelling organisms do.</td>
</tr>
<tr>
<td><strong>dels er det arter [som svømmer aktivt inn mot revet i den “hensikt” å slå seg ned der] –</strong></td>
<td><strong>&lt;s199&gt;</strong> aber aktiv anschwimmend in der “Absicht”, sich auf dem Riff selbst niederzulassen],</td>
<td></td>
</tr>
<tr>
<td>[slik millioner og atter millioner av de frittstående larvene til alle revets innvånere gjør.]</td>
<td>[wie das die Millionen und Abermilionen der freischwimmenden Larven aller riffbewohnenden Organismen tun.]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The table contains translations of the text into different languages, indicating the different perspectives and uses of language in information packaging and translation.
The source text in (20) has a quite transparent discourse structure: The opening sentence is a general statement which is elaborated by the following two parallel sentences, each giving an example of a fish species specializing in a certain way. The first of these two sentences, <s101>, contains a subordinate (finite) clause, opened by the subjunctor indem ‘in that, by’, that specifies what it means for a fish species to ‘support itself as an “unskilled labourer”’, and following the indem-clause we find a relative clause elaborating on the description of the KLEINERE LEBEWESEN ‘smaller creatures’ introduced by its matrix. In the second sentence, <s102>, the noun LEBEWESEN ‘creatures’ – the object of fressen ‘eat(ing)’ – is modified by a prenuclear participal phrase (italics) and a postnuclear relative clause as well. Literal translation into English:

<s102> On the other hand, A FISH SPECIES can specialize in eating on the reef itself living and then always in some way protected CREATURES WHOSE protective mechanisms IT then will have to render harmless in some way.

The English version in (20) adheres to the global structure of the source text but reduces the hierarchical depth of the second sentence by translating the nested subordinate clauses as parallel ing-clauses; the second occurrence of hunting, in fact, has the sole function of making it possible to elaborate on the description of the object referent declared as CREATURES THAT ... in the first hunting-clause which, in its turn corresponds to the German indem-clause. In addition, the information that plankton is driven passively by wind and waves is left untranslated, conceivably because it can be considered redundant, i.e., retrievable from the context or general background knowledge. Altogether, then, the English translation shortens and syntactically simplifies the source sentence <s101> by way of “implicitation” strategies which are the reverse of the explicitation techniques mentioned in section 1 and 2. The third English sentence is a fairly straightforward translation of <s102> that needs no further comments in the present context.

The Norwegian translation of <s100>–<s102> runs, in principle, as follows:
Here a FISH SPECIES ... can take up very different professions.

THE FISH can support itself quite well as an “unskilled labourer” doing what any average fish can do.

It can hunt SMALLER ORGANISMS WHICH are not poisenous and not armour-plated and cannot defend themselves.

SUCH MICRO-ORGANISMS come to the reef in huge masses from the open sea,

partly we have to do with “plankton” that floats passively with wind and waves,

partly it is species which swim actively towards the reef with the purpose of settling down there, as millions and millions of free-swimming larvae of all coral-dwelling organisms do.

On the other hand, A FISH SPECIES can specialize in living on ORGANISMS that live on the reef themselves.

These are always protected or armoured in some way or other, and this defence THE FISH will have to render harmless.

The Norwegian translation, in the first place, splits the second German sentence into a single independent sentence – <s125>, the Principal Counterpart (S*) – and a paratactic sequence of three sentences – <s126> – corresponding to the complex subordinate event-specifying indem-clause. <s125> contains a relative clause SOM ... which translates the premodifying participal adjunct (italics) of LEBEWesen. This has triggered the sententialization of the postmodifying relative clause DIE ... which turns up as <s126> (cf. section 5.1) with the two parallel specifying adjuncts teils als „Plankton” ..., teils aber aktiv anschwimmend ... sententialized by the specifying constructions det dreier seg om ... and det gr ... mentioned in section 4.3.

The third German sentence has been split into a single independent sentence <s127>, the Principal Counterpart of <s102>, and two coordinated main clauses <s128> which correspond to the second part of the premodifying adjunct (italics) of the object LEBEWesen and the (postmodifying) relative clause DEREN ..., respectively.

Thus, by way of forward information extraction, a sequence of two parallel source sentences is rendered as a sequence of seven sentences which gets segmented to a certain degree through punctuation (comma vs. full stop) and explicit coordination, but evidently not enough to compensate for the hierarchical structure inherent in the source text. As a consequence, the elaboration on the two types of “Lebewesen”-referents, which is packed into adjuncts in the source text, gets too much ‘weight’ or attention, interrupting the main structure that concerns the evolutionary specialization of fish. What is worse: Senten-

33 This holds not only for the text passage in (20) but also for the continuation that is left out in the example.
tializing the event-specification described by the subordinate indekse-clause in <s101> means removing it from the scope of the modal operator kann 'can, may' in the main clause. So in order to embed that information in the proper modal context, the corresponding Norwegian modal verb kan has to be repeated in the independent sentence (<s125>) translating the subordinate clause. And that obscures the discourse relation between <s124> and <s125> as compared to the main and the subordinate clause in <s101>: To someone who does not know very much about how average fish make their living, <s124> and <s125> may very well be interpreted as parallels – as the first two sentences in a list of sentences specifying possible fish professions; and it is not until encountering the discourse marker på den andre siden ‘on the other hand’ in <s127> that the intended interpretation presents itself unambiguously, forcing readers on the discourse garden path to revise their interpretation. That is, upon reading <s125> one may be inclined to build the discourse structure shown in fig. 7 (a) as opposed to the intended structure shown in fig. 7 (b)34 – which is the one that mirrors the syntactic structure of the source text.

34 For an explication of the notation used in fig. 8, the reader is referred to Asher (1993:256ff) and Fabricius-Hansen (1996).
Fig. 7 Assigning discourse structure to Norwegian target text in (20)

I hope to have shown that, taken separately, repeated backward and forward sentence splitting may both cause processing problems at the global discourse level which somehow reduce the positive effect of increased incrementality. Obviously, the risk of structural indeterminacy at discourse level will rise further when the translation combines forward and backward information extraction. In such a case, target sentences occurring between two text focal sentences S*i and S*j (principal counterparts of two neighboring source sentences) could, in principle, either be preparatory with respect to S*j, being ‘derived’ through backward information extraction, or they could relate to S*i by way of forward sentence splitting; cf. fig. 6(c) above. But we would, on the other hand, expect any sequence of source sentences to be mapped on a sequence of target sentences in such a way that in order for a target sentence Sj to relate to S*i (with i < j or i > j) through information extraction, the same must hold for all sentences between Sj and S*i. That is, information extraction should probably be conceived as being executed recursively both backward (cf. section 4.2) and forward, as shown in fig. 8. I shall not, however, pursue this issue further but try to round up.

6. Summary and research perspectives

The main bulk of this paper has been concerned with sententialization or sentence splitting in the stricter sense, i.e., the relation holding between a source sentence, on the one hand, and a sequence of sentences translating that sentence, on the other hand. This should be distinguished from ‘clausalization’, i.e., rendering non-clausal source constituents as subordinate clauses without changing sentence boundaries. Sentence splitting in the stricter sense may or may not increase informational explicitness (or redundancy) at the propositional level, depending on whether it is a non-clausal constituent or a subordinate clause that is translated as an independent sentence. But normally, it does increase incrementality at the level of discourse information, portioning the discourse information out in smaller chunks to be processed and integrated into the mental representation (DRS) of the text one by one.

Defining the Principal Counterpart S* of a source sentence *S as a designated member of the target sentence sequence translating *S allows us to distinguish between backward and forward sentence splitting – or information extraction. I have claimed that backward information extraction will be the natural choice when the information in question has presuppositional status in S* whereas forward information extraction is probably more preferred otherwise. But the validity of that claim has to be checked against more data than my observations are based on. I have also argued that repeated information extraction from nested source structures can be conceived as a top-down recursive procedure, at least normally; that means that it is the information extracted from the most deeply embedded positions in *S that will be encoded in the sentences farthest away from S*, i.e., on the left periphery of the target sequence for backward sentence splitting and on the right periphery

35 Like, for instance, the corpus used by Solfjeld (1998).
for forward information extraction, as shown in fig. 8. But here, too, much empirical work has to be done; in particular, one will have to consider the possible interplay between syntactic structure, on the one hand, and the linear order of the constituents carrying the information, on the other hand.

Fig. 8  Top-down recursive information extraction

Whereas the direction information extraction takes depends heavily on the discourse status of the information in question, the exact form it takes is determined by (i) syntactic and semantic properties of the source constituent encoding that information, including the ontological type (Kamp/Reyle 1993, Asher 1993) or reference domain (von Stutterheim 1997) of the discourse referent involved, (ii) lexical and syntactic options or preferences of the target language, and (iii) the target context the ‘new’ sentence has to fit into. As far as (i) and (ii) is concerned, Solfjeld (1998) distinguishes between Finitisierung ‘finitization’ and Ausbau ‘expansion’. In the first case, the verbal predicate of the ‘new’ target sentence $S$ is a lexical translation of a non-finite or nominalized verb contained in the source constituent $S$ may be said to correspond to, i.e., what Solfjeld calls the Quellenkonstruktion ‘source construction’; this is the normal sententialization strategy applied to nominalizations and non-finite verb phrases, including premodifying participial adjuncts (Solfjeld 1998); cf. for instance (4)–(6), <s78> in (9), and <s125>, <s127> in (20). In the second case, the source construction itself does not contain a non-finite or nominalized verb to be exploited as a predicate under sententialization, so the translator has to follow more creative strategies when construing $S$, drawing more heavily on information that is not expressed explicitly in the source sentence; (8), (11)–(12), <s153> in (16), and (18) offer some examples of the expansion strategy. It is an interesting question – not really taken up by Solfjeld but hinted at in section 4.3 above – how the encoding of sententialized information is influenced by the direction of the information extraction.

Finally, I have argued that, given a carefully written source text, extensive information extraction – which may be demanded on syntactic or stylistic grounds – will tend to have a negative effect on discourse processing at the more global level: It may be more difficult to assess the discourse functions of the individual sentences and, in particular, to identify the sentences that are intended to be focal in the sense that they belong to the ‘main structure’ – contribute to answering the ‘quaestio’ (Klein/von Stutterheim 1991, von Stutterheim 1997) – of the source text. At this point, our findings fall well in line with the conclusions von Stutterheim (1997) arrives at concerning the function of clausal subordination (hypotaxis) in spoken language:
„Fassen wir die Ergebnisse zusammen, so kann man festhalten, daß die Wahl subordinierter Ausdrucksformen weder individuell geprägten stilistischen Kriterien folgt, noch im wesentlichen auf die Schriftsprache beschränkt ist [...]. Vielmehr ermöglichen hypotaktische Ausdrucksformen Transparenz des Textes durch den Erhalt einer obersten Textebene und erleichtern folglich das Verstehen.” (von Stutterheim 1997:247)

Sentence splitting may also create attachment ambiguities for non-focal sentences due to the fact that a non-focal sentence can, in principle, either have a preparatory function with respect to a subsequent focal sentence or some kind of supplementary function with respect to a precedent focal sentence, depending on whether it is ‘derived’ by backward or forward information extraction.

Obviously, our findings are preliminary and contain many loose ends. But on the whole, they raise some interesting questions in addition to the questions asked in the discussions above. For instance:

- Do English and Norwegian translations differ substantially and in principled ways as to how they solve translational problems caused by hierarchical information packaging in German source texts? And in case they do: in which respects and for what reasons? Or are the differences observed above accidental, to be explained by idiosyncratic translation styles – or differences in translation proficiency?

- Given that Norwegian target texts resulting from extensive and recursive sentence splitting tend to have the characteristics and deficiencies suggested above, viewed as self-contained texts, one might ask whether that holds for authentic Norwegian texts written in an incremental style, too. And if not, how do authentic texts compensate for the lack of overt hierarchical structure? Could it be that incremental writing in Norwegian follows other strategies of referential movement or has other preferences, using e.g., preparatory sentences – right-focal discourse structure – to a larger extent than can be extrapolated from translations? These are questions that have to be settled by comparing authentic parallel texts, i.e., within the field of corpus-based contrastive textology or discourse studies.

- Is it feasible to construct algorithms of backward and forward sentence splitting along the lines suggested above that would allow sentence splitting under automatic or computer-aided translation from German into Norwegian or English?

---

36 'Summarizing our results, we conclude that the choice of subordinated expressions neither is determined by individual stylistic criteria nor is it restricted to written language [...]. On the contrary, by maintaining a top level of the text, hypotactic encoding contributes to text transparency and thus facilitates understanding.' (Translation by CFH).

37 It is, incidentally, not quite clear to me whether preparatory sentences which, in a way, ‘look to the right’ can be handled adequately within theories of discourse structure that, like Asher (1993), only allow discourse attachment to the left.

38 See e.g., Johansson/Oksefjell (1998).
I hope, at least, to have demonstrated that by holding the information to be conveyed and the intended global discourse structure under control, so to speak, translation-based studies settled within a precise theoretical framework of dynamic semantics can contribute in interesting ways to our understanding of discourse structure, referential movement and functional aspects of hierarchical syntactic structure.

References


**Text and translations**

