

---

## A Grammar for Finnish Discourse Patterns

KRISTIINA JOKINEN

### 22.1 Introduction

This article deals with Finnish discourse oriented word-order variations, and provides their implementation in the HPSG-style typed feature structure grammar using the LKB toolkit (Copetstone, 2002). It does not present a full-coverage Finnish grammar or even a small HPSG fragment of the standard syntactic phenomena in Finnish. Rather, the aim has been to implement the Finnish discourse configuration in the Finnish Discourse Pattern Grammar (FDPG), employing typed feature structures and old and new discourse information, and thus to supply a starting point for further research in computational modelling of syntax-discourse interplay. The goal is motivated by the need for a dialogue system to analyse utterances and generate responses using semantic representation which is rich enough to encode discourse referents with different information status. The dialogue manager distinguishes old and new information, keeps track of the discourse topic, and also provides a context e.g. for the specific corrections where the speaker objects what has been stated in the previous utterance and contrasts it with a new fact. The use of topic and new information in language generation is discussed in Jokinen and Wilcock (2003) in more detail.

The interpretation of the Finnish word-order variations is based on Vilkuna (1989). She points out that the different syntactic orders reflect a discourse configurational structure of the language: constituents in certain positions are always interpreted as conveying particular discourse functions. In order to parse the word-order variations in the HPSG grammar formalism, I will argue

in favour of discourse *patterns*. These are fixed orders of the main sentential constituents based on Vilkuna's discourse configuration and used for presenting and interpreting discourse information in utterances. I have extended the head-complement and head-specifier rules in the HPSG grammar with a set of combination rules that concern discourse patterns, so that the patterns can be effectively used in parsing the various word orders.

The article is organized as follows. I will first review Vilkuna's discourse configuration for simple transitive sentences and discuss its relation to the information structure. This is followed by a short introduction to HPSG, the LKB formalism, and typed feature-structures. I will then present the implementation of the discourse patterns in LKB, and finally discuss some points for further research.

## 22.2 Finnish Discourse Syntax

### 22.2.1 Word-order variations

Vilkuna (1989) defines the following discourse configuration for Finnish:

| Kontrast | Topic | Verb | Rest |

The main verb divides the sentence into two parts. The positions in front of the verb carry special discourse functions while the Rest-field after the verb contains constituents in no particular order. (The end of the sentence, however, marks new information, see below.) The two specific discourse functions are Kontrast (K) and Topic (T), assigned to the elements occupying the sentence-initial position (K) and the position immediately in front of the main verb (T). The T-position marks the current discourse topic, i.e. what the sentence is about. The K-position can be occupied by a discourse referent which is contrasted with the topic of the previous sentence. It is always a marked position with contrastive emphasis, and it can also be empty.

In order to determine the information status of the constituents, the Prague school question-answering method is used: one seeks for a suitable question that the sentence provides new information for, and the information status of the constituents is determined in relation to this context. Notice that in dialogues, answers typically realize only the new information, since Topic and discourse-old information can be inferred from the previous utterance and discourse context (Jokinen and Wilcock, 2003). If the utterance has K-position filled, the underlying discourse context does not contain a question but rather a statement that is contrasted or corrected, see examples below and in Section 22.4.2.

For a simple transitive sentence, the following alternatives are possible:

| Kontrast | Topic    | Verb     | Rest     | English equivalent                  |
|----------|----------|----------|----------|-------------------------------------|
| 1        | Karhu    | pyydysti | kalan    | The bear caught the fish            |
| 2        | Kalan    | pyydysti | karhu    | The fish is caught by the bear      |
| 3        | Kalan    | karhu    | pyydysti | It is the fish that the bear caught |
| 4        | Karhu    | kalan    | pyydysti | It is the bear that caught the fish |
| 5        | Pyydysti | karhu    | kalan    | The bear DID catch the fish         |
| 6        | Pyydysti | kalan    | karhu    |                                     |

Sentence (1) represents the canonical word order for Finnish: it has subject in the T-position and object in the Rest-field. Statistically it is also the most common word order, supporting the fact that the subject usually encodes the topic. As for the information structure, three alternatives are possible: the whole event can be new as in the presentation sentence (“What happened?”), the verb phrase can be new (“What did the bear do?”), or only the object can be new (“What did the bear catch?”). The sentence (2) is analogous, except that the constituents have now swapped places: the object is Topic while the subject introduces new information in the discourse. The utterance matches the question “Who caught the fish?”

Sentences (3) and (4) signal correction in regard to the previous discourse. They pair up so that the sentence initial K-position is occupied by the object/subject which is contrasted with another object/subject mentioned earlier in the discourse: e.g. “It is the fish that the bear caught, not an otter”, and “It is the bear that caught the fish, not the wolf”<sup>1</sup>. The sentences (5) and (6) have a special argumentative character, too, since the main verb is in the K-position. In (5), the speaker insists on the truth of the statement (“indeed the bear did catch the fish”), but the word-order is also used if the speaker presents the state of affairs as new, something surprising and contrary to expectations (*no, pyydystin minä pienen kalan* “well I did catch a small fish”). The alternative (6), however, with the object occupying the T-position, is awkward in simple sentences. Obviously this is due to the clash of the two specially marked word order patterns: the preposed and contrasted verb does not fit with the marked word order that indicates the subject as new information.

### 22.2.2 Information structure

Discourse configuration bears similarity to information packaging (Engdahl and Vallduví, 1996), although it does not exactly correspond to the sentential information structure. As Vallduví and Vilkuna (1998) point out, contrastiveness is orthogonal to information structure. While the elements in the Rest-field are new (rheme) and the elements in the T-position are old and carry presupposed information (theme), the information status of the K-position is not so clear; cf. also the failure of the question-answer method to directly provide a

---

<sup>1</sup> Kontrast can also be expressed by intonation in the neutral SVO order: *Karhu pyydysti KALAN*, or *Kalan pyydysti KARHU*. I will not discuss them further here.

context for the sentences (3)-(6) above: the context contains statements rather than queries for new information. Kontrast is of course new with respect to the sentential content, but it can also be old, if the referent has already been introduced in the discourse context. For instance, (4) can occur after the discourse like "I saw a wolf and a bear by the lake" - "and it was the wolf that caught a fish?" - "No, not at all, it was the bear that caught the fish, not the wolf". In fact, in this case we have a curious situation where a discourse referent is simultaneously old and new; Vilkuna (1989) calls these Topic-Focus cases. In FDPG, discourse referents in the K-position are regarded as new, since to the hearer, contrast is new information, and the discourse referent that turns the proposition into a new fact is the one occupying the K-position.

I have previously (Jokinen, 1994) introduced Topic and NewInfo as two mutually exclusive features to distinguish two types of discourse referents: Topic represents what the utterance is about and NewInfo what is new in the discourse context. NewInfo is related to Topic: it describes something new with respect to the discourse topic. If the whole event is new, the discourse referent for the verb is marked as NewInfo, and we have a presentation sentence. The distinction agrees with that proposed by Vallduví & Vilkuna (1998), who describe topic as an anchor to the focus (new information). I will not go into details of semantic representation of Topic and NewInfo, but refer to Wilcock (this volume) who discusses different representations for information structure and indicates how Minimal Recursion Semantics can be extended to take information structure into account.

## 22.3 LKB, HPSG, and FDPG

### 22.3.1 Preliminaries

The first implementation of the basic Finnish word-order variations is presented in Karttunen and Kay (1985). They describe a parser for free-word order languages, and use functional unification grammar marking topic and new information as specific features on the constituents. For FDPG, I have used LKB<sup>2</sup> as the development tool. This is an open source grammar toolkit for implementing natural language grammars in the typed feature structure formalism. Most implementations in LKB use HPSG, but the LKB itself is powerful enough to allow grammars in any constraint-based linguistic formalism to be developed. The grammar files include lexicon (lexical entry definitions), rules (feature structures describing how signs can be unified), and types (type specifications that constrain on sign unification). The toolkit consists of various tools for the developer to write and debug grammars, and it comes with several sample grammars as well as a full stepwise course for learning how to build grammars.

---

<sup>2</sup><http://www.delph-in.net/lkb/>

HPSG (Head-Driven Phrase Structure Grammar, Pollard & Sag 1994) is a strict lexicalist approach to human language modelling. It assigns rich information structures to words, and using various constraints, it projects phrasal categories and sentences from the words. Lexical heads specify information like part-of-speech and dependency relations, and also encode the basic semantic information of their phrasal projections.

The representation of lexical items, like that of the projected phrases is a uniform feature structure called the *sign*. A sign consists of attributes and their values, encoding phonological, morpho-syntactic, semantic and pragmatic information of the entries. In the ERG grammar<sup>3</sup>, a sign contains the following features: ORTH (orthographical realization of the lexical sign), SYNSEM (syntactic and semantic information), LEX (lexical status), NON-LOCAL (non-local information), and HEAD (head information). The representation of signs is made more compact by organizing them into an *inheritance hierarchy*, according to which the signs can inherit certain properties from the more general entries above them in the hierarchy. Inheritance hierarchies are based on *typed feature structures*: each sign is associated with a type which constrains free unification of the otherwise compatible signs.

In HPSG, the arguments of lexical entries are divided into complements and specifiers. The two main rules that can be applied to lexical entries to form phrasal projections are the *head-complement rule*, which unifies the sign of the lexical head with the signs of its complements, and the *head-specifier rule*, which forms a saturated phrase by unifying the phrasal sign with the sign of the specifier. The sentential specifier is the subject, and the syntactically saturated phrase is a sentence.

The complement list is encoded in the lexical signs COMPS-feature, and the specifier specification into the SPR-feature, both of which are SYNSEM-features. The complement list and the application of the head-complement and the head-specifier rules are ordered, so possible word order variations must be described by other means. A simple solution is to allow multiple verb entries, one for each different word order that the lexical item can project. However, this explodes the lexicon, and for languages like Finnish, it is not a reasonable alternative. One can also introduce a special permutation rule that produces necessary variations in the COMPS-list for any lexical entry. The problem with this alternative is that it increases grammar processing time. Yet another solution is to use unordered sets as COMPS-features instead of lists. For instance, the Japanese JACY grammar<sup>4</sup> orders possible argument structures into a type hierarchy and allows different head-complement rules to pick up the arguments in the COMPS-list in any order.

---

<sup>3</sup>The LinGO English Resource Grammar, see <http://www.delph-in.net/erg/>

<sup>4</sup>Homepage: <http://www.dfki.de/~siegel/grammar-download/JACY-grammar.html>

In Finnish, however, the driving force in sentence analysis and generation seems to be discourse configuration rather than syntactic constituency. FDPG thus uses discourse patterns, defined in terms of the discourse functions Kontrast, Topic, and NewInfo as the main elements in describing Finnish word-order. Moreover, in a well-formed utterance, NewInfo must always be explicitly present.

In their HPSG account for English and Catalan information packaging, Endahl and Vallduví (1996) suggest that the specific feature INFO-STRUCT, comprising the features FOCUS and GROUND, with the latter being further divided into LINK and TAIL, be added in the context field of lexical signs so as to encode sentential information structure. Following this, I also assume that the discourse configuration is a separate dimension in sentence analysis. However, instead of the feature INFO-STRUCT, I use DISC-STRUCT with the features KONTRAST, TOPIC, and NEWINFO<sup>5</sup>.

Furthermore, I have replaced the head-complement and head-specifier rules by a set of special rules that describe how various discourse patterns can be combined and interpreted as the speaker/hearer (incrementally) parses the particular word order variations. The patterns can be ordered in a type hierarchy, although this is not done in the sample grammar. Since the focus of the article is on word order only, I have also made some simplifying assumptions about the morphosyntactic properties of Finnish:

- Morphology is encoded in the lexical entries,
- NPs require determiner.

### 22.3.2 The Finnish grammar categories

FDPG regards *utterance* as the smallest unit in syntax, emphasizing its occurrence as part of the discourse and being uttered by a speaker. The utterance sign has two fields: DISC-STRUC for discourse configuration, and SYNSEM-STRUC for the syntactic arguments and their semantics. The sign is of type *utt-struc*, and defined as follows:

```
utterance ::= utt-struc &
[ DISC-STRUC disc-struc,
  SYNSEM-STRUC synsem-struc,
  ARGS *list* ].
```

The DISC-STRUC contains the following features (EVENT encodes the event reference denoted by the main verb):

```
disc-struc ::= utt-struc &
[ KONTR *list*,
  TOPIC *list*,
```

---

<sup>5</sup>NEWINFO relates to FOCUS, and TOPIC to LINK, but KONTRAST has no apparent counterpart.

```
NEWINFO synsem-struc,
EVENT evtype,
RESTFIELD *list* ].
```

The SYNSEM-STRUC contains feature representations for the lexical head, its syntactic dependents and semantics, and the DEP-feature includes the typical HPSG specifier and complement-lists:

```
synsem-struc ::= utt-struc &
[ HEAD pos,
  DEP dependents,
  SEM semantics ].

dependents ::= feat-struc &
[ SPR *list*,
  COMPS *list* ].
```

Finally, phrases are projections of lexical items such that all of their complements have been found, i.e. the COMPS-list is an empty list.

```
phrase ::= utterance &
[ SYNSEM-STRUC [ DEP [COMPS < > ] ] ].
```

FDPG also introduces a special subj-phrase which has an empty SPR-list. Although specifiers do not mark saturated phases as is the case in HPSG in general, their separate marking has been retained, however, since in some discourse patterns the subject can be combined with the main verb before the other complements, and thus it is necessary to distinguish signs which have been unified with their subject from those that still have to find it.

```
subj-phrase ::= utterance &
[ SYNSEM-STRUC [ DEP [SPR < > ] ] ].
```

Some lexical verbs do not usually take nominative-case subjects (emotional and physiological states, nature descriptions), and accordingly, their lexical entries have SPR-feature instantiated to null. Some sentence types (existential and possessive sentences) do not have nominative-case subjects either, and in this case, it is the main verb *olla* "be" that has SPR-feature as null.

## 22.4 Discourse Patterns

The discourse patterns are rules which define how the lexical signs can be unified with their arguments, and they can be seen as an extension of the HPSG head-complement and head-specifier rules. The FDCG idea is to unify the lexical verb sign first with the immediate left/right adjacent NP sign, regardless of whether the NP has subject- or object-marking. This unification

serves as the basic pattern for marking discourse functions. The rest of the discourse functions are filled in when the second argument is unified with the basic pattern, resulting in a saturated phrase.

#### 22.4.1 Head-complement patterns

The head-complement patterns deal with the pair-wise combination of the main verb (lexical head) with the adjacent NP sign (be it an NP functioning as subject or object), and they correspond to the HPSG head-complement rules. For space restrictions, the full feature structure is given only for the first pattern. Subj stands for a nominative case NP and Obj for an accusative NP. All examples are based on the event “the bear caught the fish”.

##### 1) Subject-Topic pattern:

Subj + V    *karhu pyydysti*    “the bear caught”

The rule combines the main verb with a nominative case NP on the left. The result is a feature structure for an utterance with both the NP and the main verb marked as Topic: Topic(Subj) + Topic(V)

```
subj-top-rule := utterance &
[ DISC-STRUC [ TOPIC < #1, #2 > ],
  SYNSEM-STRUC [ORTH #orth,
                 HEAD #head,
                 DEP [ SPR < >,
                       COMPS #comps ],
                 SEM #sem ],
  ARGs < phrase & #1 & [SYNSEM-STRUC
                           [HEAD noun & [ AGR [ CASE nom ] ],
                            DEP [ SPR < > ] ] ],
  word & #2 &
  [ORTH #orth,
   HEAD #head & verb,
   DEP [ SPR < #1 >,
         COMPS #comps ],
   SEM #sem ] > ].
```

##### 2) Object-Topic pattern:

Obj + V    *kalan pyydysti*    “the fish was caught by”

The rule is parallel to (1), but it combines the main verb with an accusative case NP (object). Also in this case, the resulting utterance has both the NP and the main verb marked as Topic: Topic(Obj) + Topic(V)

**3) Verb-Kontrast pattern:**

V + Subj *pyydysti karhu* “caught the bear”

This rule combines the verb with a nominative case NP which is now on the right. The result is an utterance where the main verb is marked as Kontrast and the subject-NP as Topic, i.e. something the utterance (and the correction) is about: Kontrast (V) + Topic (Subj). The pattern anticipates the speaker’s disagreement and wish to express a correction of what has been stated before.

**4) Verb-New pattern:**

V + Obj *pyydysti kalan* “caught the fish”

The rule is parallel to (3) and combines the verb with an accusative case NP on the right. However, the result has both the verb and the object-NP as new information: NewInfo(V) + NewInfo(Obj). Notice that the verb is not interpreted as Kontrast like in (3), since this would lead to awkward interpretations when the subject-NP is unified with the phrase (see discussion about the sentence (6) in Section 22.2.1, and the impossible combinations in Section 22.4.3).

**22.4.2 Head-specifier patterns**

The head-specifier patterns are analogous to the HPSG specifier rules which saturate the phrasal sign with the specifier, i.e. in the case of a verbal sign with the subject-NP. However, in FDPG, the rules do not apply only to subject-NPs but also to object-NPs, and also the discourse context comes to play a role in the unification. A rule can be applied only if the information status of the NP accords with that of the underlying discourse pattern. Thus the head-specifier patterns guide the parsing and constrain acceptability of a particular word-order with respect to the appropriate information structure of the utterance. Suitable discourse contexts are shown after each rule by an underlying question or statement that the resulting utterance addresses to. The abbreviations SV, OV, VS, and VO refer to the constituents formed by the patterns (1-4) above, and the utterances to the example utterances in Section 22.2.1.

**1a) [Subject-Topic] Object-New pattern (utterance 1):**

SV + Obj *karhu pyydysti + kalan* “the bear caught + the fish”

The rule produces one of the three information structures for the canonical word-order as discussed in Section 22.2.1 (the other two word-orders are produced by Rule 4a below). The basic SV pattern is already determined as Topic

(Rule 1), so the resulting utterance is used to introduce a new object-NP. The utterance thus functions as a response to an underlying question "What did the bear caught?" with the object-NP as NewInfo "(the bear caught) the fish": *Mitä karhu pyydysti? – (karhu pyydysti) kalan*

**1b) [Subject-Topic] Object-Kontrast pattern (utterance 3):**

Obj + SV    *kalan + karhu pyydysti*    "the fish + the bear caught"

The unification of the basic SV pattern with the object-NP on the left results in an utterance where the object-NP is Kontrast. The SV pattern is determined as Topic as above(Rule 1), but the previous context has also presented an object which the speaker wants to contrast and correct. For instance, the previous context may contain a statement like "the bear caught the otter", and the speaker then corrects this with a new object: *Eipäs, kalan (karhu pyydysti)* "no, it was the fish (that the bear caught)".

**2a) [Object-Topic] Subject-New pattern (utterance 2):**

OV + Subj    *kalan pyydysti+karhu*    "the fish was caught by +the bear"

The rule is parallel to (1a) above: now the basic OV pattern is Topic (Rule 2), and the subject-NP on the right is introduced as new. The underlying question is "Who caught the fish?" with the subject marked as NewInfo "(The fish was caught by) a bear": *Kuka pyydysti kalan? - (kalan pyydysti) karhu.*

**2b) [Object-Topic] Subject-Kontrast pattern (utterance 4):**

Subj + OV    *karhu+kalan pyydysti*    "the bear + the fish was caught by"

Analogously to (1b) above, unification of the topical OV pattern with a subject-NP on the left results in a correction and a contrastive utterance. A previous statement, like "the wolf caught the fish", is contrasted with the new subject-NP as Kontrast: *Eipäs, karhu (kalan pyydysti)* "no, it was the bear that caught the fish".

**3a) [Verb-Kontrast] Object-Rest pattern (utterance 5):**

VS + Obj    *pyydysti karhu + kalan*    "caught the bear + the fish"

The rule combines the VS pattern with an object-NP on the right. The verb is already marked as Kontrast and the subject as Topic (Rule 3), and the object falls in the Rest-field.

The object can be either discourse old or new information. If the object-NP is old, the contrast concerns the actual event which the subject and the ob-

ject denote the participants of. The utterance occurs in contexts where doubts about the truth of the speaker's assertion have been raised ("I wonder if the bear caught the fish after all"), and the speaker wants to reinforce and insist on the original argument: "yes indeed the bear DID catch the fish": *Enpä usko että karhu pyydysti kalan. – Kyllä toki pyydysti karhu kalan.*

The object-NP can also be NewInfo as in the answers to nosy questions: *ja pyydystitkös mitäään? – no, pyydystin minä pienen kalan* "and did you catch anything? - well, I did catch a small fish". This is a rather common pattern, since the speaker provides new information as a response to a genuine question. The contrast in this case concerns the implicit negative presupposition of the question "you may not have caught anything", which is contrasted by the speaker's positive answer. It should be noticed that the question "What did you catch?" presupposes that the partner caught something, while the question "Did you catch anything?" lacks such a presupposition. It is interesting that Finnish reflects the difference in the presuppositions in the word-orders that the possible answers to the these questions exhibit: the former is encoded in the Object-New rule (1a) and the latter in Object-Rest rule (3a).

#### **4a) [Verb-New] Verb-Presentation pattern (utterance 1):**

Subj + VO    *karhu + pyydysti kalan*    "the bear+caught the fish"

The rule produces two of the three canonical word-orders (cf. Rule 1a). The new information in the verb-object pattern (determined by Rule 4) is combined with the two possible discourse statuses of the subject-NP. If the subject-NP is Topic, the resulting utterance is simply an answer to the question "what did S do?": *Mitä karhu teki? – (Karhu) pyydysti kalan.*

If the subject-NP is new in the discourse context, the result is a presentation sentence with all the constituents as NewInfo, answering to the question "what happened?": *Mitä tapahtui? – Karhu pyydysti kalan.*

#### **22.4.3 Impossible combinations**

##### **3b) Impossible-Verb-Kontrast-Object-Kontrast combination**

Obj + VS

A symmetrical rule for (3a) would combine a contrasted verb and a topical subject-NP with an object-NP on the left. This is impossible, however, since the K-position is always sentence-initial, and there is "no space" left for a second Kontrast in front of the already contrasted verb. It must be noticed that the Obj-V-Subj word-order is fine if the discourse configuration is different: see the Object-Topic-Subject-New rule above (2a).

**4b) Impossible-Verb-Two-News combination**

NewInfo(VO) + NewInfo(Subj)

The other impossible combinations are based on the VO-pattern with the subject-NP on the right: VO + Subj.

If the VO-pattern is discourse new (introduced by Rule 4), it would seem natural to add subject-NP as new information at the end of the utterance. However, the combination is confusing: the resulting word-order is the marked discourse pattern for a sentence-initial Kontrast, while the all-new information status of the constituents suggests a presentation sentence. There is no Kontrast, and for a presentation sentence, the canonical SVO-order is preferred (rule 4a); hence unification is impossible. Even if we assume that the verb indeed is Kontrast and object-NP is NewInfo, Kontrast(V) + NewInfo(O) + NewInfo(Subj), the combination would still lead to confusion, since there is no Topic to anchor the contrast to.

**4c) Impossible-Verb-Kontrast-Object-Topic-New-Subject combination**

Kontrast(V) + Topic(O) + NewInfo(Subj)

If the verb is Kontrast and object-NP is Topic, the combination is analogous to the contrasting VS + Obj pattern licensed by the rule (3a). However, the order seems to favour the reading of the subject as NewInfo, and thus its interpretation is again confusing between whether the utterance is about contrasting events or a new subject. In the former case, the preferred combination would be VS+O (Rule 3a) and in the latter case OV+S (Rule 2a).

**4d) Verb-Object-Topic-New-Subject combination**

Topic(VO) + NewInfo(Subj)

If the VO-pattern is Topic, the subject must be NewInfo. In simple sentences, this combination will again run against the marked sentence-initial Kontrast pattern as well as the preferred Subject-New rule for introducing new subjects (Rule 2a). However, if an adverbial is added in the beginning of the utterance, the order becomes acceptable, although there are strong expectations that the contrast now continues with respect to the adverbial: *eilen pyydysti kalan karhu, tänään koira* “yesterday it was the bear that caught the fish, today the dog”. The pattern suggests that there are two pairwise Kontrasts: the adverbials on the one hand and the subject-NPs on the other hand. The verb and the object-NP make up the topical background for the Kontrasts. In the current FDPG, which only deals with simple sentences, these combinations are not possible. Obviously, the grammar rules should be relaxed in order to allow

the unification of the constituents to take place, but they should also constrain the result to be unsaturated, so as to force the adverbial Kontrasts to be added in the sign and make the phrase well-formed from the discourse point of view. These combinations will require more detailed research.

## 22.5 Discussion

In this article I have presented a HPSG-based implementation of the Finnish word-order variations in the LKB grammar environment. Following Vilkuna (1989), the Finnish syntax is characterized by its discourse configuration, which assigns certain discourse functions (Kontrast, Topic, NewInfo) to the constituents according to their position with respect to the main verb. The Finnish Discourse Pattern Grammar (FDPG) can parse simple intransitive and transitive sentences and produce appropriate discourse interpretations of the different word-orders without spurious parses. The grammar is available on request from the author.

The combination rules are based on simple discourse patterns concerning the main verb and its adjacent NP-complements. There are different patterns for combining the main syntactic constituents: four for pair-wise combinations of the main verb and an adjacent NP, and six for producing saturated utterances with the discourse functions appropriately filled in. The former extend the head-complement rule of the traditional HPSG, while the latter extend the HPSG head-specifier rule. The patterns with their associated discourse functions can also be thought of providing guidance for the hearer about what to expect next in the on-coming discourse.

In Finnish, phrase structure thus seems like an epiphenomenon that occurs as a side effect of the lexical entries being projected into full sentences and their dependents ordered into a coherent discourse. From this view-point, it would, of course, be more natural to describe syntactic relations with the help of a dependency grammar which explicitly reveals the dependency relations between the verb and its arguments, than with a phrase structure grammar which focusses on phrasal structures. In fact, the patterns can be also seen as possible ways to combine dependency relations into surface strings.

I have effectively proposed a new approach to syntax: that of discourse configuration. In this approach the speaker's intention to exchange new information on a particular topic is taken as the driving force for communication, and this intention is not realized on the level of dialogue organisation only, but trickles down to the syntactic structure as well. Discourse information, carried by the different word orders in Finnish, is thus efficiently used by the hearers when processing the incoming utterance: the presence of certain discourse patterns directs the hearer to expectations concerning upcoming elements. Cognitive studies also seem to support psychological reality of discourse pat-

terns, their incremental processing and impact on the hearer's expectations about the yet-to-come elements. For instance, in a recent study of processing Finnish word-order variations, Kaiser and Trueswell (2004) found empirical evidence that shows how the hearers make efficient use of the non-canonical word order patterns to predict upcoming referents and their discourse status.

Finally, the FDPG is an attempt to provide a model for discourse configurational syntax that would describe the link between the syntactic-semantic representation of utterances and the information they encode of dialogue situations. The view of dialogue events as the determining factor for sentence structuring may prove useful in modelling also other "free" word-order languages like Japanese, where the discourse function Topic is grammaticalised and none of the verb arguments are obligatory in well-formed sentences. Of course, systematic investigations are needed to substantiate this hypothesis.

## 22.6 Acknowledgments

The research was conducted while I was a Visiting Fellow in Clare Hall at the University of Cambridge, associated with the NLP group in the Computer Laboratory. I would like to thank Ann Copestake for her assistance and useful discussions concerning LKB implementation, Karen Spärck Jones for providing an encouraging environment, and Graham Wilcock for discussions on HPSG grammars in general.

## References

- Copestake, A. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.
- Engdahl, E. and Vallduví, E. 1996. Information packaging in HPSG. In C. Grover and E. Vallduví, eds., *Edinburgh Working Papers in Cognitive Science, Vol. 12: Studies in HPSG*, pages 1–32. University of Edinburgh.
- Jokinen, K. and Wilcock, G. 2003. Adaptivity and response generation in a spoken dialogue system. In J. van Kuppevelt and R. Smith, eds., *Current and New Directions in Discourse and Dialogue (Text, Speech and Language Technology, Vol. 22)*, pages 213–234. Kluwer Academic Publishers.
- Karttunen, L. and Kay, M. 1985. Parsing a Free Word Order Language. *Natural Language Parsing*, eds. D. Dowty, L. Karttunen and A. Zwicky, 279–306. Cambridge: Cambridge University Press.
- Kaiser, E. and Trueswell, J. C. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94(2), 113–147.
- Vallduví, E. and Vilkuna, M. 1998. *On Rheme and Kontrast. The Limits of Syntax*, eds. P. Culicover and L. McNally, 79–109. New York: Academic Press.
- Vilkuna, M. 1989. *Free Word Order in Finnish. Its Syntax and Discourse Functions*. Helsinki: Suomalais-Uuden Seura.
- Wilcock, G. 2005. Information Structure and Minimal Recursion Semantics. (this volume)