



Conexor Functional Dependency Grammar 3.7 - User's Manual

© Conexor oy, 1997-2002

English description

Linguistic changelog for English

Linguistic changes between Conexor FDG for English 3.6 and 3.7

To ensure backward compatibility between levels 3.6 and 3.7 note the following descriptive changes.

1. Base forms of some adverbs

In former levels, some adverbs still received an adjectival base form, e.g. the adverb "terribly" received the base form "terrible". Now all adverbs receive an adverb as base form (e.g. "terribly" is given the base form "terribly").

2. Unnamed links no longer exist

Former levels produced unnamed links. In level 3.7 all dependency links are assigned a name according to the [functional type](#) inventory. Note however, that linking is still optional.

3. Linking of punctuation revised

In former levels, some spurious link names, *pnct:* and *list:*, appeared. These link names are suppressed in the current level.

Linguistic changes between Conexor FDG for English 3.4 and 3.6

To ensure backward compatibility between levels 3.4 and 3.6 in Conexor's English CG, FDG Lite and FDG note the following changes.

1. Base forms of derived adverbs

For adverbs derived from adjectives, e.g. "clearly":

- 3.4 produces an adjectival base form, e.g. "clear"
- 3.6 produces an adverb base form, e.g. "clearly"

Still, some adverbs receive an adjectival base form, e.g. the adverb "terribly" received the base form "terrible" (cf. above)

2. **ING, EN forms**

Level 3.4 produces a verbal base form for all ING and EN forms (present participles, past participles, and the corresponding adjectives and nouns), e.g. "walk" for:

Walking is tiresome.
He is walking.
He walked.
The walked distance was long.

Level 3.6 produces a verbal base form for ING and EN forms used as verbs and a nominal base form for ING and EN forms used as nominals, e.g. "walk" for:

He is walking.
He walked.

"walking" for:

Walking is tiresome.

"walked" for:

The walked distance was long.

3. **Omission of <*> tag as superfluous**

The <*> tag signifies that the word contains upper case letters. As this information can be derived from the word directly, <*> is no longer retained in the output of 3.6.

4. **The <??> tag is replaced by the <?> tag**

Both tags mean that the token was morphologically analysed by the guesser, i.e. the token was not represented by the lexicon itself.

- 3.4 produces <??>
- 3.6 produces <?>

5. Lowercase base forms

Some of the base forms produced by 3.4 contained upper case letters; 3.6 consistently produces lower-case base forms.

6. Empty readings

3.4 sometimes produced empty readings (which emerged as the sequences of two or more tabs). 3.6 produces no empty readings.

7. Omission of <ADV-N> tag as redundant

3.4 produced the <ADV-N> tag for nouns often used adverbially; 3.6 no longer produces this tag. At the level of FDG analysis, this seems irrelevant, and users of the more shallow analysers are probably more interested in words, terms as such, rather than in clause-level syntax.

8. Omission of <>CC> tag as redundant

3.4 produced the <>CC> tag as introducer of coordination, such as *both*, *either* and *neither*. 3.6 no longer produces this tag.

9. Omission of morphological tags concerning number

Level 3.4 produced some tags that have become redundant and will be removed from the FDG and FDG Lite analysers in level 3.6. In table 1 are listed the tags and some of the forms they were associated with.

SG2/PL2	you
-SG1,3	are
SG1,3	was
-SG3	almost anything not 3 person singular, e.g. run

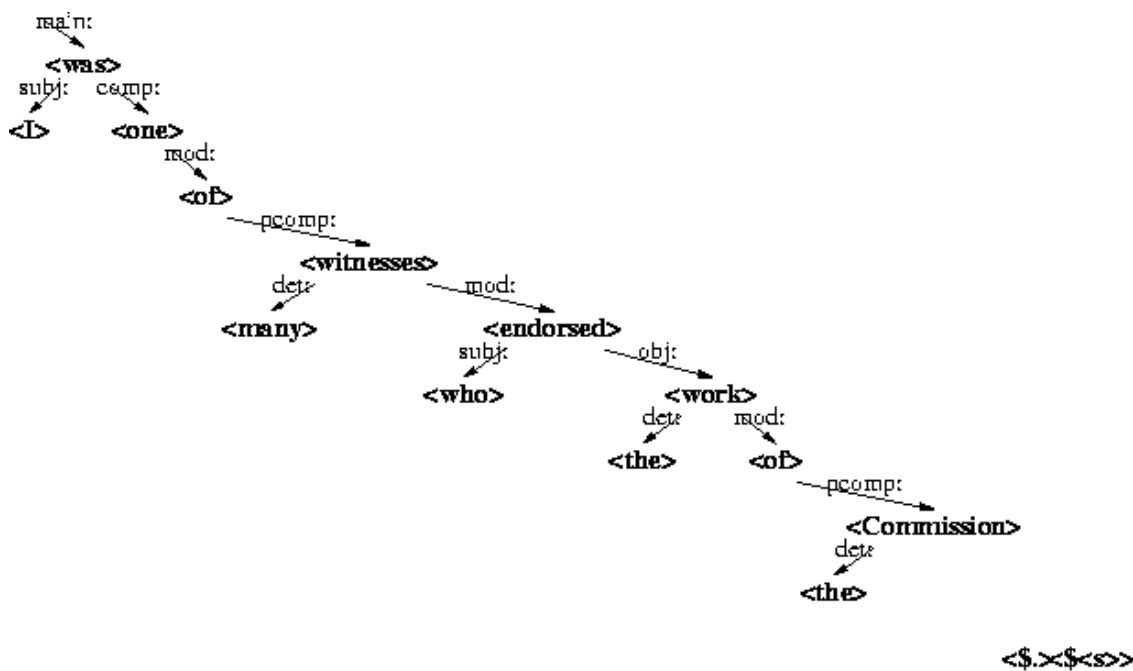
SG/PL	the, what, ever, people, customs, names: Netherlands etc.
-------	--

Table 1. Removed morphological tags

Linguistic representation

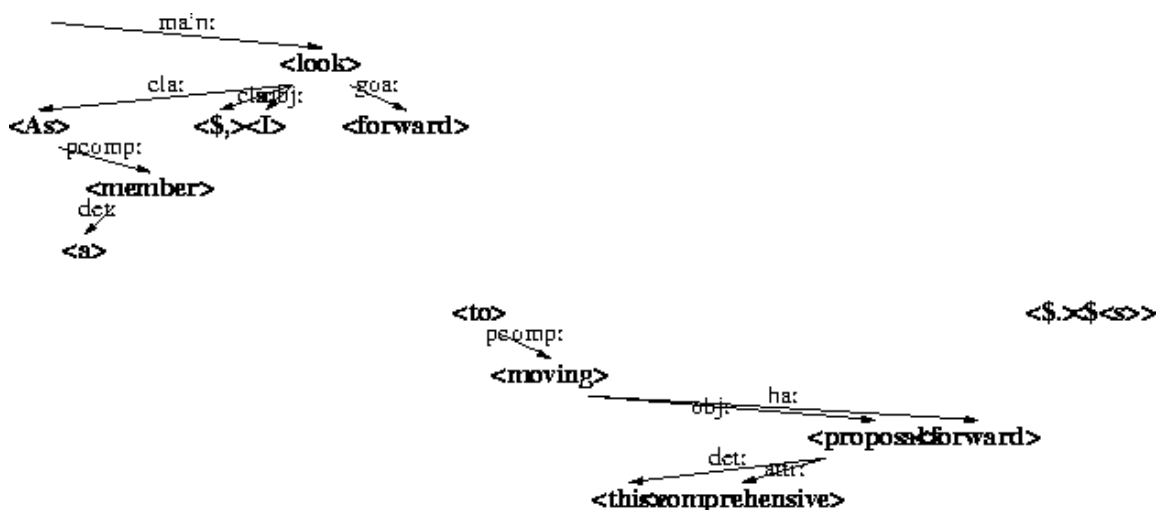
The Conexor Functional Dependency Grammar (FDG) parser builds [functionally labelled dependencies](#) between most words and assigns morphosyntactic tags to all words. The linguistic representation consists of [morphological tags](#), [surface syntactic tags](#) and [functional tags](#).

Here is a sample analysis of a dependency parser in visual format (which shows the labelled links only):



I was one of many witnesses who endorsed the work of the Commission.

When building a full dependency tree does not succeed, the parser produces partial descriptions that correspond to partial trees:



As a member, I look forward to moving this comprehensive proposal forward.

Partial descriptions often result from unresolved ambiguity.

For the sentence *As a member, I look forward to moving this comprehensive proposal forward*, the parser produces the following analysis:

1	As	as	copred:>6	@ADVL %EH PREP
2	a	a	det:>3	@DN> %>N DET SG
3	member	member	pcomp:>1	@<P %NH N NOM SG
4	,	,		
5	I	i	subj:>6	@SUBJ %NH PRON PERS NOM SG1
6	look	look	main:>0	@+FMAINV %VA V PRES
7	forward	forward	goa:>6	@ADVL %EH ADV
8	to	to		@ADVL %EH PREP
9	moving	move	pcomp:>8	@<P-FMAINV %VA ING
10	this	this	det:>12	@DN> %>N DET DEM SG
11	comprehensive	comprehensive	attr:>12	@A> %>N A ABS
12	proposal	proposal	obj:>9	@OBJ %NH N NOM SG
13	forward	forward	ha:>9	@ADVL %EH ADV
14	.	.		
15	<p>	<p>		

The analysis consists of columns separated by tabs. Column 1 gives a numerical address to each word; columns 2 and 3 show the word-form and the lemma; column 4 shows a dependency function (if any); column 5 shows syntactic function tags (beginning with @); surface syntactic tags (beginning with %) and morphological tags. In the case of unresolved ambiguity, additional columns will be shown for the type of information given in column 5.

Thus the preposition *As*, as head of the prepositional phrase *As a member*, is a clause adverbial for word number 6 (*look*); *a* is a determiner of word number 3; word number 3 (*member*) is a nominal head functioning as complement of the preposition *As*; etc.

English dependency functions (for FDG only)

Tag	Explanation	Example
main	main element: main nucleus of the sentence; usu. main verb of the main clause	The Berkeley UNIX <i>mechanism</i> for creating a virtual connection between processes. Sockets <i>form</i> the interface between UNIX standard I/O and network communication facilities.
qtag	tag question	It is cold, <i>isn't</i> it?
v-ch	verb chain: auxiliaries + main verb	If you're running the mess-dos emulator, control-alt-insert <i>will</i> cause a soft boot of the emulator, while leaving the rest of the system running.
pm	preposed marker: grammatical marker of a subordinated clause. The marker (subordinating conjunction) itself doesn't have a syntactic function in the subordinated clause.	Others go further and define software <i>to</i> be programs plus documentation <i>though</i> this does not correspond with common usage.
pcomp	prepositional complement: the head of a nominal construction (NP or non-finite clause or nominal clause) that, together with a preposition, forms a prepositional phrase. Usually a preposition precedes its complement, but also topicalised complements occur.	They are in that red <i>car</i> . She is fond of <i>walking</i> long distances. <i>What</i> are you afraid of?

phr	verb particle: certain preposition-adverb homographs that form a phrasal verb with a verb	She looked <i>up</i> the word in the dictionary.
subj	subject: the head of an NP that agrees in number with the verb in the clause. Often signals the semantic category called agent.	<i>John</i> is in the kitchen.
agt	agent: The agent by-phrase in passive sentences.	The dog was chased <i>by</i> the boys.
obj	object: the head of the other main nominal dependent of transitive verbs (and ditransitive verbs, together with indirect objects)	John saw an <i>apple</i> . John gave him an <i>apple</i> .
comp	subject complement: the head of the other main nominal dependent of copular verbs.	John remains a <i>boy</i> . What you see is what you <i>get</i> . John is <i>foolish</i> .
dat	indirect object: Ditransitive verbs can take three nominal dependents: subject, indirect object, object.	John gave <i>him</i> an apple.
oc	object complement: a nominal category that occurs along with an object for object complementiser verbs.	John called him a <i>fool</i> . John considers him <i>foolish</i> .
copred	copredicative	John regards him <i>as</i> foolish.
com	comitative	Drinking <i>with</i> you is nice.
voc	vocative	<i>John</i> , come here!
ins	instrument	He sliced the salami <i>with</i> the knife.
tmp	time	If you're running the mess-dos emulator, control-alt-insert will cause a soft boot of the emulator, <i>while leaving</i> the rest of the system running.

dur	duration	The OECD praises the relative stability of US unemployment as "remarkable", given the 50 per cent increase in the American US labour force <i>in</i> the past 25 years.
frq	frequency	It <i>often</i> involves the use of CASE tools.
qua	quantity	Singapore says <i>more</i> about this than Hong Kong.
man	manner	If Europe is so wonderful, they argue, why does its job creation record compare so <i>poorly</i> with that of the United States?
loc	location	That exacerbates a key problem <i>in</i> America, the skills gap.
sou	source	Policymakers in both seem to be moving away <i>from</i> the characteristics that defined them.
goa	goal	Virgin is expected to try to move <i>to</i> a full anti-trust trial.
pth	path	He travelled from Tokyo <i>to</i> Beijing.
cnt	contingency (purpose or reason)	The DTI was unable to say last night <i>why</i> the approach for Frances Colliery had been rejected.
cnd	condition	If Europe <i>is</i> so wonderful, they argue, why does its job creation record compare so poorly with that of the United States?
meta	clause adverbial	<i>So far</i> , the OECD has refused to disclose its country-by-country studies.
cla	clause initial adverbial	<i>Under</i> President Clinton, the highly flexible US labour market is becoming more regulated.

ha	heuristic prepositional phrase attachment	Eventually the beam will escape <i>through</i> the partially reflective mirror.
qn	quantifier	IFA Promotion, which represents more than <i>15,000</i> independent financial advisers, commissioned a random poll of new year resolutions.
det	determiner	Nearly <i>a</i> third of East Anglians resolved to stay healthy in 1995.
neg	negator	What a blessing the releases were <i>not</i> in Finnish!
attr	attributive nominal	By <i>Philip Bassett, Industrial Editor</i> .
mod	other postmodifier	Ministers will see the OECD report <i>on</i> the UK labour market as international recognition <i>of</i> their reform <i>of</i> one <i>of</i> the economy's most difficult areas.
ad	attributive adverbial	<i>So</i> much for modern technology.
cc	Coordination: The coordinating conjunction and one coordinated element are linked to the other coordinated element. Multiple coordinated elements are chained together. The upmost element in a chain shows the functional role of the coordinated units.	<i>Jack and Jill</i> bought some pins, <i>nails and needles</i> .

English morphological tags

Part of speech	Subfeature	Explanation	Example
N		noun	These integrated <i>algorithms</i> are stored on the <i>computer's</i> hard <i>disk</i> .
-- case	NOM	nominative	These integrated <i>algorithms</i> are stored on the computer's hard <i>disk</i> .
	GEN	genitive	These integrated algorithms are stored on the <i>computer's</i> hard disk.
-- number	SG	singular	These integrated algorithms are stored on the <i>computer's</i> hard <i>disk</i> .
	PL	plural	These integrated <i>algorithms</i> are stored on the computer's hard disk.

With nouns, the obligatory tags include 'N' and case. In Conexor FDG, the obligatory tags for nouns include 'N', case, and number.

ABBR	abbreviation	"SODA Manual of Operation", R. C. Brigham and C. G. Bell, School of Elec <i>Eng</i> , U New S Wales, Sydney, NSW (1958)
-- case and number like in nouns		

With abbreviations, the obligatory tags include 'ABBR' and case. In Conexor FDG, the obligatory tags for abbreviations include 'ABBR', case and number.

A		adjective	These integrated algorithms are stored on the computer's <i>hard</i> disk, from which they are downloaded into the DSP board's <i>random</i> access memory (RAM).
-- comparison	ABS	absolute	<i>big</i>
	CMP	comparative	<i>bigger</i>
	SUP	superlative	<i>biggest</i>

With adjectives, the obligatory tags include 'A' and degree of comparison.

NUM		numeral	Software can be split roughly into <i>two</i> main types - system software and application software or programs.
	CARD	cardinal	<i>2010</i>
	ORD	ordinal	<i>first</i>
-- number	SG	fraction, singular	<i>one-third</i>
	PL	fraction, plural	<i>two-thirds</i>

With numerals, the obligatory tags include 'NUM' and either 'CARD' or 'ORD'.

PRON		pronoun	<i>Others</i> go further and define software to be programs plus documentation though <i>this</i> does not correspond with common usage.
-- case and related features	NOM	nominative	<i>others</i>
	GEN	genitive	<i>other's</i>
	ACC	accusative	<i>him</i>

	INDEP	the independent genitive form functioning always as head of a noun phrase	<i>theirs</i>
-- number	SG	singular	<i>other</i>
	SG1	singular, first person	<i>me</i>
	SG3	singular, third person	<i>him</i>
	PL	plural	<i>others</i>
	PL1	plural, first person	<i>us</i>
	PL3	plural, third person	<i>them</i>
-- comparison	ABS	absolute	<i>much</i>
	CMP	comparative	<i>more</i>
	SUP	superlative	<i>most</i>
-- other pronoun subfeatures	PERS	personal	<i>us</i>
	DEM	demonstrative	<i>these</i>
	RECIPR	reciprocal	<i>each other</i>
	WH	relative or interrogative pronoun beginning with the letters 'wh' or 'how'	<i>which</i>
	<Interr>	interrogative	<i>why</i>
	<Refl>	reflexive	<i>herself</i>
	<Rel>	relative	<i>which</i>

With pronouns, the only obligatory tag is 'PRON'. The anglebracket tags

occur in front of the 'PRON' tag, when relevant. After the 'PRON' tag come the tags 'PERS', 'RECIPR', 'WH', 'DEM', or degree of comparison, when relevant. Next is the place for case, then number, and last, 'INDEP', when relevant.

DET		determiner	If you're running <i>the</i> mess-dos emulator, control-alt-insert will cause <i>a</i> soft boot of <i>the</i> emulator, while leaving <i>the</i> rest of <i>the</i> system running.
-- case	GEN	genitive	<i>whose</i>
-- number	SG	singular	<i>an</i> option
	PL	plural	<i>these</i> options
-- comparison	ABS	absolute	<i>many</i> options
	CMP	comparative	<i>more</i> options
	SUP	superlative	<i>most</i> options
-- other subfeatures of determiners	DEM	demonstrative	<i>this</i> option
	WH	determiner beginning with the letters 'wh' or 'how'	<i>which</i> option

With determiners, the only obligatory tag is 'DET'. In Conexor FDG, the obligatory tags for determiners include 'DET' and number. In the ordering of the morphological tags, the number tag is the last one and the case tag second to last, when relevant.

ADV		adverb	Others go <i>further</i> and define software to be programs plus documentation though this does not correspond with common usage.
-- comparison	ABS	absolute	<i>far</i>
	CMP	comparative	<i>further</i>
	SUP	superlative	<i>furthest</i>

-- other subfeatures for adverbs	<Ex>	existential <i>there</i>	<i>There</i> are various models of the software life-cycle, and many methodologies for the different phases.
	WH	adverb beginning with the letters 'wh' or 'how'	<i>why</i>

With adverbs, the obligatory tags include 'ADV'. The anglebracket tag again occurs in front of the 'ADV' tag, when relevant.

ING		present participle	The Berkeley UNIX mechanism for <i>creating</i> a virtual connection between processes.
EN		past participle	These <i>integrated</i> algorithms are <i>stored</i> on the computer's hard disk, from which they are <i>downloaded</i> into the DSP board's random access memory (RAM).

The tags EN and ING are used in Conexor FDG Lite for participles in all syntactic functions, whereas in Conexor FDG, they are used only for participles functioning as a verb. In Conexor FDG, participles in nominal functions are classified as adjectives or nouns. Thus, in the above example, the word *integrated* receives the tag EN in Conexor FDG Lite, whereas in Conexor FDG, it is classified as an adjective.

V		verb; used only for finite verbs and infinitives	Others <i>go</i> further and <i>define</i> software to <i>be</i> programs plus documentation though this <i>does not correspond</i> with common usage.
	AUXMOD	modal auxiliary	<i>would</i>
	INF	infinitive	would <i>be</i>
	IMP	imperative	John, <i>come</i> here!
	SUBJUNCTIVE	subjunctive	The casket <i>be</i> brought in.

-- tense	PRES	present tense	<i>are</i>
	PAST	past tense	<i>were</i>
-- person	SG1	singular, first person	<i>am</i>
	SG3	singular, third person	<i>is</i>
-- other subfeatures for verbs	<N+>	noun-verb combination	<i>India's got..</i>

With verbs, the obligatory tags include 'V' and one of the following: 'AUXMOD', 'INF', 'IMP', or tense. With tense, person is possible. The subfeature <N+> is placed before the 'V' tag when relevant.

INTERJ	interjection	<i>Hey, so-and-so needs an instruction to do such-and-such.</i>
CC	coordinating conjunction	<i>and</i>
CS	subordinating conjunction	<i>if</i>
PREP	preposition	<i>of</i>
NEG-PART	the negative particle	<i>are not, aren't</i>
INFMARK>	infinitive marker	<i>to do this in order to do that</i>
<?>	mark for unknown word; occurs in front of a part-of-speech tag	<i>mechansim</i>

English functional tags

Tag	Explanation	Example
@+FAUXV	Finite auxiliary predicator	If you're running the mess-dos emulator, control-alt-insert <i>will</i> cause a soft boot of the emulator, while leaving the rest of the system running.
@-FAUXV	Nonfinite auxiliary predicator	Software can <i>be</i> split roughly into two main types - system software and application software or programs.
@+FMAINV	Finite main predicator	Sockets <i>form</i> the interface between UNIX standard I/O and network communication facilities.
@-FMAINV	Nonfinite main predicator	If you're <i>running</i> the mess-dos emulator, control-alt-insert <i>will cause</i> a soft boot of the emulator, while <i>leaving</i> the rest of the system <i>running</i> .
@SUBJ	Subject	<i>Sockets</i> form the interface between UNIX standard I/O and network communication facilities.
@F-SUBJ	Formal subject	<i>There</i> are various models of the software life-cycle, and many methodologies for the different phases.

@OBJ	Object	If you're running the mess-dos <i>emulator</i> , control-alt-insert will cause a soft <i>boot</i> of the emulator, while leaving the <i>rest</i> of the system running.
@I-OBJ	Indirect object	John gave <i>him</i> an apple.
@PCOMPL-S	Subject complement	A statistic that is <i>content-free</i> , or nearly so.
@PCOMPL-O	Object complement	This downloading, or "booting" process of the PC-installed software algorithms occurs as part of the computer's power-up initialisation process in less than 100 milliseconds, making it <i>transparent</i> to the user.
@ADVL	Adverbial	Others go <i>further</i> and define software to be programs plus documentation though this does <i>not</i> correspond <i>with</i> common usage.
@O-ADVL	Object adverbial	She let him walk the <i>streets</i> in the cold and in the rain.
@APP	Apposition	Software can be split roughly into two main types - system <i>software</i> and application <i>software</i> or <i>programs</i> .
@NH	Stray noun phrase	The Berkeley UNIX <i>mechanism</i> for creating a virtual connection between processes.
@VOC	Vocative	<i>John</i> , come here!
@A>	Premodifier of a nominal	These <i>integrated</i> algorithms are stored on the <i>computer's hard</i> disk, from which they are downloaded into the <i>DSP board's random access</i> memory (RAM).

@DN>	Determiner	If you're running <i>the</i> mess-dos emulator, control-alt-insert will cause <i>a</i> soft boot of <i>the</i> emulator, while leaving <i>the</i> rest of <i>the</i> system running.
@QN>	Premodifying quantifier	This downloading, or "booting" process of the PC-installed software algorithms occurs as part of the computer's power-up initialisation process in less than <i>100</i> milliseconds, making it transparent to the user.
@AD-A>	Intensifier	An optical laser works by bouncing photons back and forth between two mirrors, one <i>totally</i> reflective and one <i>partially</i> reflective.
@<NOM-OF	Postmodifying prepositional phrase beginning with <i>of</i>	If you're running the mess-dos emulator, control-alt-insert will cause a soft boot <i>of</i> the emulator, while leaving the rest <i>of</i> the system running.
@<AD-A	Postmodifying intensifier	Compuserve developed the GIF format for graphical images many years <i>ago</i> .
@<NOM	Postmodifier of a nominal	The Berkeley UNIX mechanism <i>for</i> creating a virtual connection <i>between</i> processes.
@INFMARK>	Infinitive marker <i>to</i>	Others go further and define software <i>to</i> be programs plus documentation though this does not correspond with common usage.
@<P-FMAINV	Nonfinite clause as preposition complement	The Berkeley UNIX mechanism <i>for creating</i> a virtual connection between processes.

@<P	Other preposition complement	If you're running the mess-dos emulator, control-alt-insert will cause a soft boot of the <i>emulator</i> , while leaving the rest of the <i>system</i> running.
@CC	Coordinating conjunction	Others go further <i>and</i> define software to be programs <i>plus</i> documentation though this does not correspond with common usage.
@CS	Subordinating conjunction	Others go further and define software to be programs plus documentation <i>though</i> this does not correspond with common usage.
@DUMMY	Unspecified	<i>Hey</i> , so-and-so needs an instruction to do such-and-such.

English surface syntactic tags

There are two parallel tagsets of English surface syntactic tags. The FDG parser outputs a tagset beginning with '%' and the FDG Lite parser outputs a tagset beginning with '&'. The tags and their explanations are listed in the table below. Note that the tagsets differ also in other details besides their prefixes: in the surface syntactic tags of FDG, the letter E stands for 'adverb', therefore '%E>' pro '&>A'.

The surface syntactic tags of FDG Lite for English are allocated through a fast surface analysis: on a 233 MHz Pentium PC running Linux, FDG Lite for English analyses text at the speed of 2,000 words per second. The surface syntactic tags of FDG for English, on the other hand, are allocated based on full dependency parsing of the sentence structure. Consequently, their accuracy is better, their ambiguity is smaller, and the analysis takes more time: on a 233 MHz Pentium PC running Linux, FDG for English analyses text at the speed of 300 words per second.

Tags in FDG Lite	Tags in FDG	Explanation	Example
&>N	%>N	determiner or premodifier of a nominal	<i>These integrated</i> algorithms are stored on <i>the computer's hard</i> disk, from which they are downloaded into <i>the DSP board's random access</i> memory (RAM).
&NH	%NH	nominal head	<i>Sockets</i> form the <i>interface</i> between UNIX standard <i>I/O</i> and network communication <i>facilities</i> .
&N<	%N<	postmodifier of a nominal	The Berkeley UNIX mechanism <i>for</i> creating a virtual connection <i>between</i> processes.

&>A	%E>	premodifying adverb	An optical laser works by bouncing photons back and forth between two mirrors, one <i>totally</i> reflective and one <i>partially</i> reflective.
&AH	%EH	adverbial head (besides adverbs, applies to interjections, prepositions, and the negative particle)	Others go <i>further</i> and define software to be programs plus documentation though this does <i>not</i> correspond <i>with</i> common usage.
&A<	%<E	postmodifying adverb	He knew it well <i>enough</i> .
&AUX	%AUX	auxiliary verb or particle	Software <i>can be</i> split roughly into two main types - system software and application software or programs.
&VP	%VP	main verb in a passive verb chain	These integrated algorithms are <i>stored</i> on the computer's hard disk, from which they are <i>downloaded</i> into the DSP board's random access memory (RAM).
&VA	%VA	main verb in an active verb chain	If you're <i>running</i> the mess-dos emulator, control-alt-insert will <i>cause</i> a soft boot of the emulator, while <i>leaving</i> the rest of the system <i>running</i> .
&>CC	%EH>	Introducer of coordination	<i>both</i> Harry and Bill came
&CC	%CC	coordinating conjunction	Others go further <i>and</i> define software to be programs <i>plus</i> documentation though this does not correspond with common usage.
&CS	%CS	subordinating conjunction	Others go further and define software to be programs plus documentation <i>though</i> this does not correspond with common usage.