

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

Clusters

Prefabs

Bundles

Chunks

– or simply: repeated identical phrases

such as

a lot of

the fact that

at the end

at the end of the day

etc.

The question:

How do Danish students use them – and how do native speakers use them?

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

Two quotations:

John Sinclair:

Words enter into meaningful relations with other words around them, and yet all our current descriptions marginalise this massive contribution to meaning. (Sinclair 1996: 76).

Göran Kjellmar:

While the typical moderately fluent native speaker makes considerable hesitation pauses between often quite long sequences of words [...], the typical moderately fluent learner pauses after every two or three words. It seems reasonable to believe that the difference between them in this regard can be ascribed largely to a difference in the automation of collocations. (Kjellmar 1991: 124)

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

Danish Student Corpus:

App. 75.000 words

111 essays

thus average length: app. 675 word (range 300 – 1000 words).

Students' age: 18+/- 1

Grade : 11th – one year before graduation.

"Gymnasiet"'s language line.

Reference corpus:

British A-level from

The Centre for English Corpus Linguistics (CECL) at Université Catholique de Louvain in Belgium – the so-called Loeness

See

<http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/cecl.html>

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

The exam paper

The student gets a modern short story (usually around 2000 words). The assignment:

Write an essay about the text in which you include the points listed below. Your essay must also include a couple of quotations which are relevant to your interpretation. (Your essay should not include a summary of the text.)

- the setting
- a characterization of
 - {e.g. the main character}
 - {e.g. other characters}
- the theme(s)

They get 3 hours to write the essay (and a short (80 word) translation from Danish into English). They can use dictionaries, grammar books, old essays etc. But they must write the essay individually.

More information:

Undervisningsministeriet: Censorvejledning,

<http://us.uvm.dk/gymnasie//almen/eksamen/censor/skr/cvejleng01.html>

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

Method:

Assemble strings of for example 3 words:

A)

- 1) Word #1 + word #2 + word #3
- 2) Word #2 + word #3 + word #4
- 3) Word #3 + word#4 + word #5
- 4) etc. etc.

B)

Sort them,

C)

Count number of identical clusters.

Software: *PhraseContext* – see <http://www.hjkm.dk/PhraseContext>

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.

Odense, Oct 2004.

3-word clusters in Danish and British essays

	Students' essays			A-Level essays		
	Different clusters	Tokens	% of all tokens	Different clusters	Tokens	% of all tokens
3-word min 2	5513	54087	77%	3131	27903	46,5%
3-word min 5	734	19.914	28,4%	336	8.319	13,8%
4-word	2186	22.600	32,3%	964	9396	15,7%
5-word	742	8.830	12,6%	263	2915	4,9%
6-word	275	3.646	5,2%	70	876	1,5%

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

Students' 15 most frequent clusters:

63	a lot of
62	that she is (????)
59	the fact that
*49	<i>she wants to</i>
*47	<i>the narrator is</i>
43	that he is (????)
42	he is a
*40	<i>the short story</i>
*40	<i>Di and Andrea</i>
*39	<i>the whaler is</i>
*39	<i>takes place in</i>
39	she does not (????)
39	I think that
*36	<i>of the story</i>
35	it is a

A-level students' 15 most frequent clusters

*36	<i>the human brain</i>
36	one of the
33	due to the
32	the number of
31	that it is
30	there is a
*29	<i>the national lottery</i>
28	the fact that
26	such as the
26	a lot of
24	there is no
24	it would be
*23	<i>the lottery is</i>
23	should not be
23	have to be

* *italics*: content specific clusters – roman: "discourse" markers/rhetorical devices

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

6 words - the 7 most frequent clusters

Danish students:

11	the story takes place in a
8	a home for single teenage girls.
7	a home for single teenage girls
6	is a home for single teenage
5	pushed into a world that cannot
5	into a world that cannot accommodate
5	in the beginning of the story

British A-level students:

4	the number of cars on the
3	to choose the sex of their
3	the head of the church of
3	redundant by the invention of the
3	made redundant by the invention of
2	would not be able to cope
2	we would not be able to

- A-level: much lower over-all frequencies
- Even here: two (the last two) "discourse markers"
- Danish students: no "discourse markers" a high degree of dependence on the texts/subject

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

British university students' literary essays (app. 50.000 words)

4 word clusters:

32	the end of the
21	at the end of
*18	<i>the people of argos</i>
*18	<i>in le mythe de</i>
14	the role of the
14	the beginning of the
13	the fact that he
12	end of the play
12	at the beginning of
11	on the other hand
*10	<i>the absurdity of life</i>
8	the way in which
*8	<i>role of the president</i>
*8	<i>for the good of</i>
*8	<i>came to power in</i>
*7	<i>themes of guilt and</i>
*7	<i>the themes of guilt</i>
7	the start of the
*7	<i>the citizens of argos</i>
*7	<i>of guilt and remorse</i>
*7	<i>le mythe de sisyph</i>
7	is shown to be
7	is seen to be
*7	<i>guilt and remorse are</i>
*7	<i>an act of bad</i>

6 word clusters:

3	towards the end of the play
4	the themes of guilt and remorse
3	the themes of guilt and judgement
3	the role of the prime minister
3	of the play he is seen
4	of no use to the party.
4	at the start of the play,
3	at the end of the play,
8	at the end of the play
3	at the beginning of the play,
4	at the beginning of the play

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

Medical articles on medication of cardiovascular diseases

(app. 55.000 words)

4 word clusters

16	treatment of high blood
16	prevalence of hypertension in
16	of hypertension in the
16	joint national committee on
16	and treatment of high
13	the joint national committee
13	of the joint national
13	an ace inhibitor or
12	zur bekämpfung des hohen
12	report of the joint
12	national committee on detection,
12	liga zur bekämpfung des
22	of high blood pressure
10	whose initial therapy was
10	of hypertension in belgian

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.

Odense, Oct 2004.

Individual variations

	Grade 10 Excellent		Grade 8 Average		Grade 6 Just passed		Grade 5 Not passed	
	Clusters	Tokens	Clusters	Tokens	Clusters	Tokens	Clusters	Tokens
3 words	5	36	6	39	17	177	35	249
4 words	1	8	0	0	3	24	15	128
5 words	0	0	0	0	0	0	6	60
6 words	0	0	0	0	0	0	1	12

Grading done by the Danish Ministry of Education

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.
Odense, Oct 2004.

Conclusion(s)

- **Danish students *do* use clusters and prefabs to a degree which is comparable with – or above - that of native speakers *but***
- **Their prefabs are usually repetitions of the texts they are writing about to a higher degree than that of native speakers – often comparable with highly specialized scientific literature – not “discourse markers” or rhetorical devices**

Sylvie de Cock:

.... results show that advanced learners [of English] use prefabs, and in some cases even more prefabs than NSs [Native Speakers]. Consequently, they can be said to apply the idiom principle, but the chunks they use (1) are not necessarily the same as those used by NSs, (2) are not used with the same frequency, (3) have different syntactic uses, and (4) fulfil different pragmatic functions.” (de Cock, S et al: ”The Phrasicon of EFL Learners”, in Granger, S: *Learner English on Computer*, Longman, London 1998.)

Hans J. Klarskov Mortensen:

Clusters and prefabs in intermediate/lower intermediate English essays written by Danish students.

Odense, Oct 2004.

Consequences for teaching.

4 word clusters in Bank of English (100 mil. words) (compiled by D. Willis, quoted from Hunston 2003)

Number

2074	the end of the
1834	a lot of people
1650	nice to talk to
1531	that sort of thing
1189	a lot of the
1098	quite a lot of
1089	a bit of a
896	end of the day
654	of the things that
608	the rest of the
595	a lot of money
570	a little bit of
565	in terms of the
549	to go to the
514	an awful lot of
477	in the first place
441	that kind of thing
437	this sort of thing
392	per cent of the
389	got a lot of
383	a little bit more
366	a couple of years
351	a lot of time
346	a lot of things
346	most of the time
337	used to go to
323	think a lot of
324	to make sure that

SO:

The importance of these bundles or phrases is, firstly, that a syllabus designer working with a pedagogic corpus would wish to ensure that the corpus reflected these sorts of figures, if necessary differentiated by register, and secondly, that a materials designer would wish to draw attention to them as useful formulae for learners to use. Hunston (2003: 192):

Teach

- **frequency of clusters**
- **pragmatic function of clusters**