

CORPUS OF SPECIALISED LANGUAGE

Roberta Pittau

The corpus refers to triathlon race reports. The articles have been downloaded from the net.

QUANTITATIVE ANALYSIS

1. Most significant items in the list

- The following list contains the 80 most frequent nouns in the corpus
NOUNS FREQUENCY LIST

Order/original order	Word	Frequency
1. [13]	race	58
2. [15]	bike	45
3. [17]	time	43
4. [19]	run	40
5. [26]	course	29
6. [30]	swim	27
7. [32]	championship	25
8. [33]	world	25
9. [40]	year	23
10. [47]	downhill	20
11. [60]	place	16
12. [61]	pro	16
13. [75]	transition	13
14. [82]	series	12
15. [88]	mile	11
16. [89]	minute	11
17. [90]	minutes	11
18. [91]	races	11
19. [94]	water	11
20. [104]	top	10
21. [105]	tour	10
22. [107]	way	10
23. [113]	champion	9
24. [115]	flat	9
25. [116]	forest	9
26. [128]	women	9
27. [130]	athletes	8
28. [133]	champ	8
29. [143]	men	8
30. [144]	mountain	8
31. [147]	season	8
32. [149]	uphill	8

33.	[150]	win	8
34.	[151]	winner	8
35.	[155]	end	7
36.	[163]	trail	7
37.	[164]	triathlon	7
38.	[167]	climb	6
39.	[173]	lake	6
40.	[174]	line	6
41.	[176]	part	6
42.	[181]	split	6
43.	[185]	thing	6
44.	[188]	title	6
45.	[192]	years	6
46.	[193]	battle	5
47.	[203]	division	5
48.	[205]	final	5
49.	[221]	ride	5
50.	[222]	runner	5
51.	[223]	things	5
52.	[224]	trails	5
53.	[225]	victory	5
54.	[226]	war	5
55.	[233]	advantage	4
56.	[234]	age	4
57.	[236]	amateur	4
58.	[238]	bikes	4
59.	[240]	city	4
60.	[241]	competition	4
61.	[244]	director	4
62.	[245]	distance	4
63.	[246]	double	4
64.	[251]	events	4
65.	[253]	feet	4
66.	[254]	field	4
67.	[257]	hill	4
68.	[258]	history	4
69.	[259]	hometown	4
70.	[262]	lap	4
71.	[266]	name	4
72.	[274]	points	4
73.	[275]	purse	4
74.	[278]	section	4
75.	[279]	sections	4
76.	[282]	stairs	4
77.	[290]	tire	4
78.	[294]	week	4
79.	[301]	airlines	3
80.	[308]	area	3

- The following list contains the 80 most frequent verbs in the corpus
VERBS FREQUENCY LIST

Order/original order	word	frequency
1. [8]	was	117
1. [13]	race	58
2. [19]	run	40
3. [22]	is	34
4. [30]	swim	27
5. [44]	had	21
6. [51]	going	18
7. [57]	said	17
8. [58]	have	16
9. [64]	are	15
10. [66]	were	15
11. [69]	finish	14
12. [77]	be	12
13. [78]	could	12
14. [80]	got	12
15. [87]	made	11
16. [95]	will	11
17. [94]	release	11
18. [96]	did	10
19. [114]	do	9
20. [132]	came	8
21. [136]	finished	8
22. [138]	get	8
23. [140]	lead	8
24. [141]	like	8
25. [150]	win	8
26. [158]	go	7
27. [162]	started	7
28. [167]	climb	6
29. [168]	didn	6
30. [171]	has	6
31. [177]	put	6
32. [178]	ran	6
33. [183]	take	6
34. [191]	went	6
35. [194]	been	5
36. [197]	can	5
37. [198]	climbing	5
38. [200]	crashed	5
39. [204]	done	5
40. [206]	hurt	5
41. [209]	kept	5

42. [211]	knew	5
43. [212]	know	5
44. [215]	making	5
45. [218]	passed	5
46. [219]	placed	5
47. [221]	ride	5
48. [228]	winning	5
49. [242]	couldn	4
50. [249]	enjoy	4
51. [263]	leading	4
52. [273]	pass	4
53. [277]	running	4
54. [281]	should	4
55. [283]	starts	4
56. [289]	thought	4
57. [293]	try	4
58. [297]	would	4
59. [313]	beat	3
60. [314]	being	3
61. [316]	bit	3
62. [317]	bunch	3
63. [318]	called	3
64. [319]	catch	3
65. [322]	close	3
66. [323]	combined	3
67. [324]	coming	3
68. [325]	cross	3
69. [332]	feel	3
70. [337]	getting	3
71. [342]	heading	3
72. [344]	held	3
73. [350]	led	3
74. [351]	left	3
75. [352]	let	3
76. [354]	loose	3
77. [355]	make	3
78. [358]	move	3
79. [370]	push	3
80. [371]	racing	3

- The following list contains the 80 most frequent adjectives in the corpus
ADJECTIVES FREQUENCY LIST

Order/original order	Word	Frequency
1. [18]	that	41

2. [21]	this	36
3. [33]	world	25
4. [35]	two	24
5. [37]	all	23
6. [43]	second	22
7. [45]	her	21
8. [47]	downhill	20
9. [48]	first	20
10. [50]	my	20
11. [52]	last	18
12. [53]	back	17
13. [59]	his	16
14. [68]	down	14
15. [71]	seconds	14
16. [74]	three	13
17. [79]	good	12
18. [98]	fastest	10
19. [102]	third	10
20. [104]	top	10
21. [111]	best	9
22. [112]	black	9
23. [115]	flat	9
24. [117]	hard	9
25. [121]	right	9
26. [122]	steep	9
27. [123]	straight	9
28. [124]	strong	9
29. [131]	big	8
30. [134]	european	8
31. [135]	fifth	8
32. [137]	fourth	8
33. [142]	little	8
34. [145]	new	8
35. [149]	uphill	8
36. [161]	same	7
37. [163]	trail	7
38. [169]	faster	6
39. [170]	female	6
40. [180]	south	6
41. [199]	cold	5
42. [202]	different	5
43. [205]	final	5
44. [217]	old	5
45. [218]	passed	5
46. [219]	placed	5
47. [227]	well	5
48. [232]	able	4
49. [237]	american	4

50. [246]	double	4
51. [252]	fast	4
52. [255]	full	4
53. [265]	middle	4
54. [269]	next	4
55. [272]	overall	4
56. [276]	real	4
57. [280]	short	4
58. [285]	technical	4
59. [306]	annual	3
60. [321]	clear	3
61. [322]	close	3
62. [335]	former	3
63. [340]	hardest	3
64. [343]	heavy	3
65. [345]	high	3
66. [362]	nine	3
67. [376]	seven	3
68. [379]	single	3
69. [381]	some	3
70. [382]	specialist	3
71. [386]	steady	3
72. [388]	sure	3
73. [390]	those	3
74. [391]	tired	3
75. [392]	total	3
76. [400]	whole	3
77. [402]	extreme	3
78. [425]	available	2
79. [426]	bad	2
80. [428]	beautiful	2

- The following list contains the 19 most frequent prepositions in the corpus

PREPOSITIONS FREQUENCY LIST

Order/original order	Word	Frequency
1. [4]	to	137
2. [7]	in	119
3. [9]	on	82
4. [14]	with	48
5. [16]	for	44
6. [24]	at	32
7. [25]	up	31
8. [34]	out	24
9. [38]	from	23
10. [41]	by	22
11. [53]	back	17
12. [68]	down	14
13. [109]	along	9
14. [119]	off	9
15. [120]	over	9
16. [269]	next	4
17. [292]	towards	4
18. [315]	between	3
19. [564]	onto	2

- The following list contains the 13 most frequent adverbs in the corpus

ADVERBS FREQUENCY LIST

1.	as	37
2. [25]	up	31
3. [34]	out	24
4. [53]	back	17
5. [65]	behind	15
6. [68]	down	14
7. [109]	along	9
8. [119]	off	9
9. [120]	over	9
10. [129]	after	8
11. [195]	before	5
12. [269]	next	4
13. [631]	throughout	2

2. Core vocabulary of the target specialised language for each of the four word classes.

NOUNS	VERBS	ADJECTIVES	ADVERBS/PREP.
age	arrived	able	after
amateur	be	available	along
athletes	beat	best	as
battle	bit	big	at
bike	can	challenging	back
champ/champion	catch	close	before
championship	challenge	close	behind
climb	climb	comfortable	between
competition	climb	cool	by
course	close	double	down
director	combine	down	for
distance	come	downhill	from
division	crash	extreme	in
downhill	cross	fast	next
end	defend	fast	off
event	descend	favourites	on
field	do	final	onto
final	enjoy	first	out
flat	feel	flat	over
forest	finish	former	throughout
hill	get	global	to
lap	go	good	towards
line	have	hard	up
men	head	hard	with
mile	hit	heavy	
place	hurt	high	
point	jump	high	
pro	keep	last	
purse	know	late	
race	lead	main	
ride	leave	overall	
run	like	right	
runner	make	second	
season	move	single	
section	pass	specialist	
series	pass	steady	
split	place	steep	
swim	push	straight	
time	put	strong	
tire	race	technical	
title	ride	terrific	
top	run	third	
tour	say	tired	
trail	should	top	

trail transition triathlon uphill victory war way winner	star start swim take think try walk want will win would	total tough unique uphill whole	
---	---	---	--

QUALITATIVE ANALYSIS

1. Typical linguistic patterns of the target language

While narrating their courses athletes use series of clauses with a number of linkers and conjunctions:

This race was... and I knew.... because.. / I was trying to...and... but...and finished with.../ I realized that...and ...and ...because of... so...

They form long sentences with few relative pronouns to introduce subordinate clauses, while the conjunction “and” is very frequent. This is a typical feature of the spoken language.

2. Forms of nominalization

I could not detect any specific form of nominalization, apart from

*division
the swim
the run*

3. Specific language functions in the target language

The essential language functions occurring in this texts are DESCRIPTION and REPORTING.

4. Passive forms

In this kind of texts the passive form is not used. As we are dealing with race reports, the subjects are most of all people; people are more important than actions. Then, active

sentences prevail.

5. What are the most common verbal tenses to be found?

- The most common verbal tense is the SIMPLE PAST without any doubt, as we are dealing with reports, so with past events:

got a flat and crashed, broke a pedal, ran out of energy fighting a bad back, crashed, and countless others did the same / He made two minutes / he just really took off / By the time I caught him / he managed to put / had another great race / she led on the way up / repeated his fourth place finish / just ran flat out...

- The PAST CONTINUOUS tense is also very frequent but what is quite interesting is that it is nearly exclusively present in the direct speech of athletes interviewed:

I was making little mistakes / I was sweating it a little bit / I was flying downhill and staying real loose / I was riding through / I was going / I was losing time...

- There is also a discrete recurrence of CONDITIONAL forms of the verbs:

If I didn't put time on him in the swim there was no way I would have beaten him / Who would have ever thought that my weakest discipline would make the difference in a race?

They mostly recur in the athletes' direct speech.

- In a few cases, the –ing form is used:

In doing so / for winning / Following a quarter-mile run / posting the top swim split...

6. Typical compound noun phrases

two-loop bike leg
three-in-a-row

7. Some more information ...

- COMPARISON is very common in this target language:

I rode as hard as I could / The course, arguably the toughest and most scenic / the fastest of the day / two full minutes better than / one of the hardest races I've ever done / had the fastest bike split / The bike course was longer than usual / She was stronger than me / I was faster than her...

Comparatives and superlatives are mostly used while describing race results and performances.

- ORDINAL NUMBERS recur quite frequently as well, referring to placements.
- For the same reason, place PREPOSITIONS and ADVERBS recur, such as:

next / before / after / along / down / up / behind

- It is also worth observing the great number of PHRASAL VERBS which are used:

slip away / ride through / turn over / take off / go out onto / fly down / come down / come in / head into / climb up ...

Conclusions

I noticed something which I consider quite interesting, that is a language difference between what the journalist writes and the direct athletes' discourse which is reported in the articles. While referring what racers said, the language is "colourful", with a consistent number of clauses linked by the conjunction "and" and a consistent number of adjectives and phrasal verbs. In the journalists' discourse it is different: the language is plain and more standardized. Phrasal verbs can be hardly found and sentences are shorter.

The kind of vocabulary which prevails is obviously related to the triathlon but it includes frequent descriptions of places (steep climbs, frozen lake, hilly bike course, steep wooden stairs, heavy technical single track...).

While describing races, prepositions and adverbs indicating places are frequently used (after, behind, next, down, up...), together with numbers as related to results and distances.

It is necessary to pay attention to the specific vocabulary concerning race reports: phrasal verbs are used very often, qualifying adjectives and verbs of action most of all.

As far as verbal tenses are concerned, it is fundamental for students to know the past simple/past continuous, the conditional and the gerund. The simple present is a prerequisite.

Prepositions and adverbs need also to be taught, together with linkers.

The rules of comparison must be taught as well, as comparison is very common while referring to performances.

Finally, special attention must be paid to teaching reported speech.

Roberta Pittau