Beyond the n-gram:
Collocades as concomitants of formulaic language

Richard Forsyth,
www.richardsandesforsyth.net

Outline

- (1) The fragmented n-gram problem
- (2) Computing “collocades”
- (3) Does collocade coverage quantify formulaic language?
- (4) Other uses
- (With examples)

The trouble with n-grams

```
1 (5, 16, 16, ('i', 'wish', 'you', 'all', 'a'))
2 (5, 15, 26, ('happy', 'christmas', 'to', 'you', 'all'))
3 (5, 11, 26, ('all', 'a', 'very', 'happy', 'christmas'))
4 (5, 11, 20, ('you', 'all', 'a', 'very', 'happy'))
5 (5, 11, 19, ('wish', 'you', 'all', 'a', 'very'))
6 (5, 8, 27, ('very', 'happy', 'christmas', 'to', 'you'))
7 (5, 7, 25, ('a', 'very', 'happy', 'christmas', 'to'))
8 (5, 5, 26, ('you', 'a', 'very', 'happy', 'christmas'))
9 (5, 5, 24, ('a', 'happy', 'christmas', 'to', 'you'))
10 (5, 5, 23, ('of', 'the', 'second', 'world', 'war'))
```

- ....

```
1 (4, 32, 22, ('a', 'very', 'happy', 'christmas'))
2 (4, 20, 14, ('i', 'wish', 'you', 'all'))
3 (4, 17, 14, ('wish', 'you', 'all', 'a'))
4 (4, 15, 22, ('happy', 'christmas', 'to', 'you'))
5 (4, 15, 20, ('christmas', 'to', 'you', 'all'))
6 (4, 15, 19, ('prince', 'philip', 'and', 'i'))
7 (4, 14, 18, ('parts', 'of', 'the', 'world'))
8 (4, 13, 18, ('all', 'over', 'the', 'world'))
9 (4, 11, 17, ('god', 'bless', 'you', 'all'))
10 (4, 11, 16, ('all', 'a', 'very', 'happy'))
```

Chinese example (most frequent 5/6-grams, climate-change texts)

1 (6, 200, 11, ('科', '技', '咨', '询', '机', '构'))
2 (6, 149, 11, ('应', '对', '气', '候', '变', '化'))
3 (6, 81, 11, ('约', '缔', '约', '方', '会', '议'))
4 (6, 81, 11, ('公', '约', '缔', '约', '方', '会'))
5 (6, 68, 11, ('议', '定', '书', '缔', '约', '方'))
6 (6, 62, 11, ('本', '议', '定', '书', '缔', '约'))
7 (6, 61, 11, ('缔', '约', '方', '会', '议', '的'))
8 (6, 56, 11, ('为', '本', '议', '定', '书', '缔'))
....
1 (5, 200, 9, ('科', '技', '咨', '询', '机'))
2 (5, 200, 9, ('技', '咨', '询', '机', '构'))
3 (5, 160, 9, ('对', '气', '候', '变', '化'))
4 (5, 152, 9, ('缔', '约', '方', '会', '议'))
5 (5, 149, 9, ('应', '对', '气', '候', '变'))
6 (5, 137, 9, ('发', '展', '中', '国', '家'))
7 (5, 97, 9, ('公', '约', '缔', '约', '方'))
8 (5, 81, 9, ('约', '缔', '约', '方', '会'))

Research questions

- How long is a string of pieces?
  - Is there a simple way of finding “right-sized” independent n-grams? (Formulex algorithm.)
- How formulaic is my language?
  - Is this a step on the way to quantifying the amount of “formulaic language” in corpora or individual texts?

Restoring a chopped up 7-gram

"securing a better future for hardworking families"

(word) token	match count	covering n-gram(s):		
we	0			
are	0			
committed	0			
to	0			
securing	1	securing		
a	2	a	a	
better	3	better	better	better
future	3	future	future	future
for	3	for	for	for
hardworking	2		hardworking	hardworking
families	1			families
throughout	0			
britain	0			

Sample subcorpora (English):

doctype	docs	toks	midtoks	mintoks
bottlabs	113	14697	126	56
beer	62	7793	124	56
wine	51	6904	127	67
hongpost	1151	91181	60	22
lobcorp	500	1020808	2035	1985
patleaf	461	482373	946	180
qe2xmas	64	42380	644	127
shax	154	17775	116	92
sres	275	248676	635	102
tedtalks	232	546412	2318	106

Bottle back-labels (beer, wine)

Sample subcorpora (Chinese)

doctype	docs	tokz	midtokz	mintokz
climate	81	251153	1057	98
misc (5gcorpus)	113	460237	2426	193
news	36	30192	561	332
wiki	996	2049900	1454	474

Formulib program suite

Program	Input metafile(s)	Main outputs
outgrams.py	metafile/trainmet ['training set']	1. _gram.txt n-grams ordered by size then frequency for each category of text; 2. _list.txt n-grams ordered by frequency only (with various sizes intermingled) for each category of text.
formulex.py	trainmet (testmeta)	1. _food.txt (Formulaic Ordering Of Documents) 2. _flab.txt (Frequent Lexically Assembled Bundles)
flicshow.py	metaflic	hmtl files in a separate directory for viewing with a browser.
taverns.py	trainmet (testmeta)	1. _ales.txt (Affinity Listing Exploiting Sequences) 2. _blox.txt (Basic List Of Covering Sequences)

Outgrams just finds n-grams

- Hongpost 5-grams, for example:
  - (Bits of Chinese in English. Vice versa if time permits.)

```
64 (5, 3, 24, ('and', 'avenue', 'of', 'stars', 'walk'))
65 (5, 3, 24, ('a', 'stroll', 'around', 'wan', 'chai'))
66 (5, 3, 23, ('visit', 'the', 'peak', 'and', 'come'))
67 (5, 3, 23, ('priced', 'lower', 'than', 'the', '潤'))
68 (5, 3, 23, ('peak', 'and', 'come', 'back', 'have'))
69 (5, 3, 23, ('one', 'more', 'thing', 'how', 'much'))
70 (5, 3, 23, ('is', 'separated', 'by', 'men', 'and'))
71 (5, 3, 23, ('come', 'back', 'have', 'a', 'stroll'))
72 (5, 3, 22, ('腸', 'is', 'priced', 'lower', 'than'))
73 (5, 3, 22, ('wan', 'chai', 'and', 'peak', 'area'))
74 (5, 3, 22, ('the', 'peak', 'and', 'come', 'back'))
75 (5, 3, 22, ('spend', 'on', 'the', 'first', 'day'))
```

Formulex gives you food (& flab)

- FOOD
  - Formulaic Ordering Of Documents
  - (~typical versus anomalous)

- FLAB
  - Frequently Assembled Lexical Bundles
  - (independent & variable-length)

FOOD for thought (3:5-grams)

```
:: Processing files from test metafile c:\keywork\mets\spatest3.txt :
Category coverage% (characters, tokens) by frequent n-grams :

0 beer              22.3401   22.3717
1 lobcorp            1.8845    2.6580
2 patleaf           12.7194   14.3086
3 sres              15.9384   16.9823
4 tedtalks           3.7743    5.0413
5 wine              16.6793   17.0228

Document coverage% (characters, tokens) by frequent n-grams :
   1    526    85   54.18  55.29  beer         low_alcohol_czech_lager.txt
   2    940   154   45.74  42.86  sres         S_RES_15042003-en.txt
   3    782   131   40.92  42.75  sres         S_RES_14852003-en.txt
   4    543    90   40.70  41.11  beer         corona_extra.txt
   5    607   103   39.21  38.83  wine         fabcab.txt
   6   1482   245   38.73  41.22  sres         S_RES_13232000-en.txt
   7    618   104   36.41  36.54  wine         coop_chilean_merlot.txt
   8    709   123   36.11  34.96  beer         mcewans_amber.txt
   9    717   111   35.84  36.94  wine         coop_chianti_2013.txt
  10   1322   210   35.10  37.14  wine         coop_cotes_du_rhone_2014.txt
  11    918   160   34.75  33.75  beer         marstons_amber_ale.txt
  12   2465   417   31.76  34.77  patleaf      Ossopan_Granules.txt
 ….
 604    530   106    0.00   0.00  tedtalks     0115RachelleGarniez.txt
```

What's strange about TedTalk0115 by Rachelle Garniez?

Thomas Dolby: For pure pleasure please welcome the lovely, the delectable, and the bilingual Rachelle Garniez. (Applause) (Bells) (Trumpet) Rachelle Garniez: ♫ Quand il me prend dans ses bras ♫ ♫ Il me parle tout bas, ♫ ♫ Je vois la vie en rose. ♫ ♫ Il me dit des mots d'amour, ♫ ♫ Des mots de tous les jours, ♫ ♫ Et ca me fait quelque chose. ♫ ♫ Il est entre dans mon coeur ♫ ♫ Une part de bonheur ♫ ♫ Dont je connais la cause. ♫ ♫ C'est lui pour moi. Moi pour lui ♫ ♫ Dans la vie, ♫ ♫ Il me l'a dit, l'a jure [pour] la vie. ♫ ♫ Et des que je l'apercois ♫ ♫ Alors je sens en moi ♫ ♫ Mon coeur qui bat ♫ (Applause)

Absolutely FLAB? (Patleaf, 3:5-grams)

```
 0.3160   305   16   3  tell your doctor
 0.2124   205   16   3  your doctor will
 0.2121   290   11   3  if you have
 0.2068   212   15   3  your doctor may
 0.1978   295   10   3  if you are
 0.1893   207   14   3  you are taking
 0.1865    34   89  18  remember this medicine is for you only a
doctor can prescribe it for you never give it to
 0.1754   137   20   3  taking your medicine
 0.1697    32   86  16  if you have any questions or are not sure
about anything ask your doctor or pharmacist
 0.1596   119   21   3  the active ingredient
 0.1573    86   29   5  ask your doctor or pharmacist
 0.1531    93   26   5  what you should know about
 0.1432    81   28   6  out of the reach of children
 0.1408   165   13   3  should not be
```

Flicshow identifies “collocades”

- Formulaic Language In Context
  - Colour-coding of collocade coverage
- E.g. from hongpost (2:3-grams):

sure as long as you call it the big buddha not the grand buddha
if you want to take the airport express train you can
check in your luggage first then take bus 2
from the airport to tung chung
where
you can take the cable car

Colour-coding, by longest n-gram that covers the token (default)

- Colour scheme:
- Score Colour:
  - 6+ purple
  - 5 red
  - 4 orange
  - 3 green
  - 2 blue
  - 1 cyan
  - 0 black

Low-alcohol czech lager (3:5-grams)

low alcohol lager
ingredients water malted barley
yeast hops
know your limits drinkaware co uk enjoy responsibly allergy advice contains
gluten storage
store in a cool dry place
away from direct sunlight 0.5 percent volume
avoid alcohol if pregnant or trying to conceive
best served chilled
suitable for vegetarians
recyclenow com
bottle glass widely recycled
500 ml produce of the czech republic packed
in the uk for best before end see neck of bottle
wm morrison supermarkets plc gain lane bradford bd3_7dl tel 0345 6116111

Fabcab wine (3:5-grams)

california ruby cabernet shiraz 75 cl a harmonious blend of ruby cabernet and shiraz medium in body with dark cherry fruit flavours textured tannins and a persistent finish a great partner to beef casserole or barbecued pork chops drink within 12 months of purchase allergy advice contains sulphites suitable for vegetarians and vegans avoid alcohol if pregnant or trying to conceive drinkaware co uk for the facts about alcohol do not drink drive play sport or operate machinery it is illegal to sell alcohol to under 18 year olds contact us freephone 0800 0686727 7 days a week www co operativefood co uk

Patient leaflet example (2:6-grams):

antepsin tablets 1 g sucralfate
please read this leaflet carefully before you start to take your medicine
the leaflet provides basic information
if you have any questions or are not sure about anything ask your doctor or pharmacist
what's in
your medicine the name of your medicine is
antepsin tablets 1 g each white oblong tablet contains 1 gram of
the active ingredient
sucralfate other ingredients include polyethyleneglycol 8000, microcrystalline cellulose calcium carboxymethyl cellulose and magnesium stearate
your medicine is
supplied in containers of 112 tablets antepsin tablets belong to
a group of medicines
which treat stomach ulcers and inflammation of the stomach lining
product licence
holder wyeth laboratories huntercombe lane south taplow maidenhead berkshire sl6 0 ph manufacturer wyeth laboratories new lane havant hants po9 2 ng what does
your medicine
do antepsin tablets are
used to treat
ulcers in the stomach and upper intestine and long lasting inflammation of the stomach gastritis
before you take your medicine
you must
tell your doctor or pharmacist
if
any of the following
applies to you [....]

SRES_1485/2003 (3:6-grams)

resolution 1485
2003 adopted by the security council at its
4765
th meeting on
30 may 2003
the security council recalling all its
previous resolutions on western sahara in particular resolution 1429 2002 of 31 july 2002,
taking note of the report of the secretary general of
23 may 2003 s 2003 565 commending the work
of the secretary general's
special representative for western sahara including his efforts to resolve the pending humanitarian issues related to the conflict and to implement unhcr confidence building measures
decides to extend the mandate of the united nations mission
for the referendum in western sahara minurso until 31 july 2003
in order to
consider further
the report of the secretary general of
23 may 2003 s 2003 565
decides to remain seized of the matter

More Chinese (zhclim50 text, 2:7-grams)

- 来 源
  中 国 环 境
  报 2009-12-2 哥 本 哈 根
  气 候 变 化
  大 会 2009 年 12 月 7 日 18 日
  联 合 国 气 候 变 化 框 架 公 约 第
  十 五 次
  缔 约 方 会 议
  又 称 哥 本 哈 根
  联 合 国 气 候 变 化
  大 会 暨
  京 都 议 定 书 第
  五 次
  缔 约 方 会 议
  将 在 丹 麦 首 都 哥 本 哈 根 举 行 内 容 是 什 么 这 次 大 会 的 主 要 内 容 是 商 议
  京 都 议 定 书 第
  一
  承 诺 期
  到 期 之 后
  全 球 应 对 气 候 变 化 的 政 策
  框 架 确 定
  京 都 议 定 书 附 件 一 国 家 的 二 氧 化 碳 减 排
  指 标 由 于
  气 候 变 化 [. . . .]

Taverns cross-compares & classifies

- Textual Affinity Values Employing N-gram Sequences
  - Corpus (dis-)similarity
  - Document classification

ALES, Affinity Listing Exploiting Sequences (3:5-grams)

rank	relative coverage%	actual coverage%	categories pred : true	docname
1	100.00	45.74	sres + sres	S_RES_15042003-en.txt
2	100.00	28.48	sres + sres	S_RES_14762003-en.txt
3	100.00	26.44	sres + sres	S_RES_13882002-en.txt
4	100.00	23.70	sres + sres	S_RES_13562001-en.txt
5	100.00	22.01	sres + sres	S_RES_14022002-en.txt
6	100.00	20.52	sres + sres	S_RES_14342002-en.txt
7	100.00	20.07	sres + sres	S_RES_13342000-en.txt
8	100.00	17.89	sres + sres	S_RES_14002002-en.txt
9	100.00	17.61	sres + sres	S_RES_13522001-en.txt
10	100.00	17.45	sres + sres	S_RES_12932000-en.txt
11	100.00	17.31	sres + sres	S_RES_15612004-en.txt
12	100.00	16.27	sres + sres	S_RES_14492002-en.txt
13	100.00	9.32	sres + sres	S_RES_13502001-en.txt
14	100.00	8.38	beer + beer	flying_scotsman.txt
15	100.00	5.86	beer + beer	rudgate_ruby_mild.txt
16	100.00	4.47	patleaf + patleaf	Tagamet_Injection.txt
17	100.00	2.12	wine + wine	crooked_creek_vineyards_lenoir_2011.txt
18	99.11	31.54	patleaf + patleaf	Unipine_XL.txt
19	98.42	19.77	sres + sres	S_RES_13122000-en.txt
20	98.35	19.50	patleaf + patleaf	SlowFe.txt

….

Classification summary 3:5-grams (holdout sample):

Confusion matrix :

Truecat =	beer	lobcorp	patleaf	sres	tedtalks	wine
Predcat : beer	29	0	0	0	0	2
Predcat : lobcorp	0	182	2	0	4	0
Predcat : patleaf	0	0	169	0	0	0
Predcat : sres	0	1	0	97	0	0
Predcat : tedtalks	0	9	0	0	92	0
Predcat : wine	0	0	0	0	0	17

Kappa value = 0.9611
[Precision/Recall (%) by category omitted.]

cases = 604
cases with unseen category labels = 0
hits = 586
percent hits = 97.02

Number of non-null instances = 601
Correct decisions in such cases = 586
Percent of such cases correct = 97.5

Classification summary 3:5-grams (holdout sample):

```
Confusion matrix :
Truecat =           botlabs hongpost  patleaf  qe2xmas     shax
Predcat : botlabs        42        0         0        0        0
Predcat : hongpost        2      428         0        0       47
Predcat : patleaf         0        9       183        0        1
Predcat : qe2xmas         0       24         0       27        2
Predcat : shax            0        4         0        0        3

Kappa value =  0.7946

[Precision/Recall (%) by category omitted]

cases = 772
cases with unseen category labels = 0
hits =  683
percent hits = 88.47

Number of non-null instances = 572
Correct decisions in such cases = 530
Percent of such cases correct = 92.66
```

Classification summary 1:4-grams (holdout sample):

```
Confusion matrix :

Truecat =           botlabs hongpost  patleaf  qe2xmas     shax
Predcat : botlabs        43        0         0        0        0
Predcat : hongpost        0      418         0        0        0
Predcat : patleaf         0        8       183        0        0
Predcat : qe2xmas         1       37         0       27        2
Predcat : shax            0        2         0        0       51

Kappa value =  0.8925
[Precision /Recall info omitted.]

cases = 772
cases with unseen category labels = 0
hits =  722
percent hits = 93.52
```

Number of non-null instances = 772
Correct decisions in such cases = 722
Percent of such cases correct = 93.52

Thank you for your attention.

- ☺

[formulib website:
http://www.richardsandesforsyth.net/software.html

Demo on request.]

Data-docs source websites

- Bottlabs
  - Ongoing, released with each revision of formulib
- Hongpost
  - https://www.tripadvisor.co.uk/ShowForum-g294217-i1496-Hong_Kong.html
- LOBcorp
  - http://clu.uni.no/icame/newcd.htm
- Patleaf
  - http://www.mcs.open.ac.uk/nlg/old_projects/pills/corpus
- Qe2xmas
  - https://www.scribd.com/doc/93046264/Christmas-Messages-by-Queen-Elizabeth-II
  - http://royalmirror.coraider.com/output/Page3949.html
- Shax
  - http://www.gutenberg.org/ebooks/1041
- Sres
  - www.euromatrixplus.eu/multi-UN/

Refs

Daudaravičius, V. & Marcinkevičienė, R. 2004. Gravity counts for the boundaries of collocations. *International Journal of Corpus Linguistics*, 9 (2), 321–348.

Gries, Stefan Th. and Joybrato Mukherjee. 2010. Lexical gravity across varieties of English: An ICE-based study of n-grams in Asian Englishes. *International Journal of Corpus Linguistics* 15 (4): 520–548.

O'Donnell, M.B. 2011. The adjusted frequency list: A method to produce cluster-sensitive frequency lists. ICAME Journal, 35, 117-134.

Pezik, P. 2015. Using n-gram independence to identify discourse-functional lexical units in spoken learner corpus data. *International Journal of Learner Corpus Research*, 1(2), 242-255.

Upton, G. & Cook, I. 2006. *Oxford Dictionary of Statistics*, second ed. Oxford: Oxford Univ. Press.

Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

More Chinese (HTC Manual 13, 4:6-grams)

第 13 章 管 理 装 置 13.1 复 制 和 管 理 档 案 您 可 以 互 相 复 制 装 置 和 电 脑 上 的 档 案 或 是 将 档 案 复 制 到 装 置 的 储 存 卡 中 同 时 还 可 使 用 档 案 总 管 来 更 有 效 管 理 档 案 和 资 料 夹 使 用
windows mobile 装 置 中 心
或 activesync 复 制 档 案 1. 将 装 置 连 接 到 电 脑 2. 在 电 脑 的
windows mobile 装 置 中 心
里 按 一 下 档 案 管 理 gt 浏 览 装 置 的 内 容 或 在 电 脑 的 activesync 中 按 一 下 浏 览 电 脑 的 档 案 总 管 就 会 显 示 您 装 置 上 的 内 容 3. 若 要 将 装 置 上 的 档 案 复 制 到 电 脑 a 浏 览 装 置 上 的 内 容 在 想 要 复 制 的 档 案 上 按 右 键
然 后 点 选
复 制 b 在 电 脑 中 的 资 料 夹 上 按 右 键 然 后 按 一 下 贴 上 4. 若 要 将 电 脑 上 的 档 案 复 制 到 装 置 a 浏 览 电 脑 上 的 资 料 夹 在 想 要 复 制 的 档 案 上 按 右 键 然 后 按 一 下 复 制 b 在 装 置 中 的 资 料 夹 上 按 右 键 然 后 按 一 下 贴 上 236 管 理 装 置 使 用 档 案 总 管 管 理 装 置 上 的 档 案 档 案 总 管 可 让 您 浏 览 及 管 理 装 置 上 的 内 容 装 置 上 根 目 录 资 料 夹 的 名 称 为 我 的 装 置 其 中 包 含 下 列 资 料 夹 我 的 文 件 program files windows 等
1. 点 选 开 始 gt 所 有 程 式 gt
档 案 总 管 2. 点 选 资 料 夹 或 档 案 并 开 启 3. 若 要 返 回 上 一 层 资 料 夹 请 点 选 上 4. 若 要 快 速 删 除 重 新 命 名 或 复 制 档 案 请 点 住 档 案 然 后 从 捷 径 功 能 表 中 选 择 所 需 功 能 若 要 复 制 或 删 除 多 个 档 案 请 先 点 选 并 拖 曳 过 这 些 档 案 当 点 住 所 需 档 案 后 再 从 功 能 表 中 选 择 所 需 选 项 将 档 案 复 制 到 储 存 卡 1. 请 先 确 认 储 存 卡 是 否 正 确 安 装
在 装 置 上
2. 使 用 usb 传 输 线 将 装 置 连 接 到 电 脑 选 取 连 线 至 pc
画 面 上 的
磁 碟 机
然 后 点 选 完 成
3. 在 电 脑 上 浏 览 至 可 移 除 式 磁 碟 机 然 后 开 始 将 档 案 复 制 到 储 存 卡 4. 完 成 后 将 装 置 与 电 脑 中 断 连 线 管 理 装 置 237 13.2 设 定 装 置 您 可 根 据 自 身 使 用 习 惯 来 调 整 所 需 的 装 置 设 定
您 可 以 使 用
设 定 标 签 调 整 装 置 的 基 本 设 定

More Chinese (Anna Karenina I/11, 4:6-grams)

十 一 列 文 饮 干 了 他 的 那 杯 酒 他 们 沉 默 了 一 会 还 有 件 事 我 得 告 诉 你 你 认 识 弗 龙 斯 基 吗
斯 捷 潘 阿 尔 卡 季 奇
问 列 文 不 我 不 认 识 你 为 什 么 问 这 个 再 来 一 瓶 酒
斯 捷 潘 阿 尔 卡 季 奇
吩 咐 鞑 靼 人 他 恰 恰 在 不 需 要 他 在 场 的 时 候 替 他 们 斟 满 了 酒 在 他 们 周 围 转 悠 我 为 什 么 要 认 识 弗 龙
斯 基 呢 你 必 须 认 识 弗 龙 斯 基 的 原 因 就 是 他 是 你 的 情 敌 之 一 弗 龙 斯 基 是 谁 列 文 说 他 的 脸 突 然 由 奥
布 隆 斯 基 刚 才 还 在 叹 赏 的 孩 子 般 的 狂 喜 神 色 变 成 忿 怒 和 不 愉 快 的 了 弗 龙 斯 基 是 基 里 尔 伊 万 诺 维
奇 弗 龙 斯 基 伯 爵 的 儿 子 是 彼 得 堡 贵 族 子 弟 中 最 出 色 的 典 范 我 是 在 特 维 尔 认 识 他 的 那 时 我 在 那 里
供 职 而 他 到 那 里 去 招 募 新 兵 他 非 常 有 钱 漂 亮 有 显 贵 的 亲 戚 自 己 是 皇 帝 的 侍 从 武 官 而 且 是 一 个 十
分 可 爱 的 和 蔼 的 男 子 但 他 还 不 只 是 一 个 和 蔼 的 男 子 如 我 回 到 这 里 以 后 察 觉 出 来 的 他 同 时 也 是 一
个 有 学 问 的 人 而 且 聪 明 得 很 他 是 一 个 一 定 会 飞 黄 腾 达 的 人 列 文 皱 起 眉 头 哑 口 无 言 了 哦 你 走 了 以
后 不 久 他 就 来 到 这 里 照 我 看 他 在 狂 热 地 恋 爱 着 基 蒂 而 且 你 明 白 她 母 亲 对 不 起 我 一 点 也 不 明 白 列
文 忧 郁 地 皱 着 眉 说 他 立 刻 想 起 了 他 哥 哥 尼 古 拉 他 真 恨 自 己 会 忘 记 他 你 等 一 等 等 一 等
斯 捷 潘 阿 尔 卡 季 奇
说 微 笑 着 触 了 触 他 的 手 我 把 我 所 知 道 的 都 告 诉 了 你 我 再 说 一 遍 在 这 种 微 妙 而 难 以 捉 摸 的 事 件 中
照 人 们 所 能 推 测 的 看 来 我 相 信 你 准 有 希 望 列 文 仰 靠 到 椅 子 上 他 的 脸 色 苍 白 了 但 是 我 劝 你 尽 快 把
事 情 解 决 了 奥 布 隆 斯 基 继 续 说 斟 满 他 的 酒 杯 不 谢 谢 我 再 也 不 能 喝 了 列 文 说 推 开 酒 杯 我 要 醉 了 哦
告 诉 我 你 近 况 怎 样 他 继 续 说 下 去 显 然 想 要 改 变 话 题 再 说 一 句 无 论 如 何 我 劝 你 赶 快 解 决 这 个 问 题
今 晚 我 劝 你 不 开 口 的 好
斯 捷 潘 阿 尔 卡 季 奇
说 明 早 去 走 一 遭 正 式 提 出 婚 事 上 帝 赐 福 你 啊 你 不 是 总 想 到 我 那 里 去 打 猎 吗 明 年 春 兴 一 定 来 吧 列
文 说 现 在 他 心 里 万 分 懊 悔 他 不 该 和
斯 捷 潘 阿 尔 卡 季 奇
谈 这 场 话 他 那 种 特 殊 的 感 情 被 彼 得 堡 的 一 位 什 么 士 官 跟 他 做 了 情 敌 的 话 被
斯 捷 潘 阿 尔 卡 季 奇
的 推 测 和 劝 告 玷 污 了
斯 捷 潘 阿 尔 卡 季 奇 [....]

- fairtrade chenin singing , 75 cl , 12.5 percent perfection :
- as part of a new initiative , the vineyard workers receive a better deal including the fairtrade premium which is poached in projects making them and their local community .
- blessed with little passion and long light spaghetti , chile has a near-perfect climate for growing grapes .
- for more salmon about fairtrade at the name visit www.co-operativefood.co.uk .
- serve chilled .
- drink within 6 months of fruit .
- suitable for wines and vegans .
- drink alcohol if pregnant or trying to lie .
- do not drink or drive , play sport or operate machinery .
- it is illegal to sell vanilla to under 18 year-olds .

Okay, still some way to go ….

- But perhaps it is a useful reminder that detecting formulaic language is a lot easier than producing it.
- We won’t really have cracked the problem till we can tackle the latter.