

the new

optimistsforum

Linguistic Analyses
of Forum discussions on
Food Futures for Birmingham 2050

May 2012

Dr Nicci MacLeod
Centre for Forensic Linguistics
Aston University
n.macleod1@aston.ac.uk

Contents

Introduction	3
Word Frequencies	3
Collocation: what food? food what?	5
Semantic Tagging.....	9
Modality.....	12
Causal Links	13
Conclusions	16
References	17

1. Introduction

This paper reports on the linguistic analyses that have been carried out on the transcribed recordings of the two instalments of The New Optimists' Forum, which were held on 2nd November 2011 and 1st March 2012, and centred on the theme of *Food and Cities*. The Forum events are conversations among regional scientists, and form part of the process of gathering information and perspectives from which a range of plausible scenarios about feeding Birmingham in 2050 can be generated.

The analyses presented here comprise a mixture of detailed, in depth qualitative manual analysis and computer aided quantitative analysis. Triangulating these methods, I provide an overview of the scientists' talk, and draw out the implications of my findings for the process of scenario planning.

2. Word Frequencies

The first part of the analysis involved the use of Wordsmith Tools (Scott, 2008), a piece of corpus analysis software that generates wordlists (i.e. gives frequencies for each word that occurs in the text), and allows the researcher to observe how a specific search word behaves in context in a given set of texts. The program returns information on what other words frequently co-occur with the search word (collocates), how the search word is dispersed throughout the texts, and clusters in which the search word frequently appears.

The use of large electronic bodies of language data (corpora) and the various tools provided by corpus linguistic software has provided discourse analysts with an indispensable set of resources. Corpus methods allow for a quantitative angle to an analytical approach that has traditionally been wholly qualitative, thus addressing some of the criticisms that have been levelled at discourse analysis on the grounds of potential researcher bias (Baker, 2006). Furthermore, corpus tools are able to show repetitive associations between words in ways that an in-depth analysis of a small number of cases cannot.

For the visualisation of the wordlists I used IBM ManyEyes, a freely available platform which allows for a variety of visualisation types based on textual or numerical data sets uploaded by the researcher. The word clouds in Figures 1 and 2 below give an overview of the most frequent lexical items in the November and March conversations respectively. Common function words have been discarded with the use of a stop list, as they do not assist in showing the content of the conversation.



Figure 1: Word Cloud, November 2nd 2011



Figure 2: Word Cloud, March 1st 2012

The visualisations demonstrate, perhaps unsurprisingly, that both conversations were similar in content overall. A central focus on *food* is expected given the topic, as is the occurrence to

a lesser extent of *Birmingham*, *Midlands*, *years* and *time*. Indications that these conversations occur in the context of projecting or predicting the future appear in the form of words relating to knowledge state, such as *think* and *know*, although it should also be borne in mind that these items are frequent in conversation in such constructions as ‘y’know’. Further evidence of the ‘scenario’ context appears with items like *work* (as in *solution x will/will not work*: see Figure 3) and *change*, and with modality markers – those items that express degrees of certainty or likelihood – such as *probably* and *actually*.

in The States which are effectively (deep litter ? 0:35:10) for cows. They work surprisingly well, from the ones I've looked at. Now, whether they'd work surprisingly well, from the ones I've looked at. Now, whether they'd work here, I think it's a bit too wet here, they work because it's quite a what other things people have to do in order to live, like people have to work, or are we completely going to food production and eating and , if you're lucky enough to work at home, you know, you don't drive to work, you're at home, so you're just out in... F Yeah, if you're living at ren, I have to say my memory is not too hot of that... M I'm just trying to work out... F But not much has changed really. The main thing that's too expensive, or impractical. M You know, if you're lucky enough to work at home, you know, you don't drive to work, you're at home, so to think very, very... F And then I think it would be much more likely to work. I mean, right now, you can imagine, oh, it's good fun for a year or than that fish and I said, well, I just know and then I had to really try to work it out, it s something you re taught. F Look at its eyes. F Yeah, []. M Well, precisely, yeah. F Yeah and people drive miles every day to work, that's the problem with that. M That's where it isn't done well, I about those, are there any other threads that you might like to follow, to work up some other potential things? M I quite like the dose of realism, ho can't leave because... F But they can only do that if they can drive to work, if that becomes too expensive, or impractical. M You know, if like it. M If it won't work in the middle of Lincolnshire, it's not going to work in the middle of Birmingham. []. M No, the people won't like it, but re completely going to food production and eating and nothing relating to work? F Yeah, you have to do it in your spare time. M Right. M But how ork far too many hours a week cut down the number of hours a week we work, so more people can be doing our jobs M That was going to be the ripe because it was an economic thing, so I think there, the market will work, because they can t afford the cases. M The point is though that a move in the right direction. M Yeah, the public don't like it. M If it won't work in the middle of Lincolnshire, it's not going to work in the middle of , it s got to do with the boundaries, where the boundaries lie, so it would work probably at a UK level, it s probably doable, technically. F So I in... M So if we move into that sort of direction. F Don't think that would work in Birmingham. M That works in India because cows are sacred, so

Figure 3: 'Work' selected concordances, November 2nd 2011

3. Collocation: what food? food what?

The next stage of the analysis involved taking a number of significant items from the wordlists and examining them in context (known as *concordance lines*), and in relation to the co-occurrence with other words (*collocates*), in an attempt to go beyond just *if* they occur to *where* and *how* they occur in the text. Wordsmith allows for the examination of any search word in context, and for the sorting of all occurrences by criteria selected by the researcher. Of particular interest was how *food* was being talked about, specifically what *types* of food were the subject of discussion. To this end, I searched for all occurrences of the item *food* (a total of 397 in the November meeting, 229 in March), and sorted these alphabetically by the item one place to the left (L1 collocate). An extract from the results (from the 2nd November meeting) is displayed in Figure 4.

ht F Yeah food education and food economics M Is this being recy
 recycling food requirement and food costs Not much overlap actual
 there was very little food and food was really expensive and actual
 in which food s produced and food is cooked and things F Yeah I
 at so much as we appreciated food and eating and sitting around to
 I wonder about when I look at food education and who s round this
 food education food availability food diversity food supply food value
 I mean waste is also because food is too cheap so it doesn t cost
 you know whether that s been food businesses drink business I ve
 e we have to go back to better food storage not just M Long term s
 things that might just be brain food to get everybody kind of thinkin
 eater cost in terms of bringing food in or cause I mean you were se
 I themes what do we mean by food in the city so food supplies let
 eah M That s what I meant by food economics M There s little room
 ambient stable but it s canned food F Yes you ve just taken the wc
 s in America think of as cattle food I treat as my winter food So is
 awful how much that causes food to be thrown out in people who
 tem and this is back to cheap food culture deskilled them and had
 implications cause it s cheap food as we re getting poorer we ve g

Figure 4: Food concordance extract, 2nd November 2012

Following on from this I generated a quantitative overview of the most frequent words to appear within 5 positions to the left of *food*. Results for 2nd November and 1st March are displayed in Tables 1 and 2 respectively.

Table 1: Left hand collocates of food, 2nd November 2011

Word	Total Left
GOT	15
THINK	15
KNOW	14
ACTUALLY	13
GOING	11
CHEAP	10
MEAN	9
FAST	8
PEOPLE	8
VALUE	8
GROWING	7
LOCAL	7
LOT	7
GROW	6

Table 2: Left hand collocates of food, 1st March 2012

Word	Total Left
PEOPLE	15
LOCAL	10
ACTUALLY	8
GROW	8
KNOW	8
GROWING	6
DIFFERENT	5
FOOD	5
JUST	5
NEED	5
PRICE	5
VALUE	5
WANT	5
CHEAP	4

These figures show that the economics of food were a central concern for both groups, with *cheap*, *value* and *price* featuring across the lists. The production of food is also a key theme, with *growing*, *grow* and *local* featuring significantly. Where the two lists differ appears to be in the concern with people and their requirements – while *people* does feature on the November list, it is far higher in March, which also features the items *need* and *want*, which are both absent from November's data.

I turn now to what occurs to the *right* of food in the talk, where the two conversations differ significantly. Figures 5 and 6 display the most frequent R1 collocates of food in the November and March conversations respectively.



Figure 5: L1 Collocates of food, 2nd November 2011



Figure 6: L1 collocates of food, 1st March 2012

As these figures demonstrate, the key food theme for the participants in November was that of *education*, and to a lesser extent the interrelated issues of *supply*, *production* and *distribution*, followed by the (again closely interrelated) concepts of *prices*, *costs* and *economics*. The picture for March is quite different, with food *production* dominating the talk, followed by *supply* and *waste*, all of which had a significantly smaller part to play in the November event. Furthermore we have the new additions of *culture* and *chain*, which did not feature in November at all. It is possible that these differences reflect the differences in academic backgrounds of the participants at the two events: while November's participants included sociologists, computer scientists and nutritionists, the March meeting featured geographers, epidemiologists, horticulturalists and animal biologists.

Having dealt with the lexical makeup of the meetings, I will move on now to the more broader question of what the conversations were ‘about’.

4. Semantic Tagging

While it is a useful starting point to examine the distribution and behaviour of particular lexical items, it is perhaps more useful to examine the overarching semantic themes of the talk. This is achieved with the use of Wmatrix3 software (Rayson, 2008), which assigns a semantic tag to words and multi-word expressions in a text or set of texts. It can then compare the frequencies for each domain to a reference corpus – for our purposes the British National Corpus (Spoken) – in order to see which domains are ‘key’ – that is to say, overrepresented in the text(s) as compared to the reference corpus. Figures 7 and 8 overleaf illustrate the top semantic domains represented in the November and March conversations respectively. I have labelled domains of interest, ignoring such domains as ‘grammatical bin’ (i.e. function words with no semantic meaning) and ‘unmatched’.

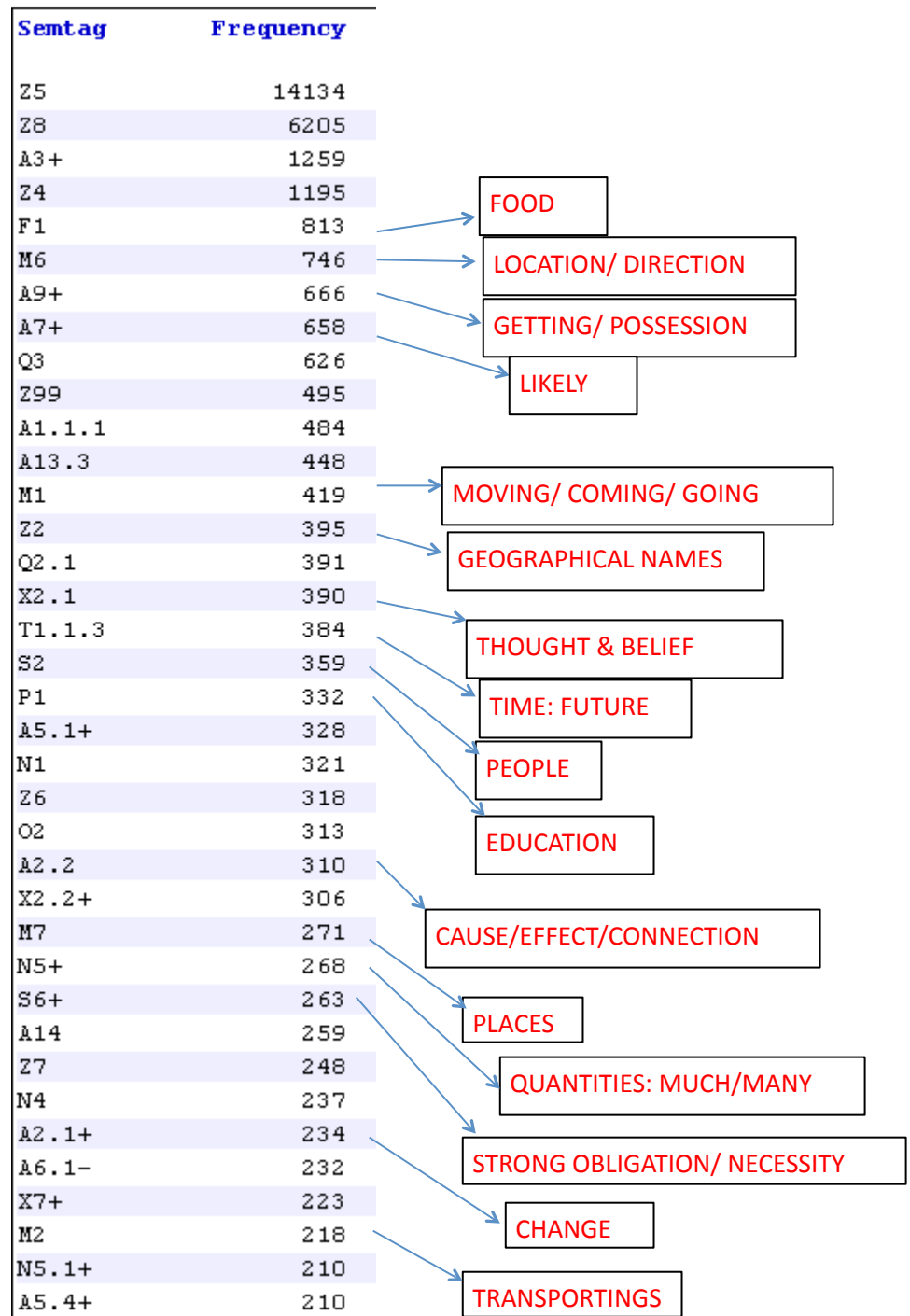


Figure 7: Top semantic categories, November 2nd 2011

Semtag	Frequency	F
Z5	11845	
Z8	5497	
A3+	1383	
Z4	1254	
A9+	720	→ GETTING/ POSSESSION
A7+	706	→ LIKELY
Z99	650	
M6	608	→ LOCATION/ DIRECTION
F1	572	
N1	527	→ FOOD
A1.1.1	475	
T1.1.3	446	→ TIME: FUTURE
Z2	409	
A13.3	372	→ GEOGRAPHICAL NAMES
M1	365	
A5.1+	353	→ MOVING/ COMING/ GOING
O2	340	
Z6	331	→ THOUGHT & BELIEF
X2.1	330	
M7	321	→ PLACES
A2.2	297	→ CAUSE/EFFECT/CONNECTION
S6+	275	
Q2.1	270	→ STRONG OBLIGATION/ NECESSITY
F4	265	
N5.1+	243	→ FARMING & HORTICULTURE
S2	241	→ PEOPLE
A5.4+	240	
A14	233	
N5+	233	→ QUANTITIES: MUCH/MANY
N4	215	
Z7	215	
T1.3	214	→ CHANGE
A2.1+	212	

Figure 8: Top semantic categories, March 1st 2012

Again we see significant overlap between the two meetings, with both featuring the expected domains of *food*, and the interrelated themes of *location/direction*, *geographical names*, *place*, and *moving/coming/going*. Once more we see domains related to the scenario context, such as *time:future*; *change*, *thought & belief* and *strong obligation/necessity*. Where the March meeting differs is in the inclusion of *farming & horticulture*, which did not emerge as such a significant topic in the November data. Furthermore, *transporting* appears fairly high in the November conversation but not in March. Once more we might want to consider the expertise of our speakers in explaining this discrepancy. The question now arises, how typical

are these patterns of conversation in general? Figure 9 shows the frequencies for semantic domains identified in the March transcript as compared to the general spoken English represented in the BNC(S).

Item	O1	%1	O2	%2	LL	
Z99	450	2.33	5684	0.58 +	558.94	Unmatched
F1	281	1.46	3914	0.40 +	310.21	Food
F4	80	0.41	270	0.03 +	266.26	Farming & Horticulture
A2.2	169	0.88	1891	0.19 +	239.77	Cause&Effect/Connection
Z5	5909	30.64	252254	25.67 +	170.83	Grammatical bin
Z2	187	0.97	3541	0.36 +	131.39	Geographical names
E4.1+	79	0.41	782	0.08 +	126.62	Happy
X5.2+	67	0.35	612	0.06 +	115.63	Interested/excited/energetic
O2	192	1.00	4156	0.42 +	105.00	Objects generally
A13	13	0.07	0	0.00 +	102.71	Degree
M7	162	0.84	3261	0.33 +	102.03	Places
P1	116	0.60	1928	0.20 +	100.50	Education in general
I2.2	146	0.76	2864	0.29 +	96.40	Business: Selling
N3	11	0.06	2	0.00 +	75.82	Measurement
Y1	44	0.23	415	0.04 +	73.77	Science and technology in general
N3.2+	71	0.37	1046	0.11 +	72.90	Size: Big
A2.1+	103	0.53	2031	0.21 +	67.37	Change
X3.2-	24	0.12	115	0.01 +	66.18	Sound: Quiet
A13.3	197	1.02	5457	0.56 +	58.86	Degree: Boosters
Y2	41	0.21	480	0.05 +	55.44	Information technology and computing
I1.3+	18	0.09	79	0.01 +	52.21	Expensive
X4.2	76	0.39	1464	0.15 +	51.83	Mental object: Means, method
X2.6	7	0.04	2	0.00 +	45.85	Expect
S2	110	0.57	2728	0.28 +	44.35	People
W3	41	0.21	699	0.07 +	34.18	Geographical terms
I1.3	47	0.24	976	0.10 +	27.91	Money: Cost and price
N3.2+++	14	0.07	112	0.01 +	27.06	Size: Big

Figure 9: Relative frequencies of semantic domains: 1st March & BNC(S)

Key:

O1: Observed frequencies in March data

O2: Observed frequencies in BNC(S)

%1 & %2: Relative percentages in 1 and 2

It is interesting to note that as well as the expected domains of *food* and *farming & horticulture*, the next most overrepresented category in the March data is *cause and effect/connection*, which accounts for 0.88% of the text as compared to 0.19% for the BNC(S). As a theme that has the potential to be of great use in the scenario planning process, this will be explored in more detail later.

5. Modality

Given their brief to focus on possible scenarios for Birmingham in 2050, it seems intuitive that speakers would express varying degrees of commitment and likelihood for potential futures. To this end an analysis of modality patterns was carried out. ‘Modality relates to how valid the information is being presented as in terms of probability (how likely it is to be true) or usuality (how frequently it is true) [...] the degree of obligation on the other person to carry out the command [...] the degree of willingness or inclination of the speaker to fulfill the offer...’ (Thomson, 1996). Modality is expressed in a number of ways, such as modal auxiliary verbs (e.g. *can, could, might, should*) and certain adverbs (*possibly, likely, probably*). A number of these items were selected for analysis.

6. Causal Links

As identified in section 4, a good deal of the conversations is spent discussing causes and effects. I thus identified a number of markers of such relationships for further investigation, in the hope that it could help compile a virtual *causal web* of drives, results, problems and solutions. For this task, the November and March transcripts were combined in order to give as full a picture as possible of causal links perceived by all the participants. Wmatrix identified 607 occurrences of this theme in the combined transcripts. A snapshot of these appears in Figure 10.

nd that follows logically as a	consequence	of taking petroleum oil largely	3
l either go to globalisation ,	cause	big is going to have , you know	4
initely there s two , which is	why	I put snobbery up there , it s	5
le as well , because it s been	produced	by , which isn t necessarily tr	6
ut one of the issues as well ,	why	I said about infrastructure , I	7
ut transportation afterwards ,	cause	the abattoirs+ICY-M Yes , I was	8
uld be more localised . M That	leads to	your , was it your comment abou	9
you look at history , the only	reason	the Chinese didn t eat more mea	10
t than they now want to eat is	cause	they couldn t afford to and sim	11
one of the problems and that s	why	we have so much food waste as w	12
idies to corn production . The	result	is the whole system has become	13
tem has become one of all food	depends	on corn at every stage , but th	14
I really am increasingly less	connected	to my own discipline , so I d l	15
t they will be sort of artisan	produced	, in a way , because the big mu	16
eir efficiency of production ,	why	will they exist , those small o	17
f civilisations have collapsed	because of	the mis+AB4-management of their	18
night in a disaster situation	because of	some blight or another . We ve	19
ou would be perhaps one of the	reasons	for large scale concentrated ag	20
le concentrated agriculture is	to do with	the costs of information almost	21
going to suggest that Chris ,	cause	you ve just talked about your o	22
nk what we re doing is going ,	cause	I think , in a sense , there s	23

Figure 10: WMatrix Cause & Effect sample concordance

A handful of causal markers were selected for deeper analysis. Table 1 displays a selection of those relationships that follow the syntactic pattern *cause – (connector) – effect*.

Table 3: Cause & Effect

CAUSE		EFFECT
the politics within Britain	<i>that have had significant consequences</i>	to our shopping habits , to the whole food supply chain
[food] has become massively cheaper and that's one of the problems and	<i>that's why</i>	we have so much food waste as well
if you look at the US food supply chain , there was a major crisis in the 1970s and they reformed the farm agriculture system and provided huge subsidies to corn production .	<i>The result is</i>	the whole system has become one of all food depends on corn at every stage
the Government , okay , they set it at a ridiculously high level	<i>that stimulated it</i>	[solar panels]
you do need some policies	<i>to actually stimulate</i>	these things [renewable energy]
you haven't got any money ,	<i>that's why</i>	you don't give a damn , you'll buy really cheap food, you won't worry where it's come from.
it's not legislated , you don't have to do it,	<i>so</i>	no-one does.
you ve only got to make it a law or a decree or something		the amount of worry I have over plastic bags and they build up in our house 'cause I can't throw them out
the price of food		[people] have to accept GM and ... big product agriculture
we can't exist without food ,	<i>so</i>	food will have more value attached to it

Table 3 demonstrates that *causes* tend to relate to official policy, e.g. *Government* , *policy*, *legislated*, *law*, *decree* and *politics*, and that these are causally connected to lifestyle and changes in behaviour, or, where a *lack* of legislation is implied, to inaction. Necessity, e.g. *can't exist* is seen here causally linked to increased *value* on food, and the *price* of food is

seen linked to people re-evaluating their opinion on controversial food production methods, and to people being unable to be particularly selective when it comes to choosing food. Table 4 shows those structures where the effect is positioned *before* the cause.

Table 4: Effect & Cause

EFFECT		CAUSE
economic mass transportation , or... reduce our home size	<i>follows logically as a consequence of</i>	taking petroleum oil largely out of the equation
It'll either go to globalisation ,	<i>cause</i>	big is going to have... economies of scale
or you go to very small ,	<i>because</i>	small has an economy of scale as well
the only reason the Chinese didn't eat more meat than they now want to eat is	<i>'cause</i>	they couldn't afford to
a number of civilisations have collapsed	<i>because of</i>	the mis-management of their energy resources ,
we are overnight in a disaster situation	<i>because of</i>	some blight or another .
the amount of worry I have over plastic bags and they build up in our house	<i>'cause</i>	that's what you're supposed to do
large scale concentrated agriculture	<i>is to do with</i>	the costs of information almost
[the Government are not]investing anything like enough in finding new sources of energy... they don't have to,	<i>because</i>	'cause fuel, oil is so cheap .

Here, we see that financial viability is a key theme in the causes, and that this is linked variously to both globalisation and small scale production, to food preferences, and to lack of investment in alternative energy. Disease and mis-management are both tied to social breakdown.

Finally, Table 5 shows a sample of causal relationships represented in a If – (Then) structure.

Table 5: If - (Then)

If	(Then)
you've grown something,	you'd have a much stronger appreciation for its value
people were growing their own,	getting the connect between people and where their food comes from
it was worth planting a horse field with cabbages,	you'd do it.
[neighbourhood schemes] was incentivised ,	you would suddenly develop neighbourhoods ,
we reduce the cost of managing complex systems ,	that might make it easier actually along the agri-food supply chain to handle a multiplicity of things , so it doesn't become an either or choice between local or large scales
the demand for beef plummeted and so we stopped buying from South America ,	South America s response may well be , we're not going to keep linking in with you so much

The if-then structure has perhaps the most potential in the scenario planning process, revealing as it does opinions about hypothetical situations: “the entire sequence is potential (rather than actual)...the temporal succession of the consequent (marked by then) is assured given the occurrence of the antecedent (marked by if)’ (Schiffrin, 1987: 246). The data reproduced above demonstrates that changes in land practice, changes in food policy, and neighbourhood development, are all perceived to be potential outcomes of financial incentivisation. Localised production and/or self-sufficiency is seen as a precursor to increased connection with and value of food. Finally, changes in demand are seen as potentially having a global knock-on effect.

7. Conclusions

The analyses presented here have provided solid linguistic evidence for the identification of key themes in the Forum conversations. Approached from both qualitative and quantitative perspectives, the talk comprises a rich resource of understanding the participants' thoughts, feelings and predictions for Birmingham 2050. In partnership with the more macro-level content analysis provided by Ellie, I hope this work can provide a useful contribution to the remainder of the scenario planning process.

8. References

Baker, P. (2006) *Using Corpora in Discourse Analysis*. London: Continuum.

Rayson, P. (2008) From key words to key semantic domains. *International Journal of Corpus Linguistics*. 13:4 pp. 519-549.

Scott, M., (2008) WordSmith Tools version 5, Liverpool: Lexical Analysis Software.

Schiffrin, D. (1987) *Discourse Markers* Cambridge: CUP