Lexical analysis of IPCC Third Assessment Reports dealing with risk and uncertainty

Electronic supplement to Swart et al. (2006) working paper on Uncertainty management in assessing climate change, impacts and responses : the evolution from the first to the fourth IPCC Assessment

Mis en ligne par Minh Ha Duong
18 juillet 2006

Top words (>4000 uses)

How to read : In the IPCC Third Assessment Report, Working Group III volume, the word "emission" is used 3967 times.

WG I	WG II	WG III	Total		WG I	WG II	WG III
5501	11269	2586	19356	<change>	3040	5180	1890
5947	9427	2159	17533	<climate>	3290	4330	1580
5894	1890	1322	9106	<model>	3260	869	970
669	4307	952	5928	<impact>	370	1980	700
2400	2273	1151	5824	<global>	1330	1040	840
1265	531	*3967*	5763	<emission>	700	240	2900
1113	3442	169	4724	<water>	610	1580	120
10	813	3546	4369	<cost>	10	370	2590
228	498	3588	4314	<energy>	130	230	2620
2305	1912	89	4306	<temperature>	1270	880	60

Left, number of occurences in TAR working group. In order to correct the size bias (reports from WG I and II were about the same size, but WG III report was 34% shorter), the right side of these tables show frequencies (% text coverage times 10000, abusively rounded).

Commment : Nothing surprising here. After "Climate change", the most frequent words used in the IPCC report vary across working groups. WG I uses "model", "global" and "temperature". In WG II we read "impact" "water" and "global" . The third working group writes about "emissions", "energy" and "cost". The word "model" is frequent also in WG II and III.

Risk & uncertainty vocabulary

Note that patterns within <single> brackets are lexical (i.e. words are put in canonical form before counting). Patterns within <<double>> brackets are morphological (i.e. counting sequences of letters).

WG I	WG II	WG III	Total		WG I	WG II	WG III
1231	695	462	2388	<<uncertain>>	680	320	340

34	1217	294	1545	<risk>	20	560	210
381	671	429	1481	<<possib>>	210	310	310
43	506	452	1001	<<strateg>>	20	230	330
17	222	590	829	<decision>	10	100	429
128	342	91	561	<<proba>>	70	160	70
62	76	271	409	<choice>	30	30	200
61	47	15	123	<<plausib>>	30	20	10
12	55	16	83	<<surpris>>	10	30	10

Comment : There is no need for statistical tests to see that each working group uses a different strategy to write about risk and uncertainty. WG I almost banished the word "risk" in favor of words in the "uncertain" and "possible" family. In contrast, WG II uses "risk" a lot. "decision" takes first place in WG III. "surpris" seems under-used compared to the real degree of concern for abrupt climate change.

Vocabulary from the guidelines

This section refers to Moss and Schneider (2000) Uncertainties in the IPCC TAR : Recommendations to lead authors figures 3 and 4.

Numbers represent upper bounds, since I did not check if in context they have been used referring to the guidelines.

WG I	WG II	WG III	Total	**III. Guidelines vocabulary**	WG I	WG II	WG III
8	240	4	252	<high><confidence>	10	330	10
0	161	1	162	<medium><confidence>	0	220	0
3	43	6	52	<low><confidence>	0	60	10
11	*444*	*11*	*466*	*Total for confidence levels*			

WG I	WG II	WG III	Total		WG I	WG II	WG III
9	22	12	43	<well><establish>	10	30	30
0	27	2	29	<establish><but>	0	40	0
10	18	8	36	<speculative>	10	10	10
0	18	1	19	<compete><explanation>	0	20	0
19	*85*	*23*	*127*	*Total for qualitative uncertainty*			

Comments : The usual bias against negative results is clearly visible. WG II used vocabulary from the guidelines much more than the other two working groups. Note that these figures do NOT include text from the Technical Summaries.

Data and methods

The full TAR text was taken from the IPCC TAR CD-ROM, also available online

Methods are formally defined in the attached script. Text was converted from HTML to 7-bit clean using the html2text script, and cat and sed standard Unix tools. Content was analysed using the « locate pattern » functions in UNITEX 1.2, an open-source corpus processing system based on automata-oriented technology.

Documents joints