

Text-mining of the IGF2014 Opening Session: An Overview

The opening session of the 9th IGF kicked off in Istanbul yesterday on 2 September with 21 speakers from government, business, academia, and civil society.

The rhyming terms **high optimism** and **low realism** have floated to the top of a textual analysis of the opening statements, as has the rather counterintuitive experience (from an IGF perspective) that usage of the terms related to **commonality** was low on the first day.

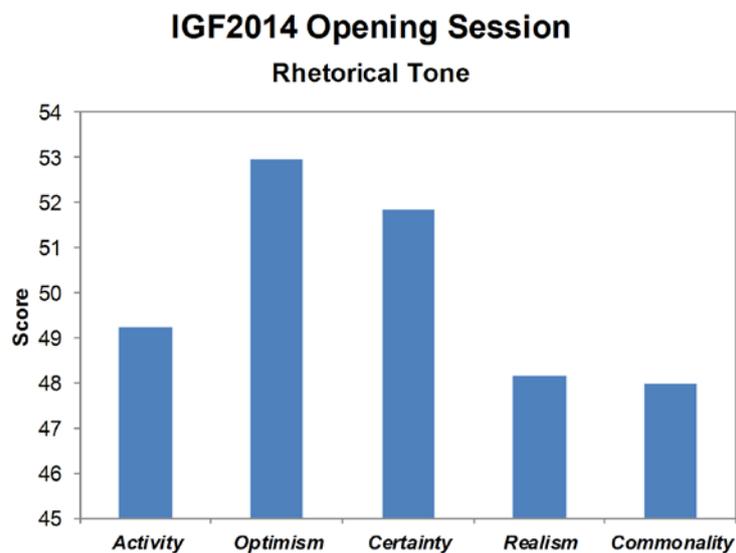


Figure 1. The overall rhetorical tone of the IGF2014 Opening Session.

The definitions of the five semantic features (*activity*, *optimism*, *certainty*, *realism*, and *commonality*) used to describe the rhetorical tone in this analysis are provided in the Appendix. High level of optimism has dominated the tone of the Opening Session as we can see how it develops from the beginning until the end of the event:

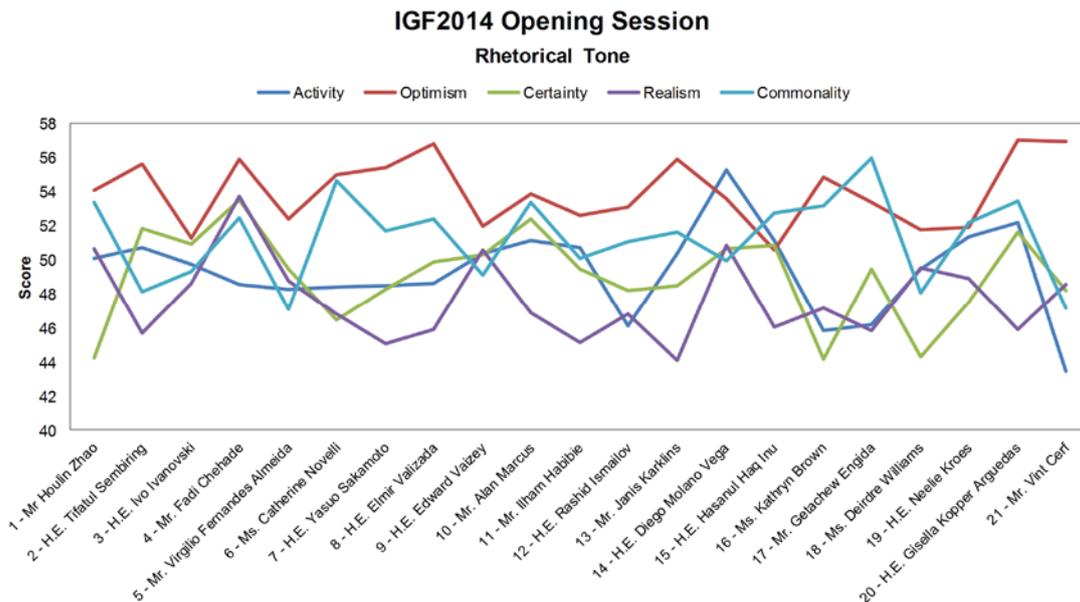


Figure 2. The rhetorical tone of the IGF2014 Opening Session as it developed during the event

In the prefix competition, one that had almost disappeared from IG language, re-emerged this year due to NETmundial. **Net** is ahead of those traditionally leading prefixes – **e-**, **cyber**, and **digital**. Another notable change, in addition to **Net**'s re-emergence, is the move from **e-** to **digital**. **E-** was the main prefix in the early 2000s and during the WSIS process, when it described e-commerce, e-health, and other WSIS follow-ups. One of the main explanations of this evolution from **e-** to **digital** is a switch in the EU's lingo. In the early 2000s, the EU's Lisbon Agenda was an **e-agenda**. The EU switched to **digital** partially to signal a new start to its digital policy and to distance itself from the mixed results of the Lisbon agenda.

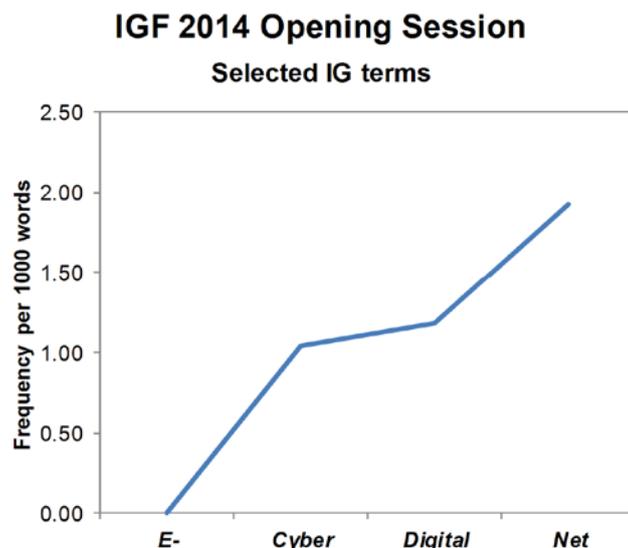


Figure 3. Frequency (per 1000 words) of selected IG words.

The distributions of word usage across the speeches delivered during the session were used to extract the most significant associations between the 15 most frequently used words and phrases:

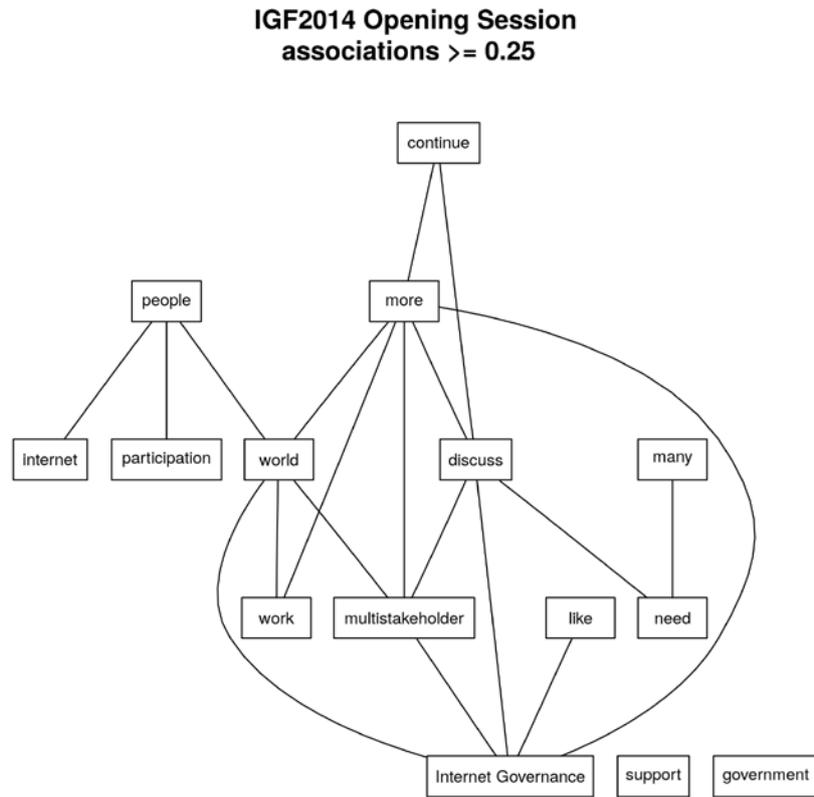


Figure 5a. Associations between the 15 most frequently used words and phrases. The usage of “Internet” was highly correlated with the usage of “people” across the speeches delivered during the IGF2014 Opening Session.

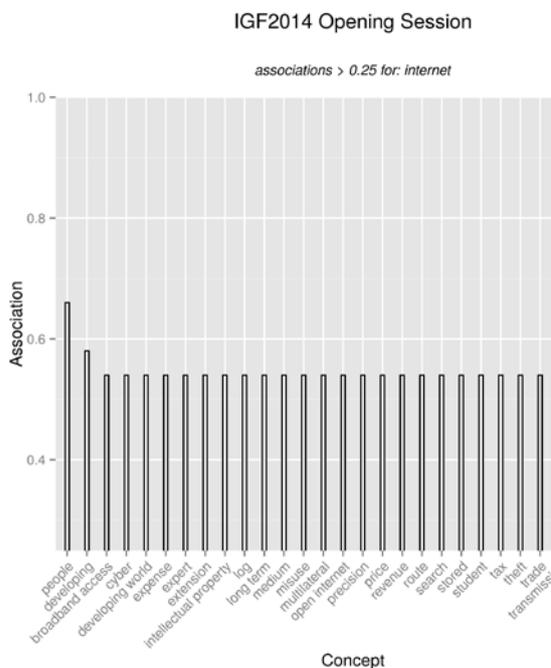


Figure 5b. Top associations for “internet”.

The following semantic space, representing the similarity of the 21 most frequently used words and phrases, was produced by examining the sentence-level co-occurrences:



Figure 6. The semantic space of the IGF2014 Opening Session. The distances between words represented similarity: the closer the two words stand in the semantic space, the more similar was the context of their usage.

Appendix

Diction 6 was used to produce the scores for rhetorical tone analyses (Figure 1 and 2) according to five semantic features: certainty, optimism, activity, realism, and commonality. These five semantic features are defined as following:

Certainty: language indicating resoluteness, inflexibility, and completeness and a tendency to speak *ex cathedra*.

Optimism: Language endorsing some person, group, concept or event, or highlighting their positive entailments.

Activity: Language featuring movement, change, the implementation of ideas and the avoidance of inertia.

Realism: Language describing tangible, immediate, recognizable matters that affect people's everyday lives.

Commonality: language highlighting the agreed-upon values of a group and rejecting idiosyncratic modes of engagement.

Diction software is widely used in the analyses of rhetorical tone. The introduction to the principles upon which it is based in found in:

R. P. Hart, "*Systematic Analysis of Political Discourse: The Development of DICTION*", in K. Sanders, et al. (Eds.), *Political Communication Yearbook: 1984* (Carbondale, IL: Southern Illinois University Press, 1985), pp. 97-134.

Text mining procedures used to produce figures 3 – 5 were developed in R, using and extending the functionality of the `tm()` package.

The R Project for Statistical Computing

<http://www.r-project.org/>

tm: Text Mining Package

<http://cran.r-project.org/web/packages/tm/index.html>

Feinerer, I., Hornik, H. & Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software*, Vol. 25, Issue 5, Mar 2008.

<http://www.jstatsoft.org/v25/i05/paper>

Semantic spaces (Figure 6) were obtained from the multidimensional scaling (`smacoff()` in R was used) of the term co-occurrence matrix previously produced in the KNIME data-mining platform. Figure 6 presents the 3D subspace from a 4D solution that has achieved a satisfying level of stress.

KNIME

<https://www.knime.org/>

KNIME Text Processing

<https://tech.knime.org/knime-text-processing>

Thiel, K. & Berthold, M. (2012). The KNIME text processing feature: An Introduction. Technical Report.

https://tech.knime.org/files/knime_text_processing_introduction_technical_report_120515.pdf

SMACOF: multidimensional scaling in R

<http://cran.r-project.org/web/packages/smacof/index.html>

J. De Leeuw and P. Mair. Multidimensional Scaling Using Majorization: SMACOF in R.
Journal of Statistical Software, 31(3):1-30, 2009.

<http://www.jstatsoft.org/v31/i03/paper>

Statistical analysis and R programming:

Goran S. Milovanović, PhD

DiploFoundation, Belgrade

Twitter: @GSMilovanovic

