The mood of the (financial) markets: In a corpus of words and of pictures[1]

Khurshid Ahmad, Pensiri Manomaisupat, David Cheng, Tugba Taskaya, Saif Ahmad, Lee Gillam and Andrew Hippisley
Department of Computing, University of Surrey, Guildford, Surrey.

Methods of corpus linguistics are typically used to study language either synchronically or diachronically. Much of the work currently being carried out under the rubric of *information extraction* benefits directly or indirectly from work in corpus linguistics: the extraction of the so-called *named entities, template elements, template relations,* and *scenario templates*, all relating to meaning bearing units of language within a text, relies on various statistical tests that have to be carried out over a corpus of texts. Some brave folk attempt to carry out these tests over a corpus of multilingual texts. One of the important developments in information extraction relates to event modelling. Here linguists look at the lexico-grammatical properties of texts and attempt to derive information about events that are supposed to be reported in the texts. We have found that event modelling requires a good understanding of the modes used in communicating the events, including natural language, graphs and images. A case study of financial market movement, where a corpus of news wires and graphical information, or a financial time series, were correlated, is described. These are preliminary results of an EU 5th Framework Project –GIDA (No. IST 2000-31123).

News streams provided by organisations like Reuters or Bloomberg comprise a range of keywords and indexical names that may change from one news item to the next; an event modeller will need to filter the news from such a diverse information resource. Specialist information providers deliver not only news texts but also supply, for example, time series of changes in value of stocks, shares, currencies, bonds and other financial instruments. We have a narrower focus than other authors in information extraction (see for example Maybury et al, 1995) in that we are looking for changes in key financial instruments that are reported in financial news-wires. The news coverage of these instruments is of two types: first, there is a daily report about changes in the value (numerical) of the instruments for instance, one can see time series comprising historic data about the changes in values of currencies; second, the manner in which the value of the instruments changes depends on the reports relating, directly or indirectly, to the instrument. The reports, for example, about war or economic uplift/downturn, affect the value of the instruments. Some authors claim that there is a correlation between 'good' or 'bad' news relating to the instrument and its potential numerical value.

Our work focuses on primary movements within a market which lasts from few months to many years and represents the broad trend within a market. We report on some initial work that attempts to changes in an index, FTSE100, with changes in 'market sentiment' as expressed in news reports about the UK economy specifically and reports about the Wall Street indices. The later has substantial influence on the UK economy. Financial analysts use sophisticated political, economic and psychological analysis to determine the reaction of market operatives and to predict the possible trading decisions of the operatives. Reports related to the sentiment use a range of metaphors to express the state of a market and its possible movements. Francis Knowles has written about the use of *health metaphors* used in the financial news reports: markets are full of *vigour* and are *strong* or the markets are *anaemic* or are *weak* (1996); most newspapers also use *animal* metaphors – there are *bull* markets and *bear* markets, the former refer to expansion, and indirectly to fertility, and the later to shy, retiring and grizzly behaviour much like that reported about bears in popular press and in literature for children. Indeed, there are fairly literal words that express the sentiment, as reported in the news wires, about the markets: financial instruments *rise, fall,* markets *boom, go bust,* and there are *gains, losses* within the markets, economies *slowdown*, suffer *downturns*, whole industry sectors maybe *hardpressed*. We created a corpus of 1,539 English financial texts from one source (Reuters) on the World Wide Web, published during a 3-month period (Oct 2001-January 2002) comprising over 310,000 tokens. The corpus comprised a blend of both short news stories and financial reports. Most of the news is business news from Britain with thirty percent of the news is from Europe and from the United Stages. We automatically extract the sentiment words and key terms from the text corpus diachronically. The correlation between the metaphorical words and the FTSE looks good at first sight:
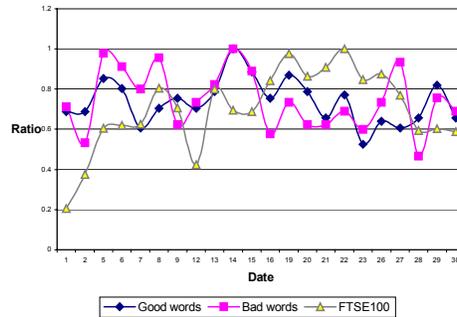
Recent work has looked beyond the frequency distributions of positive and negative sentiment keywords to the actual *relevance* of a keyword's contribution to the analysis through its grammatical properties. Specifically we look at the tense / aspect features of verbal keywords to determine their relevance to the immediate situation. Some of the categories and their relevance measures are shown below:

Greater relevance	*Less relevance*
Present Continuous e.g. stocks are rising	Present Simple e.g. stocks rise
Perfect e.g. stocks have risen	Past e.g. stocks rose

The claim is that a positive verb such as *rise* which is marked as Present Continuous will suggest greater immediacy and therefore greater relevance as an indicator of current market sentiment. In this paper we will describe our attempts to look at a much larger corpus of texts (c. 1 million token financial news stream) and perform information extraction by making greater use of the linguistic properties of the sentimental tokens thus identified.

Maybury (1995) Generating Summaries from Event Data. Information Processing and Management. 31(5) 733-751.

Knowles, F. (1996) Lexicographical Aspects of Health Metaphors in Financial Texts. In (Eds.)Martin Gellerstam et al. Euralex'96 Proceedings (Part II). Göteborg, Sweden: Göteborg University, pp 789-796.