

Vol. IV, No. II (Fall 2019) | Page: 1 – 6  
p-ISSN: 2663-3299 | e-ISSN: 2663-3841  
L-ISSN: 2663-3299  
DOI: 10.31703/glr.2019(IV-II).01  
URL: [http://dx.doi.org/10.31703/glr.2019\(IV-II\).01](http://dx.doi.org/10.31703/glr.2019(IV-II).01)

## Aspects of Discourse Stream Analysis

### Abstract

The huge proliferation of textual (and other data) in digital and organisational sources has led to new techniques of text analysis. The potential thereby unleashed may be underpinned by further theoretical developments to the theory of Discourse Stream Analysis (DSA) as presented here. These include the notion of change in the discourse stream in terms of discourse stream fronts, linguistic elements evolving in real time, and notions of time itself in terms of relative speed, subject orientation and perception. Big data has also given rise to fake news, the manipulation of messages on a large scale. Fake news is conveyed in fake discourse streams and has led to a new field of description and analysis.

### Key Words:

Big textual data, discourse, tense and time, sentiment, fake news

### Introduction

Discourse Stream Analysis (DSA) (Perkins, 2019) consists of an approach to the identification, creation and definition of concepts in text and their combinations and evolution over time in real-time flows. In my original statement of DSA (*ibid.*, p. 187ff.) I laid out certain principles, such as those relating to the subject (Searle 1992), the notion of paradigm organism (paradorg) as a meaning entity more closely defined than the notion of meme (Dennett 1993, 1995), and I later pointed out areas for further development (Perkins 2019, p.320ff.). In this paper I will focus on certain aspects of DSA such as the available data, discourse streams and discourse community, discourse streams and time, discourse streams and sentiment, and fake discourse streams in order to develop the theory further.

### Big Textual Data

In terms of the available textual data, there has been an acceleration in volume, sources, and diversity since the early 2000s. The internet offers a huge amount of data in blogs accumulated since the late 1990s (Blood 2000), social media (since the advent of Facebook in 2004 (Murphy 2013)), mainstream media and other sources, and vast amounts of text are

also held on the servers of large organisations. In tandem with the increase of accessible data, technological tools have emerged for analytical purposes (Perkins 2020).

However, there is an imbalance between the availability and history of hard data, such as economic, and soft data such as text. While hard data in the sense of figures have been available for a long time (Edvinsson et al., 2010), relatively easy to extract and record, text has been locked in the printed page, hard to extract and systematize without painstaking and time-consuming effort. Although corpus linguists began to build and store large-scale corpora computationally from the 1960s (McEnery & Hardie 2013), it was only in the late 90s that wide access to online digital text became available. The term blog was coined by John Barger in 1997 (Blood, 2000; Wortham 2007), and the creation of Facebook in 2004 marked the arrival of social media with its potential for generating huge volumes of text. According to Business of Apps (2019) 500 million tweets were sent every day in 2014, which was the last time official statistics were released by Twitter.

However, not all such big textual data is available free of charge or restrictions. Text found in online mainstream media is proprietary. As such, it cannot be downloaded and stored on a server without permission and often payment. Full access to social media text (such as Twitter feeds) can only be purchased through approved vendors (Brandwatch 2019). On the other hand, other sources of user-generated context such as blog posts are covered by creative commons, share-and-share-alike licenses, thus facilitating large-scale analysis (Creative Commons 2017). The amount of digital text available for analysis, therefore, has increased enormously in the 2000s and this has led to the development of powerful tools to that end (Perkins, 2020). This means that it is possible not only to identify and describe the content of text but also to chart developments over time. The temporal aspect of DSA is indeed extremely important, and it is one of the basic ones that I shall elaborate here and refine following my first theoretical and practical statement (Perkins 2019).

\* CEO (Chief Executive Office), Repindex Ltd., Cambridge, UK. Email: [mark.perkins@repindex.com](mailto:mark.perkins@repindex.com)

## Discourse Streams and Discourse Stream Universe

At their highest level of generality, discourse streams comprise ways of speaking about themes and topics (Foucault, 1992; Perkins, 2019). It is possible to demarcate these streams by examining setting or context, geographical, linguistic and social location, individual and group characteristics, the interactive nature of the communicative situation, and the originator(s) and receiver(s) (Hymes 1977). Suffice it to say that discourse is typically dynamic, occurs in real time (continuously or intermittently) and necessarily enjoys features to do with the interface to the world (Perkins 2019).

At a more specific level, the stream may be visualised in terms of a number of fronts which are advancing in real time. In one sense of time, the present instant, they are advancing together. Under other conceptions of time the advance is varied. Here are the main fronts which comprise the stream.

**Table 1.** Discourse Stream Fronts

Text front	Word front
	Concept front
	Phonetic front
	Semantic front (endophoric and exophoric reference)
Psychological front	Intention, Psychological condition

The discourse stream universe (DSU) is an environment in which there may be found many types of discourse stream. Each stream will have one more foci, which may interrelate (according to principles of intertextuality (Kristeva 1986; Moi 1986)) and change more quickly or more slowly over time. Although the total stream moves in real time and can be imagined in terms of the various fronts continually creating meaning, that meaning may relate internally (by means of endophoric reference) and externally (by means of exophoric reference) to events and processes in past, present and future time.

## Discourse Streams and Time

Discourse is dynamic and ever-changing. The discourse stream is manifested in a continuous flow of text, and in that flow orientations towards time are found. The semantico-syntactic system expresses the subject's time orientation.

Notions of past present and future are quite complex once grammatical form, semantics and subject orientation are taken into account (Greenbaum & Nelson 2002). Time is relative according to how it is perceived psychologically and how it is represented linguistically in the tense system.

Past forms, for example, may be used to talk about hypothetical future events, as with the second conditional.

1) If inflation rose, this would be no bad thing

Likewise, the present perfect, while employing a past form (the past participle, or part three of the verb) may be used to describe events which have a bearing on the present, as in this example.

2) Inflation has risen so far this year (and is still rising)

Moreover, when verbal aspect is considered, the speaker's view as to completion or incompleteness is revealed, as can be seen here.

3) Inflation has been rising this year (and is still rising)

Then, when the present continuous is used, it may be used to refer to events happening over extremely short durations to longer durations. The interpretation of duration can be subjective and can vary according to context. Consider this example.

4) The nucleus is splitting

This refers to an event of very short duration. But in the next example:

5) Inflation is rising

the speaker may have in mind that the event has been going on up to now, is going on now, and will be going on in the future. The precise duration of the periods before and after the present depend on the speaker's viewpoint and the context. Indeed, the expression at the moment may refer to a short-term, even instantaneous event, an event occurring over the medium-term, and to an event of long-term duration. The definition of short-term, medium-term and long-term depends on the subject and the context, as can be seen in these examples.

### Short-term

5) the atom is splitting

6) John is talking on the phone

### Medium-term

- 7) we are having a great time on holiday
- 8) Jane is enjoying her birthday

### Long-term

- 6) Venice is sinking
- 7) The sun is burning out

The subject, then, exists in the here and now but is also continually oriented towards the past and future. This means that a broad conception of the present has to be taken into account. Indeed, the present consists of an extent of time which may be longer or shorter depending on the subject and the context.

Time orientations as expressed in the above examples are found in discourse streams in terms of speed and intensity. Events can be described as occurring quickly or slowly, with varying degrees of intensity, and within different time scales. Discourse streams vary according to the discourse community and the social, economic and historical context. In the business world, an example of a very fast-moving discourse stream would be that of financial trading, and the discourse universe would comprise traders, associated people and institutions. A slow-moving discourse stream might be the real estate market. It takes a comparatively long time to buy and sell a house, whereas stocks can be traded on the market in milliseconds.

## The Creation of Future Meaning in Discourse Streams

The future consists of meaning potential which depends on the creation of meaning at the time of utterance. In the act of creating meaning about the future (as to states and events) we influence the future in the sense of the creation of one or more scenarios with various probabilities of realisation, and which may prompt others (whether people or machines) into action. Moreover, our opinion on the past also has a formative influence on meaning creation.

When sentiment is expressed a sentiment event (SE) may be said to occur. SEs draw on aspects of speech act theory Austin 1979a, 1979b; Searle 1969, 1975 ) but go much further. An SE may be internal (the subject's thought stream) or external (expressed in written or spoken form). It may or may not have an impact on real world events (RWEs) depending on whether action derives from it, although an SE, or a collection or aggregation of SEs will almost always in fact have an influence on real world events. In addition, SEs will occur before RWEs, whether fractionally or with varying time lags which might be characterised as short, medium or long relative to context, although the same or very similar SEs may occur at the same time (broadly speaking) in different geographical locations.

If this is the case, it may follow that SEs can cause, predict or influence RWEs. The predictive relationship between them may be tight or loose. This may again depend on context. Variance and degree of predictability will themselves vary according to the discourse universe under consideration.

SEs may have a causal relationship with RWEs.

- 1) The central bank intends to increase the interest rate to 2%

There has been a decision to increase the interest rate and as a result the rate will rise. While, of course, the rate may not rise (some RWE occurring in the interim period between the decision and the point of rate change might alter the eventual direction of the rate), it is highly likely that it will.

SEs may have a predictive relationship with RWEs

- 2) Most observers predict that the central bank will increase the interest rate to 2%

The RWE will probably happen, and the observers predicted that in their comments.

SEs may influence an RWE.

- 3) Many observers believe that the housing market is overheating. In that case, the central bank might increase the interest rate.

## Discourse Streams and Sentiment

Sentiment is expressed in discourse streams by means of the semantico-syntactic system as briefly outlined above. However, the term sentiment needs to be defined carefully. Only when that has been accomplished will it be possible truly to describe and understand intentional messages in the discourse stream.

In the business community, the term sentiment is used in different ways such as to indicate positive and negative stance (such as towards a brand), and also opinion as to the likelihood of future events (such as house price sentiment with regard to views on rising or falling prices). However, the use is quite simple, and sentiment can be defined in more detail.

In order to unpack the notion of sentiment, it is first necessary to identify the dimensions of sentiment. These are the following.

**Table 2.** Dimensions of Sentiment

Cognitive expression	Opinion and views on something
Emotional expression, within ranges	Feeling
Functional expression	What you can / cannot do
Intention	What you intend to do
Viewpoint	Viewpoint comparison
Engagement overall	Including degree, proximity and distance
Positivity and negativity	Cutting across all levels of analysis

These dimensions may be elicited and elaborated upon in answer to the following questions (whether posed directly or indirectly).

**Table 1.** Sentiment Elicitation Questions

What do you think about x... / believe / assert?
What do you feel about x?
What are you able to do?
What do you intend to do?
What is your view on x?
How close are you to x?
Are you positive or negative towards x?

Furthermore, sentiment has elastic properties in that it may embody positive and negative overshoot. People generally believe that the future is going to be better. When times are good, they believe that things will get even better. Evidence for this is found in decisions to invest in asset classes (such as shares on a stock market) that are rising substantially in value far beyond what would be reasonably suggested by fundamentals. In other words, people just believe that things will continue to get better and better. Parallel behaviour is found when downward trends occur. People believe that things will never get better. A good example of this is found in the political arena. The departure date for the United Kingdom to leave the European Union (EU) is set at 31<sup>st</sup> January 2020. Those advocating Brexit (the British exit from the EU) have made a number of promises regarding better times ahead in the future. Some people believe in those promises, but their beliefs may overshoot, or indeed undershoot, the reality as it may turn out. These beliefs are expressed in language and it is possible to describe the language found, quantify the level of overshoot and undershoot in sentiment and compare this with actual events. This quantification can be described as the positive-negative opinion gap (PN-OG).

### Fake Discourse Streams

While the initial analysis of big textual data centred on business uses such as brand monitoring and analysis (Perkins 2020), political applications have become prominent since the advent of social media and in particular since the rise of populism in western countries such as the USA (with President Trump), the UK (regarding Brexit) and France and Germany (regarding the rise of far right political groups) (McGonagle 2017). The creation and dissemination of messages and message manipulation by means of advanced technology has given rise to the term fake news. Interestingly, although the term fake news is new itself (Allcott & Gentzow 2017), it has already been deployed to described historical news manipulation from the American civil war to the present-day war on terror (Sirvent & Haiphong 2019).

News and message manipulation, of course, has been recurrent throughout history. One early landmark was the Latin epic poem the Aeneid in which Virgil created a story known as the myth of Rome. This poetic account of the origins and future of Rome not only offered a substantial work of poetry which was to have profound impact through the ages, but it also painted a picture politically acceptable to Octavian who later assumed the name Augustus and became known as the first Emperor (Beard 2016, pp.369-370). In more recent history, propaganda was used extensively and continuously in the communist bloc in Europe dominated by the Soviet Union. I have personal experience of living in Czechoslovakia in the early 1980s where the communist system was represented and normalised through language. Since communism was presented by the regime as logically right and good for the population, no questioning of the ideology was permitted (Reisky De Dubnic 1960). In modern times, the study of language as used to represent events for political purposes was given a major impetus by linguists such as Roger Fowler who, with others, coined the term Critical Linguistics (CL), a field which later developed into Critical Discourse Analysis (CDA) (Fowler et al. 1979; Fairclough 1995). CDA linguists attempt to provide linguistic analyses in order to reveal underlying political (and economic) agendas (Perkins 2019, pp.39-76).

What, then, is fake news, and why is it significant in the context of discourse streams and DSA? The answer lies in the technological leap made in the 2000s. Technology has made it possible to build on traditional methods of news and message manipulation. New technological methods have two main dimensions. In the first, it is possible to glean a great deal of information on people. This is done in two ways. The first may be called active information collection (AIC), and the second passive information collection (PIC). In the former, people are invited online to participate in information-gathering activities. These include online surveys (and, in particular, free text open-ended questionnaires) and other non-linguistic methods such as giving ratings by emoticons (electronic symbols of emotional state) or numerical approval scores. A vast quantity of such data is collected by corporations and governments. An example of PIC may be the collection of online user-generated text in social media platforms, but it can also include audio transcriptions (by means of voice recognition), and images found not only online but also in closed circuit TV and other photographic feeds. From these, facial image databases may be assembled.

The confluence of textual, symbolic, numeric, audio and visual technologies provides big data feeds which can be analysed in detail: there is occurring a paradigm shift in terms of quantity. However, it is the next step where it could be argued that a paradigm shift of quality is occurring. Technology has enabled the creation of a parallel world. In this world, a parallel set of discourse streams can be created, alongside new personae altogether. In this world, whole environments may be created. For example, video techniques such as the manipulation of personal images and the creation of video stories (so-called deep-fake) can be used to create alternative realities. Here, people as participants can be manipulated and convinced to vote one way or another depending on which world view they are steered into. A powerful argument can be made, for example, that the campaigns of Donald Trump and the leave organizations in the Brexit referendum were significantly aided by the technology of the company Cambridge Analytica as deployed on social media platforms such as Facebook (Kaiser 2019). If, then, fake discourse streams are being used for the purposes of manipulation, it is to be expected that counter-measures may be taken. Indeed, the detection of fake news has emerged as an activity in the late 2010s: governments and corporations are devoting increasing amounts of resource to this activity (Conforti et al. 2019).

## **Conclusions**

It is quite clear that since I first proposed Discourse Stream Analysis in 2001 (Perkins 2019) there has been an explosion of textual (and other) digital data both online and within the organisation. This enables more sophisticated and fine-grained analyses to be made which in turn prompt further theoretical developments. In this paper I have presented the notion of discourse stream fronts as basic elements of the discourse stream. It will be possible to model those elements in real time. Notions of time and the creation of future meaning are also important: time dimensions can also be modelled. A much more detailed approach to sentiment is needed, and I have given an indication of that here. The dimensions of sentiment can also, indeed, be connected to those of time. Finally, the identification and description of fake discourse streams offers a new and parallel discourse universe for analysis, which is leading to new tools for identification and analysis.

## References

- Allcott, H, & Gentzkow, M. (2017) Social Media and Fake News in the 2016 Election, in *Journal of Economic Perspectives*, Volume 31, Number 2, Spring 2017, pp.211–236. Retrieved from <https://web.stanford.edu/~gentzkow/research/fakenews.pdf>
- Austin, J.L. (1979a) *Philosophical papers*. Oxford University Press.
- Austin, J.L. (1979b) Performative utterances, in J.L. Austin *Philosophical papers*. Oxford University Press, pp.233-52.
- Beard, M. (2016) *SPQR: A History of Ancient Rome*. London: Profile Books.
- Blood, R (2000). *Weblogs: A History and Perspective*. Retrieved from [http://www.rebeccablood.net/essays/weblog\\_history.html](http://www.rebeccablood.net/essays/weblog_history.html)
- Brandwatch (2019). Retrieved from <https://www.brandwatch.com/>
- Business of Apps (2019). Retrieved from <https://www.businessofapps.com/data/twitter-statistics/>
- Conforti, C., Collier, N., Pilehvar, M. (2019) Towards Automatic Fake News Detection: Cross-Level Stance Detection in News Articles. *Language Technology Lab, University of Cambridge*. Retrieved from <https://www.aclweb.org/anthology/W18-5507.pdf>
- Creative Commons (2017). Retrieved from <https://creativecommons.org/licenses/>
- Dennett, D. (1993) *Consciousness explained*. Harmondsworth: Penguin.
- Dennett, D. (1995) *Darwin's dangerous idea*. Harmondsworth: Penguin.
- Edvinsson, R., Jacobson, T., & Waldenström, D. (eds) (2010). *Historical Monetary and Financial Statistics for Sweden: Volume 1: Exchange rates, prices, and wages, 1277–2008*. Sveriges Riksbank, Ekerlids Förlag: Stockholm.
- Fairclough, N. (1995) *Critical discourse analysis: the critical study of language*. London: Longman.
- Fowler, R., Hodge, B., Kress, G., & Trew, T. (1979) *Language and control*. Routledge & Kegan Paul.
- Foucault, M. (1992) *The archaeology of knowledge* (trans. A. Sheridan Smith). London: Routledge.
- Greenbaum, S., & Nelson, G. (eds) (2002) *An Introduction to English Grammar*. 2nd Ed. London Longman, 2002.
- Hymes, D. (1977) *Foundations in sociolinguistics: an ethnographic approach*. London: Tavistock Publications.
- Kaiser, B. (2019) *Targeted: My inside Story of Cambridge Analytica and How Trump and Facebook Broke Democracy*. London: Harper Collins.
- Kristeva, J. (1986) Word, dialogue and novel, in Moi, T. (ed) *The Kristeva reader*. Oxford: Blackwell, pp.34-61.
- McEnery, T. & Hardie, A. (2013) *The History of Corpus Linguistics*, in Allan, (ed) *The Oxford Handbook of the History of Linguistics*. Oxford Handbooks in Linguistics, chapter 34.
- McGonagle, T. (2017) “Fake news”: False fears or real concerns? In *Netherlands Quarterly of Human Rights* 35(4):203-209. Retrieved from [https://www.researchgate.net/publication/321441710\\_Fake\\_news\\_False\\_fears\\_or\\_real\\_concerns](https://www.researchgate.net/publication/321441710_Fake_news_False_fears_or_real_concerns)
- Moi, T. (1986) (ed) *The Kristeva reader*. Oxford: Blackwell.
- Murphy, L (2013) *The Fall of Buzzmetrics & Rise of The New Social Media Analytics Market Research Firm*. Retrieved from <https://greenbookblog.org/2013/04/22/the-fall-of-buzzmetrics-rise-of-the-new-social-media-analytics-mr-firms/>
- Perkins, M.C. (2019) *Discourse, evolution and power*. Cambridge: Repindex.
- Perkins, M.C. (2020) *Approaches to Text Analysis*. *Global Language Review*, IV(1), 1-7, Islamabad.
- Reisky De Dubnic, V. (1960) *Communist Propaganda Methods; a Case Study on Czechoslovakia*. New York: Praeger, 1960.
- Searle, J. R. (1969) *Speech acts*. Cambridge University Press.
- Searle, J. R. (1975) *Indirect speech acts*, in P. Cole and J. Morgan (eds) *Syntax and semantics vol 3: speech acts*. New York: Academic Press, pp.59-82.
- Searle, J. R. (1992) *The rediscovery of mind*. Cambridge MA: MIT Press.
- Sirvent, R., & Haiphong, D., (2019) *American Exceptionalism and American Innocence: A People's History of Fake News-from the Revolutionary War to the War on Terror*. New York, NY: Skyhorse Publishing.
- Wortham, J (2007) *After 10 Years of Blogs, the Future's Brighter Than Ever*. Retrieved from <https://www.wired.com/2007/12/after-10-years-of-blogs-the-futures-brighter-than-ever>