Quantitative Research in Systemic Functional Linguistics

Qingshun He[1]

[1] Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence: Qingshun He, North 2, Baiyun Avenue, Baiyun District, Guangzhou City, Guangdong 510420, China. Tel: 86-137-1523-7319. E-mail: markman1998@163.com

Abstract

The research of Systemic Functional Linguistics has been quite in-depth in both theory and practice. However, many linguists hold that Systemic Functional Linguistics has no hypothesis testing or experiments and its research is only qualitative. Analyses of the corpus, intelligent computing and language evolution on the ideological background of Systemic Functional Linguistics show that this theory focuses its research on language-in-use and is significantly quantitative in nature. It carries out both top-down and bottom-up approaches in specific studies and emphasizes on the combination of quantitative and qualitative research methods, the complementation of competence and performance data and the integration of manual and automatic operations.

Keywords: systemic functional linguistics, quantitative research, corpus, intelligent computing, language development

1. Introduction

Quantitative research is descriptive, analytical or empirical in nature. It is basically a hypothesis testing process. Currently, the research of Systemic Functional Linguistics (hereinafter referred to as SFL) has been quite in-depth in theory and has been widely applied in practice, such as translation studies, clinical discourse analyses and others. However, regrettably, many linguists outside the school of SFL mistakenly criticize that SFL has no hypothesis testing or experiments and its research is not at all quantitative (e.g. Newmeyer, 2005), even some systemicists also hold that SFL research lacks experimental proof (e.g. Berry, 1982, 1989; Butler, 1985; McGregor, 1997). In this study, we will start from the research background of SFL, to discuss relevant studies in corpus, intelligent computing and language evolution, to show the empirical nature of SFL research and to present specifically the operation methods of quantitative research through relevant typical studies.

2. The Empirical Nature of the Theoretical Background of SFL

Empirical Research can be defined as research based on experimentation or observation; it is a way of gaining knowledge by means of direct and indirect observations or experience. "Nowadays, linguistic evidence has also become a prominent topic in theoretical linguistics, where the importance of a solid empirical foundation of theoretical models is getting increasingly realized and acknowledged" (Penke & Rosenbach, 2007, p. vii).

Halliday (2002, 1985/2003, 1995/2003, 1998/2005) mentions on different occasions that his research has been affected by such scholars as Wang Li, Wittgenstein, Firth, Hjelmslev, Bernstein and Lamb, etc. SFL he founded concerns meaning, context and text and its research focuses on "language-in-use". It is just because it focuses its research on "language-in-use" that the theory of SFL cannot be separated from real language materials from which language rules and patterns are summarized. Firth (1957), Halliday's teacher, advocates observing language from the social point of view. However, emphasizing sociality does not mean at all that its research is qualitative because the research of Firth himself is to some extent empirical, for example the research on collocation. Firth (1930) proposes the formal meaning and the contextual meaning of language, the former depending on textual context, and the latter, on situational context. This appears to be subjective and hence qualitative. However, this conclusion comes from the analysis of a large number of practical language materials on the one hand, and on the other hand, the thinking of situational context comes from Malinowski, whose research method is typically on-the-spot investigation and belongs to the category of empirical research. Malinowski's academic thinking, especially his methodology with regard to the field survey had a major influence on anthropology and ethnology. The idea that the determination of the meaning of a word can only be

realized from specific cultural or situational contexts proposed by Malinowski (1923) is summarized on the basis of field survey.

Halliday inherits the thinking of both Malinowski and Firth and continues to use the research method of field survey and real language analysis. Linguistic categories and non-linguistic categories by Hasan and Martin (1989) and the three-dimensional context theory consisting of genre, register and language by Martin (1992) seemingly have nothing to do with quantitative research, but they are essentially theoretical abstractions based on rich corpus analyses and field surveys. In "Introduction: A Personal Perspective" of *Collected Works of M. A. K. Halliday, vol. 1: On Grammar*, Halliday writes:

> When I was being trained as a dialect fieldworker, by my other great teacher Wang Li (then Professor of Linguistics at Lingnan University, Canton), there were still no tape recorders. We had to transcribe responses directly into IPA script, which was excellent training for my later investigation of child language…

> Gramophone records were widely used in language teaching: when I was taught Chinese for the armed services at the University of London in 1942-43, the Department had its own recording equipment on which students could register their own performance and compare it with the recorded model. There were archives of spoken language on disk and even on cylinder, including dialect survey material in a number of different languages…

The background information fully illustrates that the research of SFL is empirical in nature since the very beginning. Halliday has carried out a long-term case study with his son Nigel and collected large amount of actual language data, with which, referring to anthropologist Malinowski and psychologist Bühler, Halliday finally generalizes the theory of metafunctions of language which is the cornerstone of SFL. The distinction of pragmatic function, magic function and phatic communion of language by Malinowski (1935) is based on field survey and the division of representational, expressive and conative functions by Bühler (1934) is based on experimental analyses. These are the exact reflection of the idea that "we shall consider language in terms of its use" (Halliday, 1970/2002). "It could also be claimed that system networks themselves, which are at the heart of Halliday's theory, constitute predictive hypotheses about what combinations of features are possible." (Butler, 2003, p. 203) "The problem, in my view, is that once formulated, such hypotheses tend to be reinterpreted by practitioners of SFG as accepted fact, rather than submitted to rigorous testing and modification or even outright rejection." (ibid, p. 204) In the following sections we will discuss the quantitative research of SFL from several specific aspects.

3. Quantitative Research in SFL

Firth started to study the systematic rules from the actual use of language, which contributes to the development of corpora and corpus-related research. The development of computer technology makes it possible to build large corpora and to process natural language quantitatively and leads to the progresses in the field of artificial intelligence, among which natural language processing plays a very important part. The research of language evolution is directly related to natural language processing.

3.1 Corpus

"Two key tools of empirical linguistics at the turn of the century are the corpus and the computer." (Sampson, 2001, p. 12) "There are, therefore, close ties between corpus linguistics and SFL" (Neale, 2006) and "corpus-based methodology and text-based research have played a central role in SFL since the beginning" (Matthiessen, 2006). The corpus-based research ties the form and meaning of language closer in certain context, and the corpus itself organically links the form, meaning and function of language together.

"It [collocation] is a central phenomenon within corpus linguistics" (Tucker, 2006). In the cline from lexis to grammar, the collocation at the lexical pole constitutes the structure at the grammatical pole, and the frequency of co-occurrence of two lexical items in a particular span along the lexico-semantic cline can be measured in the corpus. "Paradigmatically, lexical items function in sets having shared semantic features and common patterns of collocation. Thus, *tree, flower, grass* share the feature of being generic names of plants; the corpus might show that they have in common a tendency to collocate with names of colours, various forms of the item *grow* and so on." (Halliday & Matthiessen, 2004, p. 40) Collocation is also the main content of discourse analysis. Certain content words of the highest frequency extracted from certain types of discourses in the corpus can be regarded as the kernel words of this type of discourse, constituting the main source of lexical cohesion, and the collocation of these content words can be counted through the corpus. "While this study focuses on analysis of particular 'lexical items'" (Sinclair, 2004, p. 148), it does so with the aim of revealing the overall textual relationships,

meanings and coherence in the corpus" (Cheng, 2009).

According to SFL, language use is the item choosing from the language system. In a number of items that meet the entry condition, it should be counted specifically in the corpus to decide which item best fits the particular context and the frequency of occurrence of each item. That is, the corpus helps determine the probability of each item in the system of choice; this is a complex calculation process.

In addition, the construction of corpora and that of language theories are not isolated. The construction of the corpus plays a fundamental role for the construction of language theories. Halliday and Matthiessen (2004, pp. 34-35) enumerate three plusses relating to the use of the corpus: First, its data are authentic; Second, its data include spoken language; Third, the corpus makes it possible to study grammar in quantitative terms. Currently, "SFL is an 'extravagant' theory, which consciously provides a rich description. This helps to explain why SFL corpus-based work is generally slow, unmechanised and small-scale in comparison with corpus linguistics." (Thompson & Hunston, 2006) The basic situation is that there is "so much theory built overhead with so little data to support it" (Halliday, 1996/2002). To overcome these problems, we need to pay more attention to the relationship between the construction of the corpus and that of theory, to make full use of modern computer technology to build large corpora of natural language, and to apply the most advanced software programs to convert the language data into effective information resources. Although there are some urgent problems to be dealt with, corpus-based SFL research is becoming increasingly significant. Corpus-based SFL research is comprehensive, in which quantitative and scientific methods are indispensible.

3.2 Intelligent Computing

Intelligent computing is a natural language based computation process. The aim of intelligent computing is to establish language models statistically, to analyze and process natural language, and eventually to enact the communication using natural language between human and computer. Intelligent computing includes two processes, i.e., computing and reasoning. "People reason and infer with meanings, not with wordings" (Halliday & James, 1995/2005). However, reasoning is performed on semantic representations, and meaning is construed in wording. Therefore, it is necessary "to refer explicitly to the notion of 'computing meanings' to break through the current word (or words string) computing research limitations" (ibid). In fact, since 1980s, people have begun to consider establishing large-scale corpora to describe the grammar of natural language in a computable way. "Thus the concept of 'computational linguistics' already implied something closer to 'computing meanings'" (ibid).

Natural language generation and machine translation are the main areas of research of natural language processing. The Penman system is a large-scale natural language generation system developed by Matthiessen. The core of the Penman system is the "Nigel" English systemic grammar. Nigel is a system network that contains more than 700 system nodes. When generating sentences, the Penman system will continuously repeat downward choosing along the rank scale at each node in the system network based on the input information and the default settings, until the selected features are enough for a complete sentence. The features chosen from the system network are used to construct structures according to realization rules, resulting in the creation of sentences. The COMMUNAL system developed by Fawcett and the WorkBench system developed by O'Donnell are both natural language processing systems developed in the framework of SFL.

Machine translation involves descriptive comparison between the two languages. It is required to build a large-scale bilingual corpus so that the computer can automatically search translation units in the corpus of the source language and then search and configure peer units from the target language. The segmentation of translation units starts from lexical items, each of which instantiates a particular set of semantic features, forming the meaning potential system of that lexical item. For example, Sharoff (2006) takes the adjective *little* as an example to analyze the selection of semantic features of lexical items. Within the measure-type system, the semantic features of *little* include [class-property], [animate-size], [inanimate-size], [absolute-child-age], [relative-age], [mass-size], [count-size], [duration]. However, when *little* is in collocation with a noun, the modified noun imposes restrictions on the selection of semantic features. In the nominal group *little girl*, for example, the most probable feature of *little* is [absolute-child-age], and in *little table*, [mass-size] is the most likely semantic feature.

It is necessary to compute meaning in order to enable the computer to understand natural language, and to achieve a higher level of intelligence. The current meaning computing is based on either logic or statistic. At the level of theory and research, it is usually logic-based, and at the level of application, it is mostly statistic-based. Achievements have been made in both types of meaning computing, but there are yet no major breakthroughs in natural language generation and machine translation. In Halliday's vision, breakthroughs can only be achieved in

the meaning-based meaning computing.

3.3 Language Evolution

The idea of natural language processing can be used in the research of children's language development. Therefore, the study of children's language development is not always qualitative; it is also quantitative. According to SFL, the original experience of human being is the result of phylogenesis and ontogenesis. The evolution of language is reflected in the research of children's language development. The focus of the research is on how children learn the meaning system of language in social context. Halliday (1974/2003, 1978/2003, 1979/2003) and Painter (1996) both study the development process of children's language development respectively from the perspective of ontogenesis. They record and analyze children's language statistically.

Halliday (1974/2003, 1978/2003, 1979/2003) takes his son Nigel as an example and studies children's language development from the age of zero to 2.5 years old. Research shows that children begin to develop their own language system at the age of nine-month and have got a meaning system of five components, and at the age of 15 months, the number of components in the meaning system increases to 50, expressing instrumental, regulatory, interactional, personal and imaginative functions. At this stage, Children's language contains only the strata of phonology and semantics but not the lexico-grammatical stratum. That is why Halliday refers to children's language at this stage as "proto language". At this stage, children's language can express only one meaning. At the age of 15 months, two semantic systems intersect in children's language, indicating that children are no longer creating language themselves; rather they express meanings with what they have heard around them. This marks the second stage of children's language development, the transition stage towards adult language. At this stage, children are able to use multiple meanings and to play multiple roles.

Painter (1996) studies children's language development from 2.5 to 5 years of age. The result shows that at this period of time, children have experienced another two important stages of language development. From 2.5 to 3.5 years, children understand non-verbal phenomena with language. At the age of 3.5 years, children begin to understand the value relations of meaning systems themselves with language. First, the phenomenon in children's visual field is construed into experience, and once the construing process is realized, experience is summed up as semantic categories. At the beginning of the acquisition of individual words, children are not able to classify semantic categories. Then, their acquisition of language gradually transits from individual names (proper noun category) to category names (common noun category), forming their own conceptual categories.

These studies show that the children's language development is in essence a process of how to express meanings in the mutual interaction with the surrounding environment. Children's ability to master language is mainly affected by the acquired experience and the language environment (Halliday, 1975/2003; Painter, 1984, 1999, 2009). These quantitative studies are the inspiration to the study of language evolution.

4. Method of Operation

SFL studies language-in-use. The systems are formulated with the data collected from spoken and written discourses. Theories are then tested and improved in practical use. The research method is a combination of top-down and bottom-up operations. SFL emphasizes complementarities of quantitative research and qualitative research, of competence and performance data, and of manual and automatic operations both theoretically and practicably.

4.1 Top-down Operation

The basic concepts of SFL form an abstract framework or model, based on which we can propose hypotheses, select systems and determine research methods, etc. In the relevant SFL studies, this kind of top-down quantitative research is quite common. The research on grammatical categories by Halliday (1961/2002) is operated with the top-down method. SFL first assumes four basic categories of grammatical theories: unit, structure, classification and system. Thereafter, the relationships between these categories and their relations with language materials are further assumed, involving the detailing at three scales, i.e., rank, exponence and delicacy. For example, Halliday (1961/2002, pp. 59-61) presents the framework of categories of daily menu to yield an analogy with linguistic form as the following:

Units:

      Daily menu

      Meal

      Course

      Helping

         Mouthful

Unit: *Daily menu*

Elements of primary structure	E, M, L, S ("early", "main", "light", "snack")
Primary structures	EML EMLS (conflated as EML(S))
Exponents of these elements	E: 1 (breakfast)
(primary classes of unit	M: 2 (dinner)
'meal')	L: 3 (no names available; see secondary classes)
	S: 4
Secondary structures	EL_aS_aM EL_aM EML_bS_b
	EMS_aL_c
Exponents of secondary	L_a: 3.1 (lunch)
elements (systems of	L_b: 3.2 (high tea)
secondary classes of unit	L_c: 3.3 (supper)
'meal')	S_a: 4.1 (afternoon tea)
	S_b: 4.2 (nightcap)
System of sub-classes of unit	E: 1.1 (English breakfast)
'meal'	E: 1.2 (continental breakfast)

    Passing to the rank of the "meal", we will follow through the class "dinner":

Unit: *meal, Class: dinner*

Elements of primary structure	F, S, M, W, Z ("first",
	"second", "main",
	"sweet", "savoury")
Primary structures	MW MWZ MZW FMW
	FMWZ FMZW FSMW
	FSMWZ FSMZW
	(conflated as (S)MW(Z))
Exponents of these elements	F: 1 (antipasta)
(primary classes of unit	S: 2 (fish)
"course")	M: 3 (entrée)
	W: 4 (dessert)
	Z: 5 (cheese)

         ……

And so on, until everything is accounted for either in grammatical systems or in classes made up of lexical items. Like the morpheme, "Mouthful" is the smallest unit, and all eating activity can be broken down into mouthfuls.

4.2 Bottom-up Operation

The bottom-up research against the top-down research is a method focusing on the reasoning, modifying and accomplishing of hypotheses, theories and conclusions starting from language-in-use. For example, the research on the probability of tense and polarity of the finite clauses by Halliday and James (1993/2005) is a bottom-up research. The purpose of this research is to test the probability hypothesis of the options in the system.

First, they formulate the hypothesis that grammatical systems fell largely into two types: those where the options were equally probable and those where the options were skew. In the options of equal probability, there are no unmarked items, and in the options of skew probability, there is an unmarked item. Assuming a binary system, in an "equi" system, each term would occur with roughly the same frequency, while in a "skew" system, one term would be significantly more frequent than the other. The probability distribution ranges from 0.5: 0.5 to 0.99: 0.01. The polarity system is a two-term system: positive and negative; and the tense system is a three-term

system: past, present and future. Halliday and James (1993/2005) postulate that in the primary-tense system the positive terms and the negative terms form a skew system. To test this hypothesis, they first define a clause set to investigate polarity:

(1) Identify and count all finite clauses.

(2) Within this set, identify and count all those that are negative.

(3) Subtract the negative from the finite and label the remaining set positive.

(4) Calculate the percentage of negative and positive within the total set of finite clauses.

Similarly, the clause set of primary-tense can also be defined as follows to investigate tense:

(1) Identify and count all finite clauses having modal deixis.

(2) Within the set of finite clauses remaining (which therefore have temporal deixis, that is, primary tense) identify and count those whose primary tense is future.

(3) Subtract future from temporal deixis and label the remaining set non-future.

(4) Within non-future, identify and count those whose primary tense is past.

(5) Subtract past from non-future and label the remaining set present.

(6) Calculate the percentage of present and past, within the total set of non-future primary-tense clauses.

In the counting process, homographs and other similar items are excluded. For example, *can* and *will* both can be used as nouns. The result shows that the ratio of the positive to and the negative is 89.85: 10.15, that of the present tense to the past tense is 49.18: 50.82, and that of the present tense to the future tense is 88.87: 11.13.

This kind of corpus-based quantitative research not only enriches the description of the polarity system and the primary-tense system, but also provides people with an example of particular research. A similar research is the probability statistics of some basic grammar systems by Matthiessen (2006).

4.3 Complementarity and Quantitative Research

SFL pays special attention to complementarity, opposes to completely separate grammatical categories and holds that qualitative research and quantitative research both play leading roles and supporting roles as well according to the requirement of the research itself. Thus, SFL is always implementing the intercrossing between concepts and the complementarity of methods. The complementary nature of SFL research is represented in the following three aspects.

4.3.1 Complementarity between Quantitative research and Qualitative Research

The above discussion has shown that SFL is not without hypothesis testing or experimental proof and not without quantitative research. On the contrary, SFL has always been adopting a combination of qualitative and quantitative research methods.

The research of the cohesion in English by Halliday and Hasan (1976) is a typical combination of qualitative and quantitative research. The identification of the cohesion categories and subcategories and the study of cohesion distance and direction are qualitative, and the studies of the application tendency of certain cohesion devices and cohesion density in certain texts are quantitative. For example, in accordance with qualitative description, *reference* can be divided into anaphoric *reference* and cataphoric *reference*, while in accordance with quantitative description, the probability of anaphoric and cataphoric *references* can be counted. They choose seven texts of different types and count the cohesion types and cohesion density. Data show that the average cohesion densities of each sentence in different types of text are respectively: narration 2.7, conversation 2.1, sonnet 2, biography 3.2, drama dialogue 1.3, adult informal interviews 3.1 and children's informal interviews 2.1.

The research of lexical cohesion pattern by Hoey (1991) is also a typical combination of qualitative and quantitative research methods. Unlike Halliday and Hasan (1976), Hoey (1991) includes reference, ellipsis and substitute into lexical cohesion pattern, because these grammatical items have their antecedents in context. Research shows that in non-narrative texts, if there are three or more lexical cohesion items in two sentences, the two sentences are relevant in meaning. If all sentences with three cohesion items are extracted and then all sentences having connections with three or more sentences extracted, the sentences eventually extracted constitute the synopsis of the text. For example, there are 19 sentences in *Of Studies* by Bacon, in which there are 11 sentences each having three or more cohesion items with other sentences. However, in the 11 sentences, there is only one (the second) sentence having connections with three other sentences, so this sentence can be considered as the center sentence of this text.

Quantitative and qualitative analyses are complementary to each other (Bunge, 1995). Qualitative research is the basis of quantitative research, and quantitative research makes qualitative research more accurate. In actual studies, qualitative and quantitative methods are always used with each other in order to accurately qualify on the basis of quantifying.

4.3.2 Complementarity between Competence and Performance Data

Competence data are determined by the speaker's intuition, while performance data, by the producing and understanding of language. The relative advantage of competence and performance data has long been a topic of linguistic debate. In fact, the two types of data cannot be completely separated. Performance data must be derived from competence data, and competence data must be performed in spoken or written form. From that point, competence data can also be regarded as a kind of performance data. "These sources of data are best treated as complementary to one another" (Baldwin et al., 2005). Accurate grammar can be formed with competence data, and then tested with performance data.

Language-in-use is of great advantage in comparison with intuitive data. The corpus can provide us with enough actual language materials, but not grammar. While emphasizing language intuition is of help to produce grammatical but unacceptable sentences, deviating language intuition will result in taking the erroneous application of language as a marked form, or even taking the innovative development of language as the erroneous expressions. The formation of grammar relies on the complementarity of competence data and performance data. "Writing a description of a grammar entails constant shunting between the perspective of the system and the perspective of the instance." (Halliday & Matthiessen, 2004, p. 29) Language instances are included in relevant systems according to specific entry conditions, the probability of the language instances is acquired from corpus statistics, and then the validity of the language instances is justified with the native speakers' intuition. This fully reflects the nature of quantitative research of SFL in the complementarity of competence data and performance data.

Many grammarians have adopted this method of complementarity. For example, Fillmore (1992) agrees to apply corpora to maintain the authenticity of the discourse and to discover new forms of expression and then supplement native speakers' intuition with the newly discovered language materials. Similarly, descriptive grammarians such as Quirk et al. (1985), Sinclair (1990), Biber et al. (1999) and Huddleston and Pullum (2002), etc. all describe English syntactical structures using the corpus data, while determine grammatical boundaries using native speakers' intuition. Currently, "the recent trend towards a more empirical linguistics…might be characterized as the attempt to integrate the data collection and analysis techniques and competence of the more descriptively adequate school of linguistics with the explanatory ambitions and sophisticated theoretical architecture of the more rationalist school" (Featherston & Winkler, 2009).

4.3.3 Complementarity between Manual Processing and Automatic Processing

The advantage of computer processing lies in its being able to process more data quickly and to reveal some implicit language features. However, there is still a considerable distance between computer processing and manual processing in depth, accuracy, flexibility and richness, etc. "Automatic analysis gets harder the higher up we move along the hierarchy of stratification." (Halliday & Matthiessen, 2004, p. 49). That is to say, the higher the grammatical rank is, the more difficult the automatic analysis will be. For example, automatic analyses can deal with any models described in words and models of lower lexico-grammatical rank, but cannot conduct systemic functional analysis of clauses completely or conduct meaning analyses, because meaning is vague and ambiguous in essence. Many scholars propose "to have a trade-off between volume of analysis and richness of analysis: low-level analysis can be automated to handle large volumes of text, but high-level analysis has to be carried out by hand for small samples of text." (ibid)

For example, Matthiessen (2006) found from manual processing of a small text archive of about 6500 clauses that "'mental' clauses have disproportionately many selections of the circumstantial type of 'manner: degree'" (Matthiessen, 2006). The circumstantial selection "manner: degree" is more frequent in mental clauses. Most instances for the adverb *deeply* occur in mental clauses, especially in "emotive" ones. "There is thus a collocational pattern involving manner: degree: *deeply* + Process: verb of emotion" (ibid). So a search of *deeply* in a larger corpus starts from the selection of several adverbial groups functioning as "manner: degree" type of circumstance, to count their distribution and then to make sorting and classification, to distinguish degree adverbs with an adverbial group as Head and degree adverbs with a nominal group or an adverbial group as Modifier.

It can be seen that a combination of automatic analysis and manual retrieval and classification should be carried out in the data analysis, because such a combination can be helpful for manual analysis and automatic analysis to

complement each other.

5. Conclusion

Many current studies of SFL are qualitative. That does not mean that there is no hypothesis testing or experiments in the theory of SFL. Due to the complex nature of human experience itself and the dynamic nature of human social relations, language is fuzzy and dynamic in nature, hence difficult to quantify. However, quantitative research helps position certain language features more accurately and more specifically. This point has been being recognized in the development process of SFL theory since the beginning. Although certain studies place extra emphasis on qualitative discussion, the basic idea of SFL research is still the combination of qualitative and quantitative methods. In general, in the cline from instance to system, qualitative research dominates at the system pole, and quantitative research dominates at the instance pole, for example, the research of language modeling by Halliday (2005). The nature of language is hard to reveal with pure qualitative research, but the weakness of pure experimentalism is likely to show with pure quantitative research. The proper combination of the two methods will be more conducive to language studies.

Acknowledgments

References

Baldwin, T., Beavers, J., Bender, E. M., Flickinger, D., Kim, A., & Oepen, S. (2005). Beauty and the beast: What running a broad-coverage precision grammar over the BNC taught us about the grammar - and the corpus. In S. Kepser, & M. Reis (Eds), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives* (pp. 49-69). Berlin & New York: de Gruyter.

Berry, M. (1982). Review of M. A. K. Halliday, 1978, language as social semiotic: The social interpretation of language and meaning. *Nottingham Linguistic Circular, 11*, 64-94.

Berry, M. (1989). They're all out of step except our Johnny: A discussion of motivation (or the lack of it) in systemic linguistics. *Occasional Papers in Systemic Linguistics, 3*, 5-67.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Bunge, M. (1995). Quality, quantity, pseudo quantity and measurement in social science. *Journal of Quantitative Linguistics, 1*, 1-10. https://doi.org/10.1080/09296179508590027

Butler, C. S. (1985). *Systemic Linguistics: Theory and Applications*. London: Batsford Academic and Educational.

Butler, C. S. (2003). *Structure and Function: A Guide to Three Major Structural-Functional Theories: Part 1: Approaches to the Simplex Clause*. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.63

Bühler, K. (1934). *Sprachtheorie: Die Darstellungfunktion der Sprache*. Jena: G. Fischer.

Cheng, W. (2009). Describing the extended meanings of lexical cohesion in a corpus of SARS spoken discourse. In J. Flowerdew, & M. Mahlberg (Eds), *Lexical Cohesion and Corpus Linguistics* (pp. 65-83). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/bct.17.05che

Creswell, J. W. (1994). *Research Design: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage Publications.

Featherston, S., & Winkler, S. (2009). Empirical linguistics: Process and product. In S. Featherston, & S. Winkler (Eds), *The Fruits of Empirical Linguistics* (pp. vii - xiv). Berlin & New York: de Gruyter. https://doi.org/10.1515/9783110216141

Fillmore, C. J. (1992). "Corpus linguistics" or "computer-aided armchair linguistics". In J. Svartvik (Ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August, 1991* (pp. 35-60). Berlin & New York: de Gruyter. https://doi.org/10.1515/9783110867275.35

Firth, J. R. (1930). *Speech*. London: Ernest Benn Limited.

Firth, J. R. (1957). Modes of meaning. In J. R. Firth (Ed.), *Papers in Linguistics, 1934-1951* (pp. 190-215). London: Oxford University Press.

Halliday, M. A. K. (1961/2002). Categories of the theory of grammar. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 1: On Grammar* (pp. 37-94). London: Continuum.

https://doi.org/10.1080/00437956.1961.11659756

Halliday, M. A. K. (1970/2002). Language structure and language function. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 1: On Grammar* (pp. 173-195). London: Continuum.

Halliday, M. A. K. (1974/2003). A sociosemiotic perspective on language development. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 4: The Language of Early Childhood* (pp. 90-112). London: Continuum. https://doi.org/10.1017/S0041977X0009488X

Halliday, M. A. K. (1975/2003). Learning how to mean. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 4: The Language of Early Childhood* (pp. 28-59). London: Continuum. https://doi.org/10.1016/B978-0-12-443701-2.50025-1

Halliday, M. A. K. (1978/2003). Meaning and the construction of reality in early childhood. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 4: The Language of Early Childhood* (pp. 113-143). London: Continuum.

Halliday, M. A. K. (1979/2003). The ontogenesis of dialogue. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 4: The Language of Early Childhood* (pp. 144-152). London: Continuum.

Halliday, M. A. K. (1985/2003). Systemic background. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 3: On Language and Linguistics* (pp. 185-198). London: Continuum.

Halliday, M. A. K. (1995/2003). A recent view of "missteps" in linguistic theory. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 3: On Language and Linguistics* (pp. 232-247). London: Continuum. https://doi.org/10.1075/fol.2.2.07hal

Halliday, M. A. K. (1996/2002). On grammar and grammatics. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 1: On Grammar* (pp. 384-417). London: Continuum. https://doi.org/10.1075/cilt.121.03hal

Halliday, M. A. K. (1998/2005). On the grammar of pain. In J. Webster, J. (Ed.), *Collected Works of M. A. K. Halliday, vol. 7: Studies in English Language* (pp. 306-337). London: Continuum. https://doi.org/10.1075/fol.5.1.02hal

Halliday, M. A. K. (2002). Introduction: A personal perspective. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 1: On Grammar* (pp. 1-16). London: Continuum.

Halliday, M. A. K. (2005). Collected Works of M. A. K. Halliday, vol. 6: Computational and Quantitative Studies (Ed. J. Webster). London: Continuum.

Halliday, M. A. K., & James, Z. L. (1993/2005). A quantitative study of polarity and primary tense in the English finite clause. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 6: Computational and Quantitative Studies* (pp. 93-129). London: Continuum.

Halliday, M. A. K., & James, Z. L. (1995/2005). Computing meaning: Some reflections on past experience and present prospects. In J. Webster (Ed.), *Collected Works of M. A. K. Halliday, vol. 6: Computational and Quantitative Studies* (pp. 239-267). London: Continuum.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar* (3rd edition). London: Edward Arnold.

Halliday, M.A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hasan, R., & Martin, J. R. (Eds). (1989). *Language Development: Learning Language, Learning Culture. (Advances in Discourse Processes 27)*. Norwood NJ.: Ablex.

Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316423530

Malinowski, B. (1923). The problem of meaning in primitive languages. In C. K. Ogden, & I. A. Richards (Eds), *The Meaning of Meaning* (pp. 451-510). London: Routlege and Kegan Paul.

Malinowski, B. (1935). *The Language of Magic and Garden, vol. 2, Coral Gardens and Their Magic*. New York: American Book Co.

Martin, J. R. (1992). *English Text: System and Structure*. Philadelphia, Amsterdam: John Benjamins. https://doi.org/10.1075/z.59

Matthiessen, C. M. I. M. (2006). Frequency profiles of some basic grammatical systems: An interim report. In G.

Thompson, & S. Hunston (Eds), S*ystem and Corpus: Exploring Connections* (pp. 103-142). London: Equinox.

McGregor, W. B. (1997). *Semiotic Grammar.* Clarendon Press, Oxford.

Neale, A. (2006). Matching corpus data and system networks: Using corpora to modify and extend the system networks for TRANSITIVITY in English. In G. Thompson, & S. Hunston (Eds), S*ystem and Corpus: Exploring Connections* (pp. 143-163). London: Equinox.

Newmeyer, F. J. (2005). A review of CS Butler structure and function. *Functions of Language, 2*, 275-283. http://dx.doi.org/10.1075/fol.12.2.07new

Painter, C. (1996). The development of language as a resource for thinking: A linguistic view of learning. In R. Hasan, & G. William (Eds), *Literacy in Society.* London & New York: Longman.

Painter, C. (1984). *Into the mother tongue: A case study in early language development.* London: Pinter.

Painter, C. (1999). *Learning through Language in Early Childhood*. London: Cassell Academic.

Painter, C. (2009). Language development. In M. A. K. Halliday, & J. Webster (Eds), *Continuum Companion to Systemic Functional Linguistics* (pp. 87-102). London: Continuum.

Penke, M., & Rosenbach, A. (2007). *What Counts as Evidence in Linguistics.* Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/bct.7

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London & New York: Longman.

Sampson, G. (2001). *Empirical Linguistics.* London & New York: Continuum.

Sharoff, S. (2006). How to handle lexical semantics in SFL: A corpus study of purposes for using size adjectives. In G. Thompson, & S. Hunston (Eds), *System and Corpus: Exploring Connections* (pp. 184-205). London: Equinox.

Sinclair, J. (1990). *Collins COBUILD English Grammar*. London: Harper Collins.

Tayler, G. R., & Turnbull, M. (2000). Practical applications for developing research paradigms in quantitative and qualitative research. In G. R. Tayler (Ed.), *Integrating Quantitative and Qualitative Mathods in Research* (pp. 221-234). Lanham: University Press of America.

Thompson, G., & Hunston, S. (2006). System and corpus: Two traditions with a common ground. In G. Thompson, & S. Hunston (Eds), *System and Corpus: Exploring Connections* (pp. 1-14). London: Equinox.

Tucker, G. (2006). Systemic incorporation: On the relationship between corpus and systemic functional grammar. In G. Thompson, & S. Hunston (Eds), *System and Corpus: Exploring Connections* (pp. 81-102). London: Equinox.