

Empowerment through language description, empowering language description¹

(Part-I)

Christian M.I.M. Matthiessen

Introduction: empowerment

The theme of ISFC 48 is "SFL: power and empowerment". One of the most fundamental ways in which systemic functional linguists can empower communities around the world is by developing systemic functional descriptions of the languages relevant to the speakers of the communities. Such descriptions are powerful because they are couched in terms of the holistic theory of the architecture of modern (as opposed to archaic) language in context and because, empowered by this theory, descriptions of particular languages are comprehensive, meaning-oriented and text-based (e.g. Matthiessen 2007, 2023b, 2024; Matthiessen & Teruya 2024). Consequently, a systemic functional description of any particular language is an appliable description (e.g. Halliday 2008; Matthiessen 2014a): it has the potential and power to be applied to a wide range of contexts of research and application to address questions and problems that arise in the community of the speakers of the language; in today's university parlance, it can support positive impact statements. In other words, descriptive research in SFL is designed to be translational in nature (to borrow a term from medical research).

Thus, systemic functional descriptions of particular languages are pre-conditions for engaging in activities such as L1 and L2 education, language-based education in different school subjects and university disciplines, inter-language translation, intra-language translation (e.g. to make legal or medical texts accessible to the

¹ I am very grateful to Eyas Hamed and Mohamed Ali Bardi for their generous expert help with Arabic texts that I have drawn on to arrive at interpretations I present or allude to here.

general public), for undertaking language-based psychotherapy, for addressing problems with communication in healthcare, for diagnosing language disorders and developing forms of treatment. These examples are located towards the applied end of activities undertaken by linguists. One aspect they have in common is that they involve text (discourse) analysis, and text analysis can only be carried out systematically and systemically if it based on comprehensive descriptions of the linguistic systems instantiated in text, this being a fundamental way of staying clear of text analysis based on "cherry picking" (cf. Baker & Levon 2015).

At the same time, comprehensive descriptions also support activities that are, or may be seen to be, more internal to language sciences, such as language documentation, language comparison and typology, historical linguistics and grammaticalization studies, and register cartography. The activity of describing languages is a central aspect of 'doing linguistics'. Despite the fundamental importance of language description in terms of both science and benefits to the community, it was devalued and pushed to the periphery by Noam Chomsky; cf. among many commentators criticism of the Chomskyan programme, Gross (1979); Ellis (1994); Seuren (2004); Evans & Levinson (2009); Lukin (2011); Pullum (2022)². Taking

² I recall a talk Noam Chomsky gave at UCLA in the first half of the 1980s to a general academic audience. In this talk, he claimed that it used to be possible to produce a description of a language as a PhD in linguistics but that now that was no longer sufficient: a PhD in linguistics had to deal with "theoretical" questions (cf. Chomsky 1965). To anyone familiar with linguistics as a discipline, his claim was patently false — although it might have been taken as an accurate representation of the discipline by non-linguists not familiar with the breadth of linguistic research; and at the same time, it represented a completely unfounded devaluation of descriptive linguistics as a vibrant empirical strand within linguistics. When his view of the place of language description in the hierarchy of values in linguistics is taken together with his stance towards applied linguistics (e.g. de Beaugrande 1991), the effect is clearly not one of encouraging the development of theoretically empowered language descriptions to serve as appliable resources. In his extended documentation of the negative effect Chomsky has had on applied linguistics, Rajagopalan (2004, 403) writes in detail about the "sadly deleterious impact of Chomsky's revolution in linguistics on developments in AL [Applied Linguistics], especially from the 1960s to the 1980s". Later in the same chapter, he refers to Halliday's (1974, 405) "radical departure from Chomsky's

Chomsky's programme seriously, generations of PhD students focussed on his theoretical questions, and a huge amount of person power and time was diverted from the urgent task of language description, which had been put on the agenda so passionately and compassionately by Franz Boas and his followers — now again given a boost through various language documentation initiatives (see e.g. Evans 2022). But by the turn of the century, one metalanguage that had supported description of a wide range of languages, Tagmemics, had largely disappeared (cf. Pike 2001).

In this chapter, I will begin by locating language description within a system of activities undertaken by linguists, characterizing it and discussing strategies for managing the complexity of the task of describing a particular language (Section 1). Then, I will outline the role that the general theory of (modern) language can play as a resource supporting the development of descriptions of particular languages (Section 2). Next, I will identify different sources that can serve as data or material for the development of a systemic functional description of a language, differentiating and considering primary sources and secondary sources (Section 3). Once descriptions get under way, we may find that we encounter or can envisage alternative descriptions of the same phenomena (Section 4). In the final section before the conclusion (Section 5), I will illustrate ways in which the systemic functional approach to the description of a language can be applied to a few areas of Modern Standard Arabic (MSA); this will centrally involve the contrast between the approach to grammar "from below" in traditional grammar and Halliday's (e.g. 1978) trinocular vision, which enables us to view the lexicogrammar of a language in the round as a resource for creating meanings as wordings. In the Conclusion, I will summarize the main points, and relate the chapter to other directly relevant publications I have been involved in to provide an indication of where information relevant to the development of systemic functional descriptions of a particular language can be found. I have also included an Appendix

more restrictive view of linguistics as a branch of cognitive science". In contrast with Chomsky, Halliday has always highlighted the value of descriptions of particular languages in its own right and the complementarity of theoretical and applied linguistics — his concept of appliable linguistics.

with information about MSA that I have compiled from online typological databases (WALS and Grambank) so that future systemic functional descriptions of MSA can make contact with the information provided by these databases.

I hope this chapter will be helpful to researchers embarking on descriptions of particular languages! But it is not a manual in systemic functional language description — a document we need quite urgently³; rather it is a more (meta-)theoretical overview of points to consider as we set out on the description of a particular language and of the ways in which SFL can empower us.

1. Language description in a system of linguist's activities

There are many ways of characterizing linguistics as a discipline; but for present purposes, I will focus on what is involved in 'doing linguistics'— i.e. the different activities we engage in as linguists. Let me begin by locating the **linguistic activity** of *describing* a particular language such as Akan, Arabic, Bajjika, Chinese, Dagaare, English, Japanese, Korean, Mongolian, Oko, Telugu, Thai or Vietnamese within a taxonomy of non-applied linguistic activities: see Figure 1. I have represented this taxonomy as a system network, with both systems ordered in delicacy (e.g. ABSTRACTION: language / particular language > GENERALITY: generalization / description / analysis) and also with simultaneous systems (analysis / synthesis & primary data / secondary data).

³ I recall one of the long hikes Michael Halliday and I enjoyed in the coastal mountain range in Southern California around four decades ago. During one of these hikes, we decided we needed to produce an account of how to develop a systemic functional description, but this task remained on the to-do agenda as we worked on various other projects.



Figure 1: Types of (non-applied) activities undertaken by linguists with focus on options in description

The first distinction among the activities undertaken by linguists is based on **ABSTRACTION**: either we develop a theory of **language** as a general human system — a theoretical "architecture" of modern language in context — or we deal with **particular languages** at some degree of generality — i.e. particular manifestations of language as a kind of higher-order human semiotic system.

1.1. Theory of language

The theoretical architecture of language in contexts supports our engagement with particular languages; in terms of language description, it provides a *template or a grid for developing the description of a particular language* (Section 2). The systemic functional theory of language in context characterizes modern language as a higher-order semiotic system among other semiotic systems, e.g. differentiating the architecture of language from that of primary semiotic systems such as human protolanguage.

It also differentiates the architecture of modern language from that of archaic language in the evolution of humans, and, by another step back in time, from the evolutionary phase that preceded archaic language, viz. protolanguage (Matthiessen 2004a). In my hypothesis about the phases of language evolution, they correspond to Halliday's (e.g. 1975, 2003a) ontogenetic phases: protolanguage transition (archaic language) — adult, post-infancy language (modern language). This is in keeping with a general principle the gradual increase of the complexity of these semiotic systems, and at the same time their increasing meaning-making power.

In contrast with the earlier phases of language (protolanguage, archaic languages) and with other semiotic systems such as gesture, facial expression and proxemics in face to face interaction (cf. e.g. Arndt & Janney 1987; Ngo et al. 2021), modern language is organized as a **quadristratal metafunctional semiotic system**; in contrast, protolanguage is bistratal (content and expression) rather than quadristratal and microfunctional rather than metafunctional. Thus when we choose to describe a particular language, we use the theory of modern language as a resource: we are guided by the architecture of language, and we investigate the particular language under description in terms of the four strata, in terms of the metafunctional organization, and in terms of the other dimensions of the architecture.

1.2. Accounts of particular languages

Contrasting with the option 'language' — involving the activity of constructing theories of language as a general shared human semiotic system, the other option in the system of ABSTRACTION in

Figure 1 is 'particular languages'; i.e. the linguist engages with one or more *particular* languages — i.e. with one or more manifestations of modern language as a system characteristic of AMHs (Anatomically Modern Humans, i.e. Homo sapiens sapiens; cf. Matthiessen 2004a).

The nature of the engagement depends on the focus on the particular language(s) — the degree of GENERALITY — we analyse texts, describe a language, or generalize across descriptions: we choose among the three options of (i) analysing texts (always "embedded" in context), where the focus is located at the instance pole of the cline of instantiation, (ii) describing (aspects of) a particular language, where the focus is located at the potential pole of the cline of instantiation (language as meaning potential, the linguistic system)⁴ (see Figure 3, discussed below in Section 2), and (iii) generalizing across descriptions of particular languages.

Since I am concerned with the description of particular languages here, I have extended the metalinguistic system network in delicacy at this point. There are two simultaneous systems of choice with 'description of a language' as their entry condition, viz. DIRECTION and ORDER OF SYSTEM:

• DIRECTION: We develop the description by analysing (examples in) texts (either primary data or secondary data) or by constructing (synthesizing) examples, e.g. in order to check hypotheses based on our emergent description with a language consultant. When we analyse texts in order to develop descriptions of particular languages, we can choose

⁴ Intermediate between the two poles of the cline of instantiation is the region of register variation — variation in language according to context of use. We can study and account for registers by describing them as linguistic subsystems or by analysing them as text types; these two approaches are complementary, starting from the different poles of the cline of instantiation (potential and instance, respectively).

to conduct the analysis manually or to use computational tools to perform automatic analysis.

- If the direction is one of 'analysis', we can choose 0 between two options in the system of AUTOMATION, viz. 'manual' vs. 'automated' analysis. Traditionally, descriptive linguists had to rely solely on manual analysis; but as computational corpora and corpus tools began to be developed, they have increasingly been able to supplement the manual analysis with automated analysis using corpus tools. Since corpus tools have largely been restricted to the view "from below" (in terms of strata and ranks), they have remained supplementary⁵. However, there have been experiments outside SFL with fully automatic analysis, at least in the area of morphology: see Goldsmith (2001, 2006), Lee & Goldsmith (2016) using unsupervised learning. Morphology is of course, precisely "low-level" grammar being focussed on the lowest rank within the units of the rank scale of a language (e.g. Matthiessen 2015, 2023c).
- At the same time, if we are doing 'analysis', we can choose between 'systemic' analysis and 'pre-systemic' analysis. We would typically choose 'pre-systemic' analysis during the initial phase of a project concerned with the description of a "new" language" (cf. Figure 7, to be discussed below) since during this phase, we have not yet got a systemic description to reference in the analysis. This pre-systemic approach may involve 'automated' analysis, e.g. using a corpus

⁵ For the trade-off between the "level" of analysis (i.e. strata and ranks) and the potential for the automatic analysis, see e.g. Matthiessen & Teruya (2024). Unless one has a holistic theory showing what a comprehensive description of a particular language can and should be like, it may be difficult to appreciate how little of a language can actually be illuminated using just corpus tools since they provide only the view "from below" — i.e. of low-level patterns. The descriptive way forward is thus based on a clear account of the complementarity of manual and automated analysis.

tool to identify the most frequent items in a corpus since they are most likely to be grammatical ("function words") rather than lexical ("content words"), to identify collocations and Ngrams, and we can then produce concordances for these very frequent grammatical items. But we can also use 'manual' analysis, identifying patterns in texts from registers that are likely to serve as "gateways" into different lexicogrammatical domains (see further Section 3.1.2 on registers and areas of wordings "at risk").

ORDER OF DATA: We can use primary data, i.e. instances of the language under description, ranging from whole texts via passages of texts like rhetorical-semantic paragraphs (parasemes) to short examples of a clause or clause complex taken from texts; or we can use secondary data, i.e. existing descriptions of the language under description, ranging from field notes to full-fledged reference grammars⁶, or generalizations, e.g. in the shape of general freely available typological databases such as WALS and Grambank.

I will deal with these two orders of data in more detail in Section 3.

1.3. Scale of linguistic work involved: from analysis to theory construction

The different activities of doing linguistics identified in Figure 1 differ in ABSTRACTION and GENERALITY, as already noted. But these differences also have implications for other aspects of doing linguistics, importantly the size of the "material" (primary data like corpora of texts or secondary data like descriptions of languages) that we need to process and therefore also the timeframe needed for the linguistic activity. I have set these out in Table 1.

⁶ And texts included in existing descriptions can be extracted and included in the archive or corpus of primary data.

Linguistic act	tivity	Coverage	Timefra me
language (the construction)	ory	a large representative number of generalizations across descriptions of particular languages, preferably including languages that are typologically very diverse in terms of various properties	decades
particular language(s)	generalization	sample of a few descriptions of language systems for contrastive analysis or comparison in aid of historical linguistics or a large number of descriptions selected to be representative of some generalizations about languages (e.g. languages spoken in West Africa, Afro-Asiatic languages, languages spoken around the world today)	3-5 years
	description	systematic sample of texts belonging to a significant range of registers, large and varied enough to be representative of the linguistic system under description (Biber et al., 1999, base their description on a corpus of four macro-registers of 40 million words)	3-5 years
	analysis	sample of texts, sample size depending on whether they are treated as artefacts or as specimens, and, if specimens, what claims they are intended to support (centrally, how far up the cline of instantiation are the claims located?)	weeks to months (to years)

Table 1: Types of linguistic activity: coverage and timeframe

How long does it take to develop a systemic functional description of a particular language? The short answer is that we do not know since no language has been given a comprehensive description in terms of all the dimensions of the architecture of language (in context) according to SFL — not even well-described languages such as English, Japanese, Chinese, French and Spanish (and perhaps only one tenth of languages still spoken around the world have been given descriptions of reasonable coverage); in fact, the task of producing a comprehensive description is obviously endless. Needless to say, it is important to compile "meta-data" on descriptive projects so that future initiatives can be guided by past experience. For example, there are a number of PhD theses

presenting descriptions of the clause grammars of a range of languages (for recent overviews, see Teruya & Matthiessen 2015; Kashyap 2019; Matthiessen & Teruya 2024, 448-450), and they can be consulted for guidance as to descriptive coverage (and by implication, timeframe since PhD theses typically take three to five years to produce).

A little over a quarter of a century ago, based on extensive experience with non-SFL descriptions of many languages from different parts of the world, Dixon (1997, end of Chapter 9) estimated what is involved in a descriptive project:

There are 2,000 or 3,000 languages, for which we have no decent description, that will pass into disuse within the next few generations. Trained linguists are urgently needed to document them. In some cases, native speakers can be trained as linguists but in many instances an outsider is required. All this costs money. If one can hire a properly trained linguist, someone who has already written a grammar as a Ph.D. dissertation and has a thorough grounding in Basic Theory they will need salary for themselves and for their language consultants; travel funds; equipment; facilities for writing up the description, producing a dictionary and volume of texts; and so on. At least 3 years is needed to do a good job; the total cost will be (at 1997 values) around \$US200,000.

If this work is not done soon it can never be done. Future generations will then look back at the people who call themselves 'linguists' at the close of the twentieth and beginning of the twenty-first century, with bewilderment and disdain. (Dixon 1997, 138)

Dixon's estimate of 3 years corresponds to the quantum of time allocated to PhD projects in many universities — with luck perhaps extended to 5 years; in either case, a longer period than a post-doc lasting two years. Dixon's (2010a,b; 2012) notion of a description is one based on what he calls "Basic [Linguistic] Theory", which in systemic functional terms can be interpreted as descriptive generalizations across particular languages (not "theory" of language; cf. Section 3.2.3 below). To develop a systemic functional description, we need to add more time to his estimate of 3 years, but the systemic functional description, the "deliverable" of the project, would be a comprehensive, meaning-oriented and text-based description of a language — one that would be appliable to different needs in the community of its speakers, not only a document for use in linguistics⁷. So, the investment of additional time and energy is very well motivated. How much more time would be needed for such an appliable description? In a talk delivered in Nairobi in 1972, Halliday (1972/ 2007, 223) estimates that a "good description" will take "five to ten years":

From the point of view of strict scientific truth, describing a language is an endless task. There can never be a complete description of any language; it is a logical impossibility, and even English, about which more books and articles and PhD theses have been written than about any other language, is far from being fully described and interpreted. But for practical purposes a good description, which means a semantic analysis, a grammar and dictionary, a phonetic analysis and a set of recorded texts —stories, myths, dialogues, etc. — can be done by a trained linguist, preferably a native speaker of the language in question, given proper facilities, in from five to ten years. (Halliday 1972/ 2007, 223)

(Note that the semantic description of the higher of the two content-plane strata, semantics and lexicogrammar, extends all the way from the semantics of whole texts, texts being the primary units of semantics, to semantic units and unit complexes that have lexicogrammatical realizational analogues: see e.g. Martin 1992; Halliday & Matthiessen 2006.)

⁷ Franz Boas's pioneering work on language description included a grammar, a dictionary and texts — a very important conception of language description. However, as the framework of language documentation has developed in the last few decades, descriptions have expanded in scope to include audio and video recordings, enabling descriptivists to include observations dealing with semiotic systems like gesture, facial expression and gaze accompanying face to face interaction (see e.g. Evans 2022; and cf. advances in SFL covering somatic semiotic systems: Ngo et al. 2021).

1.4. Strategies for managing descriptive projects

Given the magnitude of the task of developing a comprehensive description of a particular language, we can, or even must, divide up the task of developing it using the dimensions of the architecture of language in context according to the theory of SFL. That is, we can identify strategies that will help us *manage the complexity* of the overall descriptive tasks, ways of moving towards a comprehensive description in motivated stages. I will start with the hierarchy of stratification, and then move onto other semiotic dimensions defining the architecture of language in context.

1.4.1. The hierarchy of stratification

We can divide the descriptive work in terms of the **hierarchy of stratification** into a phonetic description, a phonological description, a lexicogrammatical description, a semantic description and a contextual description. For example, Mock (1985) presents a description of Isthmus Zapotec prosodies, thus focussing on aspects of the phonological stratum of the language, and leaving open the possibility of later descriptions of the content strata.

But there are **inter-stratal interdependencies**, especially within the expression plane (phonetics and phonology, or graphetics and graphology, or the analogues in "sign languages"), within the content plane (lexicogrammar and semantics), and across the content plane and context, especially between semantics and context. Thus, it would be hard to develop a semantic description without first having made advances in the construction of the lexicogrammatical description; and in progressing with the lexicogrammatical description, we will need to shunt up into the semantics (e.g. Halliday 1984a; and cf. Martin 1983, on comparison of languages based on semantic "tasks"). Similarly, in developing the prosodic part of the description of the phonology of a particular language, we are likely to need to refer to the lexicogrammatical description to determine "how far" to take the prosodic description in delicacy, using the lexicogrammatical description of the language to differentiate emic distinctions to be captured as delicate systems from etic ones (Halliday 1967; Halliday & Greaves 2008).

While different linguistic theories, traditions and schools vary considerably in how they handle the stratal organization of language, this "compartmentalization" of the overall description is the most commonly applied as a way of managing the complexity of the descriptive task (cf. Figure 8).

1.4.2. The spectrum of metafunction

We can divide the descriptive work in terms of the **spectrum of metafunction** into an ideational (logical & experiential), an interpersonal and a textual description. For example, the descriptions of Chinese undertaken by Chinese scholars under Halliday's supervision in the first half of the 1980s dealt with systems belonging to different metafunctions (logical: TAXIS & LOGICO-SEMANTIC TYPE; experiential: TRANSITIVITY; interpersonal: MODALITY; textual: COHESION; cf. Halliday & McDonald 2004).

The textual system of THEME has been described for a number of languages, e.g. Danish (Hestbaek Andersen 2004), Korean (Kim 2007); and Williams (1989) presents a comparison of textual structures in Arabic and English, while Ethelb (2019) examines thematic patterns in translations of scientific texts in English and Arabic.

The interpersonal metafunction is the focus of Zhang's (2020) description of Khorchin Mongolian; focussing on interpersonal meaning in negotiation, he focusses on a **stratal slice** through the system, including context, semantics and lexicogrammar. Martin, Quiroz & Figueredo (2021) cover the interpersonal systems of the clause grammars of Khorchin Mongolian, Mandarin, Tagalog, Pitjantjatjara, Spanish, Brazilian Portuguese, British Sign Language and Scottish Gaelic.

1.4.3. The hierarchy of axis

Within a specific stratum, we can divide the description in terms of the **hierarchy of axis** (axiality): while an axially comprehensive description needs to be both systemic and structural, the systemic axis being the primary one (e.g. Halliday 1966; Martin 2013; Matthiessen 2023a), for certain purposes it may make sense to focus on one axis only. However, I would not recommend this approach; there are many examples of what tends to be left out in descriptive frameworks that are syntagmatically oriented — the sense of the overall system typically being a casualty.

More helpfully, in terms of the paradigmatic axis, we can use the **cline of delicacy** to control the delicacy of the description — the "granularity" (see Figure 3, to be discussed below). Thus a description can be comprehensive in coverage but low in degree of delicacy. (Here it is important to note the approach of gradual approximation in systemic functional description, from lower to higher delicacy: see Figure 5, and the discussion in Section 2. This approach is *contextual* in nature: more delicate systems are always described in the environment of the less delicate systems they are related to.)

1.4.4. The hierarchy of rank

Within a given stratum, we can divide the descriptive work in terms of the compositional hierarchy of units — the **rank scale**⁸; this represents the "division of semiotic labour" among the units of the stratum: while the general principle is that we start describing the unit of the highest in order to work "contextually", i.e. always approaching units in their linguistic environment, and thus to maximize the information available to us as we describe the different units (cf. Matthiessen 2001), we may have good reasons to move in at a lower rank — e.g. group rank in the description of the lexicogrammar of a particular language or syllable rank in the description of the phonology of a particular language (see e.g. Halliday 1992).

Good reasons will of course include the availability of existing descriptions of higher-ranking units (cf. Fontaine 2008; Fang 2015); this is the case with the descriptions of the nominal group in a number of languages in a sequence of special issues of the journal *Word* (e.g. Mwinlaaru 2021, elaborating on an area described in Mwinlaaru 2017): the clause grammars of the languages included had already been described, so the environment in which (ranking) nominal groups function had already been fleshed out descriptively

⁸ As a semiotic dimension in the multidimensional architecture of language in context, the rank scale is a theoretical notion; but ranks within any of the strata are descriptive in nature: they need to be established based on empirical evidence in the description of each particular language. Variation in rank is one of the many areas of variation across languages.

and could serve as contextual background for the descriptions of the nominal groups.

But there are examples of systemic functional descriptions of particular languages that have focussed on a rank lower than that of the clause even when the grammar of the clause had not yet been described systemic-functionally, e.g. Made's (1988) systemic functional description of the nominal group in Indonesian.

1.4.5. The cline of instantiation

Within the content plane of languages (i.e. the strata of semantics and lexicogrammar), we may focus on the mid-region of the **cline of instantiation** and base our description initially on one or more **registers**, instead of taking on the full complexity of the overall linguistic system of the language located at the potential pole of the cline. This is a natural, well-justified way of managing the complexity of the descriptive task as we begin to develop the first description. For example, in her systemic functional description of the lexicogrammar of Thai, Patpong (2005) based her description on a large corpus of traditional Thai folktales, and having developed the description on the basis of this corpus, she checked it against a smaller multi-registerial corpus.

Studies of particular registers are, naturally, important contributions also to an established general description; for example Sellami-Baklouti (2021) reports on a contrastive study of aspects of TRANSITIVITY in legal texts in Arabic and English, referring to Bardi's (2008) general systemic functional description of Arabic.

1.4.6. Summary

These strategies that can be used to manage the magnitude of describing a particular language are all, as I noted above, derived from the systemic functional architecture of language in context. They are not just a random collection of methodological add-ons. Naturally, there are links to higher-ranking and lower-ranking domains within the grammar of a particular language that will need sorting out to proceed towards full coverage of the higher-ranking system. For example, the description of the clausal system of TRANSITIVITY is related to the description of hypotactic verbal group complexes in English, "serial verb constructions" in Akan, and in Arabic also to the ten to fifteen different "measures" of the verb with

general senses such as causation that relate to different transitivity patterns of the clause (but in potentially complex ways, and with a good deal of specific lexicalizations).

2. Theory of language: template (grid) for descriptions of particular languages

In Figure 1, the basic system is that of ABSTRACTION, which is the choice is between constructing **theory of** (modern) **language** in context as a higher-order human semiotic system and engaging with **particular languages** by analysing texts, describing the system instantiated in texts or comparing descriptions of the systems of a number of languages. In language description — my focus in this chapter, we rely on the "product" of theorizing language, viz the systemic functional "architecture" of language in context (e.g. Halliday 1994; Halliday 2003b; Matthiessen 2007); we use this **architecture as a template** guiding the development of our descriptions of particular languages. In other words, the architecture of language in context serves as a schematic map when we explore a particular language in its context of culture.

In this section, I will discuss the role that the systemic functional architecture of language in context can play, but I will start with B.L. Whorf's important attempt to produce a map for descriptive linguists, viz. his "plan and conception of arrangement", which includes theory and also certain descriptive generalizations.

2.1. Whorf's "plan and conception of arrangement"

B.L. Whorf (1891-1941) was a pioneering contributor to the American anthropological linguistic tradition — originating with Franz Boas, Edward Sapir, and further developed by Mary Haas, Ken Pike and other "descriptivists" of the generation born around a decade into the 20th century. One of Whorf's concerns was to provide guidance for the development of descriptions of particular languages, and Whorf (1956, 126) presents his "plan and conception of arrangement" of language, reproduced here as Figure 2. I mention it here, in the context of the development of systemic functional descriptions of particular languages, not only because it is of historical interest — in particular in view of Whorf's influence on the development of SFL (e.g. his notions of reactances and cryptotypes), but also because, as we develop descriptions of

particular languages, we can still use it as a "checklist" together with systemic functional guidance and material from other descriptivetypological traditions.



Figure 2: Whorf's (1956, 126) "plan and conception of arrangement" of language

Source: (Whorf 1956, 126)

The editor of B.L. Whorf's (1956) posthumous book, John B. Carrol, contextualizes the plan and arrangement as follows:

In 1938, Whorf circulated this table and accompanying outline in manuscript form among selected colleagues. It was written as a supplement to the Outline of cultural materials prepared by George P. Murdock and his colleagues at the Department of Anthropology at Yale University as a guide to ethnological field workers, and is referred to in the brief "Language" section of that outline. In several places in his writings Whorf mentions the desirability of a "world survey" of languages; this outline was doubtless intended by him as a suggested standard framework for collecting the information on particular languages which would be needed for such a survey.

The reader's attention should be directed first to the table on page 126 [reproduced here as Figure 2, CMIMM], which displays the whole scheme of language as conceived by Whorf. The subsequent outline, which represents an expansion of the semasiology section of the table, is thus an appendage to the table, even though it contains most of the meat. (Carrol 2012, 159)

From a systemic functional point of view, Whorf's "plan and conception of arrangement" can be interpreted as an outline of the architecture of language in context. It is naturally less extensive than a current systemic functional "plan" would be; significantly, it does not include the metafunctional organization of language. But he provides a very helpful trajectory and his vision of a "world survey" of languages is now in part reflected in the descriptive-typological databases of WALS (the World Atlas of Language Structures) and now also of Grambank (as part of Glottobank), which I will return to below (cf. also Section A.1).

2.2. Systemic functional matrices as descriptive overviews

In SFL, the theoretical architecture of language in context serves the role that Whorf envisaged for his "plan and conception of arrangement"; and descriptive generalizations are stated in reference to this architecture. One way of viewing the architecture is by means of **matrices** representing the intersection of two semiotic dimensions or more; they include the function-stratification matrix and the stratification-instantiation matrix showing aspects of the global organization of language in context, and the function-rank matrix showing the intra-stratal organization of the strata of the content plane of language, viz. lexicogrammar and semantics. Thus, using the **stratification-instantiation matrix**, we can view the global organization of language in context, and within it we can locate the function-rank matrix displaying the local organization of lexicogrammar, as shown schematically in Figure 3 (from Halliday & Matthiessen 2014, 50).

Figure 3: The global organization of language in contexts viewed in terms of the stratification-instantiation matrix, and the local organization of the stratum of lexicogrammar viewed in terms of the function-rank matrix



(Source: Halliday & Matthiessen 2014, 50)

The **function-rank matrix** was introduced by Halliday (e.g. 1970) as a display of the total set of systems of the lexicogrammar of English, with each cell being the "semiotic address" of one or more systems such as the systems of THEME, MOOD and TRANSITIVITY; it constitutes a comprehensive overview of the total resources of lexicogrammar as a "web of systems". Thus the systemic environment of each system identified in the matrix is explicit, and it can be investigated in view of its place in the total web of systems (just as organisms are observed and investigated within the ecosystems which they are part of). The matrix has been used to provide overviews of the systems of the lexicogrammars in the descriptions of many languages, including English (e.g. Halliday & Matthiessen 2014), Japanese (Teruya 2007), Korean (Kim et al. 2023) and a number of the languages described in Caffarel, Martin & Matthiessen (2004).

In Figure 4, I have represented a schematic version of the function-rank matrix as a "grid" through which linguists can observe a given language in its manifestation as text (i.e. at the instance pole of the cline of instantiation). I have drawn the figure as an allusion to Albrecht Dürer's drawing of his "drawing machine" or "perspective machine"⁹, with the linguist as the draughtsman (Dürer) and language as the object being drawn. This simile is helpful to think with; it reminds us that what we see depends not only on the "grid" but also on where we position ourselves relative to it (and the language under description) as observers; and it can be related to Halliday's notion of trinocular vision.

⁹ See e.g. <u>https://www.npg.org.uk/learning/digital/portraiture/perspective-</u> seeing-where-you-stand/the-drawing-machine



Figure 4: The function-rank matrix as part of the linguist's "grid" used in observing language in context

As an example of a function-rank matrix serving as an overview of the systems included in the description of a particular language, I have reproduced Halliday & McDonald's (2004, 312) "metafunction-rank matrix of Chinese". They describe it as follows:

> A metafunction / rank matrix for Chinese grammar, given below in Table 6.2 [reproduced here as Table 2, CMIMM], looks very similar to that of English (Matthiessen 1995; Halliday & Matthiessen 1999). We will treat only the clause systems in detail in this account, although other systems may be briefly referred to when appropriate (the relevant sections of this chapter are added in brackets after each system). There is a difference to note between the categories of rank and metafunction in this matrix. Both the concept of "metafunction" itself and the particular subtypes of metafunction ideational (experiential, logical). _ interpersonal, and textual – are theoretical categories; they are part of the general linguistic framework of the description. But while "rank" is also a theoretical category, the particular "ranks" or units such as clause, phrase / group, word, morpheme and their complexes are treated as descriptive terms, which may vary from one language to another. (Halliday & McDonald's 2004, 311)

Table 6.2 A metafi	unction-rank matrix f	for Chinese		
metafunction	textual	interpersonal	experiential	logical
rank				[complexes at each rank]
clause	THEME (3.1) VOICE (5.1.3.1)	моор (4.1) мораlity (4.2)	TRANSITIVITY: NUCLEAR (5.1.1-4) &	LOGICO-SEMANTIC RELATIONS & TAXIS
		polarity (4.3) assessment (4.4)	circumstantial (5.1.5) aspect (5.2.1.2)	
phrase			MINOR TRANSITIVITY (5.1.5)	
group: verbal			ASPECT (5.2.1.1) Phase (5.2.2)	
nominal	DEIXIS (2.2)		NUMERATION (2.2) CLASSIFICATION (2.2)	
adverbial	COHESION	modality (4.2) comment	TIME (5.1.5) MANNER (5.1.5)	
information unit	INFORMATION (3.2)	моор (4.1)		

Table 2: Example of a function-rank matrix — Halliday & McDonald's (2004, 312) partial map of the systems of Chinese lexicogrammar

(Source: Halliday & McDonald's 2004, 312)

The systems that are identified in the cells of a function-rank matrix of any language, e.g. THEME, MOOD, TRANSITIVITY in Table 2, are identified at *primary* delicacy. Thus, if the matrix is populated completely with systems in the course of description, the description is comprehensive at primary delicacy. This comprehensive description clearly needs to be tested against texts in context — a move along the cline of instantiation from the potential pole that we focus on in the systemic description and the instance pole where we observe, sample and analyse texts: we use the description to analyse a growing number of texts that we sample in order to expand the description, checking it along the way and identifying gaps in the description (cf. Figure 7 below).

Using the function-rank matrix and other aspects of the architecture of lexicogrammar located within language in context, we can be clear about the criteria for including and excluding areas to describe (cf. Section 1.4 above); in this way, we can engage in systematic description — in addition to systemic description, just as we when we use *criteria for inclusion and exclusion* in developing systematic literature reviews. (Very often when I consult reference grammars of various languages, I find that they do not actually address the fundamental issue of what to include (or not to include) in a comprehensive description: there is no clear sense of what would constitute a comprehensive description, nor indications as to how to check whether the description on offer is comprehensive or not. It is worth noting that reference grammars are typically not written to support the systematic analysis of text in context — which is actually the most robust way of testing the coverage of the description, provided of course that we analyse a reasonably large sample of texts instantiating quite a varied range of registers¹⁰. In contrast, the description of English in Halliday & Matthiessen 2014, is explicitly designed to support the lexicogrammatical analysis of texts, and it has been tested in innumerable projects involving text analysis.)

¹⁰ This is a test that Halliday's description of the lexicogrammar of English in his *Introduction to Functional Grammar* has been subjected to for around four decades. I doubt that there is any other description of any languages that has actually been tested as extensively — in the sense of being "confronted" with new texts in analysis by researchers all around the world.

In the description indexed in the function-rank matrix, the systems are often extended only a few steps in delicacy; but if the linguist(s) developing the description have sufficient support, they can continue to be extended in delicacy (as discussed for the English system of TRANSITIVITY in Matthiessen 2014b). This is in fact a manifestation of the general approach to the development of descriptions of languages in SFL: we start by trying to cover the whole language — or, say, the whole lexicogrammar — at low delicacy; then we gradually increase the degree of delicacy of the description, as shown schematically in Figure 5 — the holistic approach to the development of comprehensive descriptions of particular languages, involving gradual approximation along the cline of delicacy¹¹.

In the case of the development of descriptions of lexicogrammars, the move towards increasing delicacy will gradually make contact with any descriptions that have been produced to take account of "constructions" in the language under description and with lexical descriptions (both of which may be based on the automatic analysis of corpora). For example, in Matthiessen (2014b), I show how the description of the system of TRANSITIVITY that takes the grammatical zone of lexicogrammar as its starting point can be related exhaustively to the "verb classes" documented by Levin (1993)¹².

2.3. Holistic approach to description: gradual approximation

The holistic approach characteristic of SFL contrasts quite sharply with the approach that has tended to be dominant in linguistics (and also in other sciences), viz. Cartesian Analysis (see e.g. Halliday & Matthiessen 2006; and cf. Capra 2006; Capra &

¹¹ The representation of corpora and corpus tools are geared towards lexical studies (cf. Halliday 2002), so using corpus-based methods, we can complement the move of gradual approximation in delicacy with a focus on lexical islands or fields; but they this approach will only be truly significant when such investigations can be located within the overall low-delicacy map of the lexicogrammatical resources. At the same time, it is important to note that the most frequent items revealed by corpus-based methods will be grammatical ("function words") rather than lexical items ("content words"). ¹² Her "verb classes" are really fragments of the transitivity structures of clauses — potential "constructions" in the sense of "construction grammar" (e.g. Hoffman & Trousdale 2014).

Luisi 2014; Noble and Noble 2023). In the holistic approach, we try to achieve descriptive coverage of the whole semiotic space of lexicogrammar (or phonology or whatever subsystem is under description). We begin by sketching the systemic outlines at a low degree of delicacy, trying to achieve **metafunctional spread** and **rank-based compositional coverage**. Then, we increase the coverage gradually by extending the description in delicacy, as shown in Figure 5^{13} . If we are basing the description on a corpus, increasing the coverage will of course place higher demands on the size of the corpus, as well on its registerial composition.

In contrast, in linguistic description informed by Cartesian methodology (cf. Descartes' 2006, 17 [1637]), linguists focus on a small area within the overall space, drilling down in detail in order to address questions about language articulated within the theory¹⁴. In the Chomskyan traditions, such questions originate in his attempt to identify (certain parameters of) UG, Universal Grammar, as a way of supporting the rationalist position in the long-standing debates framed within Western Philosophy, more specifically, Western Epistemology. The fundamental failure of this research programme has been well-documented, thoroughly and succinctly by Evans & Levinson (2009).

¹³ I have based this figure on one Michael Halliday drew for Bill Mann and me during a lunch meeting in Marina del Rey in the late spring of 1980. Fortunately, the restaurant (the Warehouse) had large place mats made out of paper, furnishing the necessary material for semiotic depictive expansion.

¹⁴ The areas considered theoretically interesting over the decades have varied considerably. When Paul Schachter taught a group of us field methods over two semesters in the mid 1980s, the language selected was Akan. We had a wonderful Akan consultant, a PhD student in geology at UCLA from Ghana. My fellow students would work him hard to try to elicit information about unbounded dependencies, since they were fashionable in generative linguistics during this period. Two of them were from Kenya, and sensed that what they were doing was not very useful in their context but explained to me that they had to pursue this line of investigation to honour their scholarships. At the end of the year, the consultant asked me if he could have a copy of my work on the lexicogrammar and phonology of Akan, telling me he understood what I was trying to achieve but had no idea what the others were up to.



Figure 5: Comprehensive description of a particular language as a gradual increase in delicacy

As I have shown, the moves in the development of descriptions of particular languages are derived from the theoretical architecture of language in context, e.g.

- **instantiation:** the move from texts at the instance pole of the cline of instantiation towards the potential pole, where the system to be described is located; in terms of linguistic activities (cf. Figure 1 above), this means analysis in support of description;
- **delicacy:** the move from low delicacy towards increasing delicacy; within lexicogrammar, this means moving from grammar towards lexis (covering "constructions") along the way;
- **metafunction:** the move across the metafunctions, possibly using one of them (often but not necessarily, the textual metafunction) as the "way into" the general description;
- **rank:** the move down the rank scale, from the most extensive environment (the clause, as the gateway to the semantics of text), towards lower-ranking units, interpreting them contextually in reference to their roles in higher-ranking units.

In terms of all of these dimensions, the phenomena under investigation are thus always viewed *contextually*, i.e. *in relation to the environment in which they can be located*. This is akin to the ecological approach to the study of biological systems. At the same time, when we describe a particular language, we need to keep **shunting** along the different semiotic dimensions that make up the overall architecture of language in context. The methodological principle of shunting was identified already by Halliday (1961). By training ourselves to shunt along the different semiotic dimensions, we ensure that we do not get stuck with a single view of the linguistic patterns that we are trying to make sense of: we shunt according to the principle of **trinocular vision**, thus ensuring that we observe, sample, analyse and describe patterns in a well-rounded way, varying our angles of vision.

For example, while the primary move along the cline of instantiation will be from instance to potential — from patterns in texts at the instance pole that we try to infer systemic generalizations on the basis of, we will also need to move in the other directions, starting with hypotheses based on the current description of the system and check them against texts at the instance pole (an approach that can be characterized as abduction, supplementing

deduction and induction: see e.g. Matthiessen & Teruya 2024) — quite possibly asking language consultants to construct examples for us (e.g. based on translation from the language that serves as the medium of investigation): this is what I have called **paradigm probing**: see Section 3.1.3 below.