Hanna Rutkowska

Orthographic Systems in Thirteen Editions of the *Kalender of Shepherdes* (1506–1656)

35

Chapter One
Theoretical preliminaries:
Orthography and corpus linguistics

1.0 Introduction

This chapter provides definitions and brief discussions of orthography and related terminology. It deals with the theoretical assumptions forming the basis for the present study, with the elements of comparison between the most important theoretical approaches to speech and writing, and their relation to language. Particular attention is paid to the so called autonomistic approach to writing systems and written language, which is assumed in this study. It also discusses reasons for treating orthographic systems and orthographic variation as proper objects of linguistic investigations. This chapter is closed by a section devoted to the definition and assumptions of corpus linguistics as well as to the main characteristics of the corpus compiled for the purposes of the present study.

1.1 Approaches to language, speech and writing

Before proceeding to the definition of orthography, it is worth considering such broader terms as *language*, *speech*, and *writing*. Writing (or script) "may be defined as a system of visual symbols whose purpose is to convey the thought of one individual or group to another" (Crossland 1956: 8). Analogically, speech can be treated as a system of phonic symbols with a similar purpose. Thus, both speech and writing are used to express ideas and feelings, as well as to transmit information, so they can be called different means or channels of communication. The differences between speech and writing comprise several aspects. For example, the former is temporary, spontaneous, and irreversible, whereas the latter is permanent, planned, and editable.[13] However, they both remain in a symmetrical relation to language, which is a term denoting a

---

13 Discussions of these differences can be found, e.g. in Crystal (1995: 291–93) and Cook (2004: 31–53).

complex system inherent in the human brain, allowing for spoken and written communication among human beings. This system consists of words which are interpretable because they are used within a regular morphosyntactic structure and with the observance of co-existing spoken and written norms.[14] These norms both make up subsystems of language elements, but one is manifested phonically and the other graphically (Vachek 1973: 15–16).

There exist two main approaches to the notions of speech and writing, referred to as *relational* and *autonomistic* by Sgall (1987: 2–3; see also Ruszkiewicz 1976: 37–44).[15] The former equates speech with language, and treats writing as an extra-linguistic phenomenon. Sapir's statement that "[t]he written forms are secondary symbols of the spoken ones – symbols of symbols" (Sapir [1921] 2006: 10; see also Lyons 1968: 38) formulates the main assumption of the relational approach.[16] It was adhered to by many leading structuralists and post-structuralists in the twentieth century, e.g. de Saussure, Bloomfield, Hockett, Sapir, and Lyons, whose opinions strongly influenced a large number of linguists, both their contemporaries and successors (see e.g. Kohrt 1986: 81). Venezky blames Bloomfield in particular for the successful promotion of "the view that writing is secondary and subservient to speech" (Venezky 1970: 27).

The arguments believed to confirm the superiority of speech over writing include, e.g., the chronological primacy of speech, from which writing was originally derived, the fact that children acquire speech without formal education before they eventually learn to write, and the fact that there exist languages which never developed a writing system (see, e.g., Bloomfield [1933] 1961: 21, Smith 1996: 56). These claims seem difficult to deny, but some scholars point to the fact that they should not be used as arguments to strip the orthographic system of its autonomy in its relations to other levels of the language (Ruszkiewicz 1976: 19, Liuzza 1996: 28). Members of the linguistic school of Prague, most conspicuously Josef Vachek, but also scholars associated with other intellectual centres, e.g. Bolinger, Stetson, McIntosh, McLaughlin, Haas, and Kohrt, belong to the best known supporters of the autonomistic approach to writing in the twentieth century. They claim that although writing cannot and should not be considered as totally unrelated to speech, its primary

---

14 This definition constitutes a modification of the ones in Rogers (2005: 2) and *OED3* (*s. v.* "language, *n.*").

15 Coulmas proposes the division into *representationalistic* and *distributionalistic* theories, which are roughly equivalent to the *relational* and *autonomistic* ones, respectively (1999, *BEWS, s. v.* "grapheme").

16 See also Rutkowska (2012: 225–26) for more details and quotations concerning the relational and autonomistic approaches.

purpose is not to represent a spoken utterance, but to be a codified manifestation of language in visible signs, which does not require decipherment into speech to be understood (Vachek 1982: 38). Instead of recording speech and constituting a continuous transcription process, the function of writing is "to communicate linguistic utterances in a different mode from speech" (Stenroos 2002: 453), with the assumption that "utterances" are of two kinds, spoken and written, where the latter "cannot be simply regarded as optical projections of the former" (Vachek [1945-49] 1976:132). This symmetry between speech and writing in their relation to language is also emphasised by Haas, who explains that "[s]poken words consist of individual sounds; and written words, of individual letters; and between the two kinds of element there is correspondence but not reference" (Haas 1970: 15).

Even if we assume that originally written utterances were "symbols of symbols", or spoken utterance put down in writing, "as soon as some writing tradition develops in the given language community, a strong tendency asserts itself towards the establishment of a direct link between the written utterance and the external reality referred to by that utterance" (Vachek [1972] 1976: 140). Vachek's opinion confirms the one formulated by Henry Bradley towards the end of the second decade of the twentieth century that "written language has to a great extent become an instrument for the direct expression of meaning, co-ordinate with audible language. The result of this has been that the written language has in part been developed on lines of its own, independent of the development of oral speech." (Bradley 1919: 15).

The autonomists point to the fact that not all the characteristics of writing systems are present in speech (see e.g. Stetson [1937] 1981: 35, McIntosh [1974] 1989: 46, Stenroos 2002: 456, Cook 2004: 32). These characteristics are most evident in the predominantly morphographic (or logographic) writing systems where the graphic symbols refer to morphemes and words, not to the phonemes of a given language, e.g. in Chinese and Sumerian (see Sampson 1985: 145–71, Rogers 2005: 272), but can also be found in phonographic writing systems (in most European languages), where the correspondence between the spoken and written norms is clearer. Characteristics unique to writing include, among others, the differentiation between homophones (e.g. *seen – scene*) and in the consistent spelling of the stems in heterophonous, but etymologically and semantically related words (e.g. *south – southern*).[17]

As a consequence of treating the written language on a par with the spoken one, autonomists consider it to be equally eligible for linguistic study. For example, Crossland claims that linguistics "should include the study of written

---

17   See Section 1.4 for more examples.

languages as well as that of spoken; the former study can and should be as scientific as the latter, and it needs its own terminology which should be basically independent of that of the study of spoken languages" (Crossland 1956: 8). A similar view was formulated by McIntosh in the same year. He asserted that "no position is satisfactory which fails to accord equality of status to spoken and written language and (...) the analysis of one is neither more nor less linguistic than the other", and also that "the problems involved are not likely to be less complex than those which confront the student of a spoken language, but (...) being in many respects very different, these problems of written language must be met with techniques specially devised to overcome them" (McIntosh [1956] 1989: 7, 10). The implication of such views is the need to apply as rigorous an analysis to the written language as is applied to its spoken counterpart. Coulmas asserts that the regularities of written language are now generally considered to constitute an independent level of linguistic structure, and the study of these regularities is construed as an independent subdiscipline of linguistics (*BEWS*, 1999, *s. v.* "graphemics").

1.2 Definition of orthography and related terms

According to *OED3* the word *orthography*, derived from Greek, appeared in English in the fifteenth century. In classical Latin *orthographia* meant "correct spelling" (Quintilian). In the sixteenth century, with reference to language, two additional senses were added, including "a system of spelling or notation" and "spelling as an art or practice; the branch of knowledge which deals with letters and their combination to represent sounds and words; the study of spelling" (*OED3*, *s. v.* "orthography, *n.*"). The earliest meaning of this word can be expanded to "a spelling norm which consists of all the standardized and codified graphic representations of a language" (Rutkowska – Rössler 2012: 214). A spelling norm would be hard to imagine without reference to an orthographic system, which implies that the short definitions quoted above are, to some extent, overlapping. This relation between the norm and systematicity are adroitly expressed in another definition of the term, which reads: "[t]he standardized writing system of a language is known as its orthography" (Crystal 1995: 257).

The term *orthography* is also related to and partly synonymous with other terms. One of these is *spelling*. The latter is defined as a "[m]anner of expressing or writing words with letters; orthography" as well as "a particular instance of this; a special collocation of letters representing a word" (*OED2*, *s. v.* "spelling, *n.*"). Here we encounter some terminological overlap and even inconsistency.

On the one hand, we can consider spelling to cover the graphic realisations of all words, and in this sense orthography, as a binding norm, would be part of the spelling of a given language. On the other hand, orthography understood as a writing system comprises not only the spelling of particular lexical items and morphemes, but also capitalisation, word division at the end of lines, and punctuation. Eventually, taking into consideration the third sense of "orthography", it is also used as a term denoting a branch of knowledge which studies all aspects of a spelling norm, including i.a. the relation between writing practices beyond the norms, codification, and language policy (Rutkowska 2012: 226, Rutkowska – Rössler 2012: 214).

This apparent inconsistency and complexity inherent in the word *orthography*, does not preclude its employment in discussions of different aspects of writing systems, because the exact meaning can usually be derived from the context, even though it may occasionally lead to confusion or misunderstanding. However, one needs to be aware of several meanings of this term in order to appreciate and compare particular scholars' claims and observations. The semantic complexity of the term under consideration is also closely connected with the fact that the writing system or orthographic system of any natural language acts as the interface of different linguistic levels, including phonology, morphology, lexicon, and even syntax and semantics.

Graphemics is another term related to *orthography*, and functioning as its near-synonym. It first appeared at the beginning of the 1950s, by analogy to *phonemics* (see Pulgram 1951: 19, Stockwell – Barritt 1951). In *OED2* it is defined as the "study of systems of written symbols (letters, etc.) in their relation to spoken languages" (*OED2*, *s. v.* "graphemic, *adj.* and *n.*"). However, according to some linguists "the term graphemics should be confined to the study of systems of writing only" (Bazell [1956] 1981: 68). A new term was proposed, i.e. *graphophonemics*, to refer to the "discipline concerned with the study of the relationship between graphemics and phonemics" (Ruszkiewicz 1976: 49). Another term is *graphotactics*, denoting a study of the syntax or combinations of graphemes (units of a writing system) as well as the laws governing them (Haas 1970: 59, Vachek 1973: 9, Carney 1994: 66–69). Here, again, some terminological overlap can be identified, because according to another view *graphotactics* is a synonym of *graphemics*, because the latter can be defined as the "linguistic study of writing systems based on a description of their elements and the graphotactic rules specifying the systematically permissible combinations thereof" (Coulmas 1999, *BEWS*, *s. v.* "graphemics").

1.3 Units of a writing system

The term *grapheme* replaced the classical notion of *littera*, introduced by the Stoic grammarians. *Littera* denoted the smallest element of language, combining three attributes, *nomen* 'name used for identification', *figura* 'shape or visual configuration', and *potestas* 'power to signify sound' (Abercrombie 1949: 59; Henderson 1985: 142, Rutkowska 2012: 229–30). The word *letter*, etymologically related to the Latin *littera*, is now in use mainly in everyday language, referring to a character of an alphabet, and also in definitions of terms connected with the description of writing systems, but some linguists consider this word unworthy of mention among "full-blooded linguists" (Kohrt 1986: 80).

Grapheme was first mentioned in the writings of Baudouin de Courtney in 1901 (Ruszkiewicz 1976: 24–37, 1981 [1978]: 20–34, Kohrt 1986: 82). In Western linguistics it is recorded in Stetson, who defines it simply as "[t]he unit of writing" (Stetson [1937] 1981: 35). Since then, grapheme has been defined in different ways, but these definitions can be divided into two main groups, related to their makers' theoretical approaches to writing systems. To the supporters of the relational approach to writing a grapheme is "the class of graphs which denote the same phoneme" (Hammarström [1964] 1981: 97) or "[t]he class of letters and other visual symbols that represent a phoneme or cluster of phonemes" (*OED2*, *s. v.* "grapheme, *n.*). In contrast, for the adherents to the autonomistic approach it would be defined as "the minimal functional distinctive unit of any writing system" (Henderson 1984: 15) or "a purely distinctive visual unit, part of an autonomous semiotic system" (Liuzza 1996: 28).

Since the term *grapheme* was coined by analogy to *phoneme*, it has often been compared to its model notion. Pulgram compiled a list of parallel features of graphemes and phonemes, and terms related to them such as *graphs* and *allographs*. He defined the graph as "[t]he hic et nunc written realization of a grapheme", and stated that "all graphs identifiable as members of one grapheme are its allographs" (Pulgram 1951: 15–16). Other definitions of a graph include, e.g. "a written character, modification of a character, or feature of arrangement in a particular segment of a particular document" (McLaughlin 1963: 29), and "a materialization of a certain grapheme in a concrete written utterance" (Kohrt 1986: 90). By comparison, for the supporters of the relational approach to writing systems, a graph would be rather "[a] visual symbol representing a phoneme or a segment or feature of speech" (*OED2*, *s. v.* "graph, *n.*[3]"), and an allograph "[e]ach of two or more letters or letter combinations representing a particular phoneme" (*OED3*, *s. v.* "allograph, *n.*[2]"). However, even those scholars who support the latter approach should be aware that defining a grapheme as a relational unit raises problems whenever "the distinction

between two words lies in a difference in graphemes that has nothing to do with phonological reference" (Liuzza 1996: 28), as in the case of homophones, e.g. *site* and *cite*.

The analogy between phonemes and graphemes has also been noticed by Vachek, according to whom "the graphemes of a given language – like its phonemes – remain differentiated from one another, i.e. (…) they do not get mixed up" (Vachek [1945–49] 1976: 128–129). Moreover, he claims that there are some limits to this analogy, because "the structuration of phonemes is indeed not paralleled by that of the graphemes, but this fact does not disagree with our basic thesis asserting the necessity of structural correspondence of the spoken and written norms on SOME language level" (Vachek 1973: 35). This view emphasises the autonomy of the writing system.

Haas (1970: 51) remarks that "[c]orrespondence between phonemes and graphemes, if it is not one-to-one, can be either one-many or many-one or many-many". The complex correspondence between graphemes and phonemes has been the subject of many scholars' deliberations. One of the strategies aiming to solve the theoretical problem posed by the description of the relations between the units of the phonological and graphemic systems has been the use of the terms "complex grapheme" as opposed to "single grapheme" (see, e.g. Hall [1960] 1981: 71). This is characteristic of the typically relational approach to writing systems where "[o]ne is forced to distinguish a series of discrete graphemes from a 'compound grapheme' only by reference to the phonic system" (Liuzza 1996: 31), so e.g. in *partition* the first <ti> would be a series of graphemes and the second one a "compound grapheme". Instead, Liuzza prefers to talk about series of graphemes forming "orthographic sets" corresponding to single phonemes (Liuzza 1996: 31, see also Haas 1970: 49). This seems more acceptable from the point of view of the autonomistic approach to writing.

Laing (1999) went even further, abandoning the division into graphemes and phonemes, inspired by the ancient concept of *littera*, and introducing the notions of *litteral substitution sets* (LSS) and *potestatic substitution sets* (PSS), denoting sets of orthographic variants corresponding to sounds, and sets of phonological (or phonetic) interpretations of spellings, respectively, when examining particular lexical items found in medieval texts (see also Laing – Lass 2003, 2006: 431–32, *LAEME*). However, such a complex description apparatus, which has proved very useful in the investigation of Middle English texts, does not seem necessary for the description of Early Modern English.

In descriptions of correspondences between graphemes and phonemes also the terms *consonantal* and *vocalic* (functioning as nouns) have been used with reference to graphemes corresponding to consonants and vowels, respectively (Wełna 1982: 10).

1.4 Orthographic principles in the English writing system

A set of rules governing orthographic systems was proposed already in the 1870s by Jan Baudouin de Courtenay. They included three principles: the phonetic, etymological, and historical, relating orthography to what this linguist believed were its three determining factors, pronunciation, origin (covering the morphemic structure of words), and tradition (referring to those cases which cannot be explained by the first two principles) (Ruszkiewicz [1978] 1981: 24–25; Sgall 1987: 2–3). The idea of orthographic principles operating at different levels has been further elaborated upon and modified by numerous linguists. For example, Firth (1935: 61) discussed the notion of "polysystem", or system of systems, referring to the co-existence of and interaction between phonological, grammatical, and lexical systems of orthographic representation. Albrow (1972: 3–51) divided the English orthographic system into three subsystems, "basic" (comprising native words), "Romance" (loanwords from Romance languages), and "exotic" (other loanwords).[18] Overviews of the investigations on the interaction of principles in English orthography and in other orthographic systems can be found, e.g., in Ruszkiewicz (1976), Sgall (1987), Liuzza (1996), Rollins (2004: 27–40), Rutkowska (2012: 227–229), and Rutkowska – Rössler (2012: 215–16).

In Present-Day English the following orthographic principles can be identified as the most important ones:[19]

(a) The phonological (phonemic) principle. From the typological point of view, the English writing system can be classified as "phonographic", and more precisely a "phonemic" one, because its graphemes correspond predominantly to the phonological units of the language (or phonemes), including consonants and vowels (Rogers 2005: 272–73, Rutkowska 2012: 227–29). This type of system is also referred to as "alphabetic" (Venezky 1999: 4). The phonemic principle implies "biuniqueness", which "requires not only that a given phoneme is represented by a constant symbol but also that the symbol involved does not represent other phonemes" (Carney 1994: 15). However, this principle is far from being consistently observed in English (in contrast to e.g. Greek and Finnish). For example, the phoneme /iː/ can correspond to different graphemes (and their sequences), including <e> (*be*), <ee> (*meet*), <ea> (*meat*); whereas

---

18   A division of English words into several types with separate spelling rules depending on their origin can be found already in Craigie (1928).

19   See Rutkowska – Rössler (2012: 215–16) for a more comprehensive list, not limited only to English.

the grapheme <i> can correspond to different phonemes, including /aɪ/ (*line*), /ɪ/ (*wind*). The following principles explain the most significant types of divergence of the English orthographic system from the phonemic principle.[20]

(b) The etymological principle, according to which morpheme constancy should be respected. This concerns mainly the etymologically related words in which the shared root is spelt the same, e.g. in *nation* and *national*, as well as *sign* and *signature*, in spite of different pronunciation. This principle has also been referred to as a "morphophonemic" principle (see Hall [1960] 1981: 74).

(c) The historical principle, which preserves the graphic forms reflecting earlier stages in the correspondence between spelling and pronunciation, e.g. in *do* the grapheme <o> corresponds to the Middle English pronunciation of the vowel in this word. This principle also refers to the so called etymological (or classicising) respellings, e.g. *debt* and *doubt*, where <b> was introduced in the Early Modern English period in order to make these forms more similar to their Latin etymons irrespective of their contemporary pronunciation (*OED2*, *s. vv.* "debt, *n.*", "doubt, *n.¹*"). Equally, the historical principle concerns "pseudo-etymological" respellings, where the modification has been applied erroneously, and not justified by historical etymons, e.g. the addition of <d> in *advance*, as well as reverse spellings (or back spellings), e.g. *could*, where <l> was introduced by analogy to *would* and *should*, even though it was not pronounced. In some cases such respellings led to changes in pronunciation, known as *spelling pronunciation*, e.g. in *adventure*, where the epenthetic <d> added in Early Modern English started to be pronounced as /d/. The historical principle has also been referred to as a "lexical" principle (see Chomsky – Halle 1968: 49).

(d) The principle of heterography which requires a visual distinction between homophones, e.g. *right – rite – wright – write*. A certain degree of homography can be found in English, e.g. *bow* can refer to several lexical items with different meanings, including 'a bend; a weapon for shooting arrows', 'an inclination of the body or head in salutation or in token of respect', 'the fore end of a ship or boat' (*OED2*, *s. vv.* "bow, *n.¹*", "bow, *n.²*", "bow, *n.³*"). The principle of heterography also asserts a graphic distinction between common and proper nouns, where the latter are spelt with the word-initial capital (compare *archer* and *Archer*).[21]

---

20 See Haas (1970: 4, 51) and Carney (1994: 15) for more discussion on phono-graphic divergence from biuniqueness.

21 Compare Rutkowska – Rössler (2012: 216), where the capitalisation of proper nouns is alternatively considered an example of a semantic principle. Moreover, linguists'