



Language in Digital Motion: From ABCs to Intermediality and Why This Matters for Language Learning

Heather Lotherington

Contents

7.1 Introduction.....	217
7.2 Language and Literacy.....	219
7.3 The Expanding Borders of Language in Digital Communication.....	222
7.4 Theorizing Multimodal Communication: Two Views.....	225
7.5 Modality, Mode, and Media in Digital Communication.....	227
7.6 Conclusion: From ABCs to Intermediality.....	235
References.....	236

7.1 INTRODUCTION

The context with which this chapter grapples is *mobile language learning*. Recent research (Lotherington 2018) indicates that top commercial mobile (m-)learning apps tend to rely heavily on outmoded structural

H. Lotherington (✉)
York University, Toronto, ON, Canada
e-mail: HLOtherington@edu.yorku.ca

models of language and tap dated behaviouristic pedagogies in their lessons, typically applying gamification veneers to attract and maintain users. Popular m-learning apps predominantly feature levelled vocabulary study, and to a lesser extent, language structure, aka *grammar*. Linguistic communication, however, has moved from simple alphabetic encoding to multimedia design in digital environments; this challenges the fit of structural theories of language to digital language learning contexts. In language teaching and learning literature, *multimodality* describes communication employing diverse semiotic resources in texts not limited to alphabetically (or logographically or syllabically) encoded language. In reality, all communication is multimodal; however, professional awareness of *multimodal communication* in language teaching emanates particularly from digital texts and discourses, which are created using a broader palette of meaning-making resources than static print texts, ergo, multimedia texts. The predominant trend has been to approach digital multimodal communication from a *social semiotics* paradigm (e.g., Bezemer and Kress 2008, 2016; Jewitt 2008; Kress 2000, 2003, 2005, 2009, 2011).

Shortly after the turn of the century, Kress (2003: 1) predicted: “the combined effects on writing of the dominance of the mode of image and of the medium of the screen will produce deep changes in the forms and functions of writing.” *Mode*, though, is not defined. Jewitt (2004: 84), clarifies the concept of *mode* as indexing “technologies of representation (the modes of ‘multimodality’),” and contrasts *mode* with *media*, which indexes “technologies of dissemination (the media of multimedia).” As social semiotic conceptualizations of multisemiotic composition have grown in concert with increasingly sophisticated, grammatically differentiated digital communication, the terrain of multimodal communication can be seen to overlap considerably with that of intermediality theorizing. The fundamental building blocks of *mode* and *media*, which, in a social semiotics reading based in linguistics, rely largely on cultural interpretation and exemplification, gain precision from an intermediality analysis based in interart studies.

Elleström’s landmark intermediality model (2010, 2020) delineates the concept of a *mode* from a semiotic perspective, illuminating how *modality* characterizes *media*, a generally nebulous and capacious concept. Elleström’s model offers a valuable resource for examining *multimodality* in digital communication from an innovative theoretical stance grounded in art, rather than linguistics.

Taking an intermediality lens, this chapter parses two selected features of mobile digital communication that are only vicariously captured in linguistic analysis. The analysis draws on Elleström's (2020) intermediality paradigm to help delineate how diverse meaning-making resources could be conceptualized from a perspective decentred from linguistics and leveraged in reconceptualizing the m-learning of language as it is used in contemporary interactive, multimedia texts and discourses.

7.2 LANGUAGE AND LITERACY

Language has traditionally been described as a *medium*: of communication, and of learning. Language is, in fact, an abstract until it is materialized: mediated physically, in speech and signed conversations, and technologically, in printed documents, social media sites, roadside signs, movies, games, and suchlike. Gershon and Manning (2014: 539) point out that curiously “the mediality of language is rarely explored in media studies.”

Language is often complexly remediated. Ong (1982) theorized *secondary orality* as the remediation of speech into writing with subsequent re-voicing, as in scripted televised newsreading (also see Eide and Schubert 2020). Have and Pedersen (2020) describe contemporary remediation in books: traditional media products that can be remediated as audiobooks and *read* aurally. In this chapter, remediation of the human voice is described as material in artificial intelligence (AI) programs that create conversational digital agents.

Language is innate to human beings, hardwired as individual cognitive capacity, though the developmental trajectory of the child learning to speak (or to sign in the case of deaf children) requires appropriate socialization to activate. Languages, thus, live in society as well as in the minds of individuals, social use constituting their lifeblood.

Speech is dynamic in character, spoken language evolving in tune with the social community/ies in which it has currency. Different accents signify regional location in spoken language populations, for example, English as spoken in Toronto, Canada, as compared to English as spoken in Edinburgh, Scotland, or in Cape Town, South Africa. Literate norms, though, tend to be much more fixed, coded into a body of literature relying on common conventions that developed for the social expectations and the mediating technologies of the time of publication.

The technological interface encoding language into literate forms for most of the history of mass literacy was the printing press, which, through pioneering use, forced the conventionalizing of spelling and sentence mechanics to make printed literature easily readable. Times have changed, however, and the predominant canvas for communication is now the screen, and increasingly, a screen that is individualized, mobile, and ubiquitously wifi-connected. Communication is now ineluctably and indelibly multimodal.

7.2.1 *The Literate Bias of Education*

In education, there is a distinct bias towards literate learning. At the grade school level, literacy is seen as the keyhole through which school learning progresses. In the province of Ontario (Canada) which is the policy jurisdiction for children in Toronto, *language* is conflated in curriculum documents with *literacy*. As such, literacy references language written down. This makes for a messy transition to digital multimodal communication.

In the case of sequential language learning (language learning following initial child language development), sometimes referred to as *second (or foreign) language learning*, the characterization of language competence in terms of four skills: *speaking–listening–reading–writing*, ostensibly mirroring child language development (as theorized in pre-digital, majority language, middle-class contexts) has tenaciously persisted for decades. Though the validity of discrete language skills (reading as separable from writing, and so forth) has been in contention for decades, formal language testing in gate-keeping tests, such as the TOEFL has been complicit in cementing a four-skills model into social and economic benchmarks, so this limited twentieth-century thinking about what language comprises and entails for learners continues to be reinforced in many teaching contexts.

7.2.2 *Mobile Language Learning*

Mobile language learning is geared to second and foreign language learners. Despite myriad possibilities in using mobile devices for language learning, the evolution of m-learning tilted towards app-based learning following the release of the iPhone in 2007. Apps are third-party software packages that are directly downloadable to mobile devices on a trial cost-free or low- cost basis. Apps for language learning take a variety of generic

approaches from language courses to games to vocabulary resources to memorization drills (Lotherington 2018), but each offers a proprietary package, marketed by a software developer. The digital marketplace for language teaching apps is a veritable wild west of unregulated products, educationally speaking, in which *learners* are treated as *users*. The user is, in fact, the product in app design, their learning behaviours, collected and sold as *data*, though there may also be a more direct profit motive in paid course upgrades, unlocked learning resources, and so forth.

Mobility in the context of digitally mediated communication references capacity for ubiquitous interactive communication untethered to physical location. Contemporary m-learning uses smart mobile devices that embed powerful multifunction toolkits for digital text making. Because of this, mobile connection has the potential to invite dynamic pedagogical approaches that put agency in the hands of the learner for collecting and creating resources towards directed language learning, contextualized in space. Mobile devices, however, frequently meet with disapproval in formal education sites, if traditional curricular agendas of language and literacy learning dominate. Worse, smartphones are banned in some schools, precluding creative classroom m-learning.

Top-selling apps for language learners, recently surveyed and road-tested (Lotherington 2018), revealed a tendency to import dated structural models of language and behaviourist pedagogies into the fluid mobility and complex functionality of a smart device. Features of current interactive multimedia communication are used here and there; most apps apply a gamification engine to mask the tedium of vocabulary drills. Some apps also utilize a social media feature to connect random chat partners selected on the basis of the home country, which is an unreliable indicator of language proficiency in any case. However, generally speaking, social media affordances were poorly utilized and digital discourse forms largely avoided.

This chapter presents data from an extended review of the research literature on how language has morphed in form and function in co-evolution with technological change. This literature review forms the basis of an active exploratory pedagogical design study to build mobile production pedagogies (Thumlert et al. 2015) for language learning that utilize the powerful resources of smart devices and activate a contemporary palette of semiotic resources. Our research team's aim is to invite interactive multimedia textual composing in ways that work agentively for language learners. To do this, we need to understand the kinds of changes language

has undergone and how we can document these changes from a theoretically rigorous standpoint. Elleström's (2020) intermediality paradigm offers analytical specificity and categorization that is helpful in understanding contemporary communication from a perspective not grounded in linguistics that can be intelligently merged with what we understand about how language works to create meaning.

7.3 THE EXPANDING BORDERS OF LANGUAGE IN DIGITAL COMMUNICATION

Ignoring the paralinguistic information embroidering the borders of spoken language (e.g., body language, facial expressions, and tone of voice) was pedagogically reinforced in both school and second language learning. A related principle was held in books where nonlinguistic material, such as illustrations and charts, was acknowledged but looked at as supportive of rather than integral to overall meaning contribution. This was particularly the case in second language learning, where the focus was traditionally on the structural elements of language. In grade schools, where language is more holistically approached, books were nonetheless assessed as to content and nature according to the density of language carrying, as it were, the message; image-centred genres, such as comic books, were derided as insufficiently serious. Nowadays, the graphic novel has taken on a new life, imparting content taught in grade school and university classes, for example, *Persepolis* (Satrapi 2003).

Digital mobile connection has redrawn the literate borders of language, challenging the print-centred skills of second language teaching and learning, and the linear linguistic encoding of text. With Web 2.0 technological upgrades, enabling interactivity, literacy has morphed from discrete reading and writing of the static page into a *multimedia read/write (R/W) capacity* underpinning social media posting, collaborative authorship (wikis); tweeting (microblogging), texting, and so forth. Furthermore, the shift from page to screen entailed radically new encoding capabilities that have dislodged the stuff of literacy from the letter to the pixel, which indifferently encodes still and moving images, sound files, and spoken language (Cope and Kalantzis 2004). The limits of theories focused on written instantiations of language: linguistic structuralism and literacy theories, focused on linear encoding and decoding of the static page, have become

inadequate to imagining the task of understanding and creating multimedia textual products.

Though the evolution of page to screen has taken place over decades, the period of most rapid change in language form and function has taken place over the past 15 years, consistent with Web 2.0 and 3.0 technological advancements. The initial iteration of the public World Wide Web in 1991, retroactively referred to as WWW 1.0., was essentially a digital bulletin board where content could be publicly posted. It was Web 2.0 circa 2004 that opened the flood gates to rapid co-evolution in social practices, economic opportunities, cultural life, and associatively, language form and function. Web 2.0, or the *semantic web*, gave birth to social media forums, for example, Facebook, LinkedIn; collaborative authoring tools or wikis; video-streaming services, such as YouTube; and novel sites for information sharing and social networking, such as blogging, podcasting, and microblogging, for example, Twitter. The evolving capacity for integrated artificial intelligence in Web 3.0 enabled *smart technologies* that yield new functionalities. These include *conversational digital agents*, AI programs that use natural human language in spoken form to respond to user voice or text inquiries in environments such as mobile phones, business call centres, and GPS systems in cars.

7.3.1 *DIY Language Norms and Conventions*

Historically, lexical and grammar conventions were based on literature, which, having met publication standards, provided the guide rails for language standards. Dictionaries and grammar books encapsulated accepted spelling, grammar, punctuation, and lexis. Figure 7.1 provides an amusing excerpt from an eighteenth-century grammar of English, which would be met with general hilarity by even the most tolerant of language teachers today.

Fig. 7.1 An excerpt from Henson's grammar (1744)



RTHOGRAPHY is that Part of Grammar, which teacheth to write Words with proper Letters, divide Syllables, and whatever else pertains to exact Writing.

A Letter is a Mark or Sign of a fingle Sound.

The established quality controls (and built-in biases) of commercial publication, however, are far less influential in an interactive digital environment enabling self and open publication, collaborative authoring and editing, and user-developed generic conventions, co-evolving with technological affordances, yielding novel spelling norms in texting (*txtng*), for example, brb, omg; tweet grammars (280 characters, incorporating semantic punctuation such as # and @ as well as other digital media), and suchlike. These new forms challenging the standard spelling, punctuation, and grammar relied on in schooling, however, are *user-driven*, emerging, ironically, from the same historical basis as literary and grammatical standards: *publication*. This creates a conundrum for formal education, which educates learners for tomorrow, not yesterday, because the stuff of literate learning that underwrites formal education continues to draw, in the main, from yesterday's norms and standards.

7.3.2 *Language in Mobile Digital Context*

How exactly is language outgrowing print era borders? In an ongoing study of evolutionary changes in language form and use co-evolving with digital technologies,¹ our team is documenting a number of key linguistic evolutions permeating everyday communication, including but not limited to:

- interactive R/W functionality,
- multisemiotic resource encoding using haptic strokes or voice commands on a keyboard or touch screen,
- modular design as the principle in composition rather than linear word choice and positioning in edited writing,
- immediate global publication and sharing of ideas and documents enabling crowd-sourcing, and real-world audience feedback,
- textual products created in a permanently editable state (with some exceptions in commercially published digital books and journals) rather than print-fixed publication (gone are *errata* sheets inserted in printed books), and
- chatting with AI.

These and other evolutions in communication have co-developed with technological advancement, generating new genres (e.g., fanfic, unboxing), discourses (e.g., #photooftheday, comments sections in social media),

and cultural practices (e.g., massive multiplayer online role-playing games [MMORPG]; blogging; podcasting), each of which grows corresponding conventions (e.g., #, lol, 😊). These become norms within the generating genres, though given the speed of technological change, the norms, too, are in a state of change, for example, the microblogging site, Twitter, has gone from 140 characters to 280 characters per tweet. Interestingly, some conventions have jumped generic borders, such as hashtags. For instance, #MeToo as a topic heading has migrated across microblogging and social media forums to protest signs and newspaper headlines, and even into speech, pronounced: “hashtag me too.”

Given the plethora of novel language forms being generated in digital fora, appropriate theorizing of what constitutes *language* and *literacy* for learners in this day and age is urgently needed.

7.4 THEORIZING MULTIMODAL COMMUNICATION: TWO VIEWS

“Multimodality’ names the field in which semiotic work takes place, a domain for enquiry, a description of the space and of the resources that enter into meaning in some way or another,” states Kress (2011: 38). He further asserts that analysing multimodal discourse “needs to encompass all modes used in any text or text-like entity, with each described both in terms specific to its material and historical affordances and in terms shared by all modes” (2011: 38). This description signals the underlying complexity of multimodality as theorized in social semiotics and indicates significant overlap with Elleström’s intermediality theory. There is commonality in conceptual range, including the modalities of basic media (material, spatiotemporal, and semiotic modalities though less of sensorial), as well as qualified (historical, social, cultural) aspects of media. Epistemologically, however, the approaches vary considerably.

Social semiotics is a theoretical perspective derived from Michael Halliday’s (1978) systemic functional linguistics (SFL), which theorized language in social use. Halliday’s SFL departed from prior structuralist theories of language form, which explained language as abstract structure, detached from social use. The concept of social semiotics was further developed by Hodge and Kress (1988), predating the evolution of digital communication. Kress’ (2000, 2003, 2009, 2011) continued work on social semiotics, multimodality, and literacy provides a dominant

theoretical lens for understanding the innovative multimedia texts that have evolved over the digital era.

In a social semiotics paradigm of multimodality, *mode* is the basic concept, indexing the means of semiotic representation. *Media* refers to technologies of dissemination (Bezemer and Kress 2008; Jewitt 2004; Kress 2005). Multimodal texts are designed and “composed of different modes, resting on the agentive semiotic work of the maker of such texts” (Kress 2011: 36). Modes are typically exemplified rather than analytically explained, for example:

If, going one step further, we compare a contemporary textbook with ‘pages’ on the Web dealing with the ‘same’ issues, we see that modes of representation other than image and writing—moving image and speech for instance—have found their way into learning resources, with significant effect. (Bezemer and Kress 2008: 167)

The affordances of modality can become exceedingly complex:

It varies in line with the affordances of each mode: here in a contrast of speech and image—of lexis vs depiction; of possession vs proximity or distance, of centrality or marginality; as a verb-form vs spatial co-location; sequence (as temporal succession in speech or linearity in writing) vs simultaneity (of appearance and arrangement) of the entities. (Kress 2011: 45)

In Elleström’s (2020) conceptualization, multimodality is a feature that helps to define intermediality. His theorizing follows Mikko Lehtonen’s explanation that “multimodality always characterises one medium at a time. Intermediality, again, is about the relationships between multimodal media” (cited in Elleström 2020: 41). Elleström explains, “intermediality is about the relationship between media having a multitude of vital traits, or modes” (2020: 41). Four modalities, that is, types of modes, form indispensable cornerstones of all media: *material*, *spatiotemporal*, *sensorial*, and *semiotic*. Together they build in physicality, perception, and cognition.

Media in a social semiotics paradigm may index what Elleström categorizes as *technical media of display* (2020: 33–40). However, a discussion of technical media invites an ontological lens on what constitutes technology, which exceeds the purview of this article, calling into question the relationship of qualified (socio-historical aspects of media) and technical

media of display. My reluctance to do this, in brief, follows Lawson's (2008: 48) argument that technology is "irreducibly social." Bijker (2010), speaking from a social construction of technology theory, elaborates on how detached views of technology are impossible. In his words, "technology is socially (and politically) constructed; society (including politics) is technically built; technological culture consists of sociotechnical ensembles" (Bijker 2010: 72). Disconnecting technical media from their contextualizing socio-history is, thus, specious.

7.5 MODALITY, MODE, AND MEDIA IN DIGITAL COMMUNICATION

Digital text creation is a process of multimedia design, not simply alphabetic encoding, and it cannot be taught as if it were. This is not a new realization, as can be seen in Kress' (2000: 339) prediction two decades ago:

The semiotic modes of writing and of image are distinct in what they permit, that is, in their affordances. Image is founded on the logic of display in space; writing (and speech even more so) is founded on the logic of succession in time. Image is spatial and nonsequential; writing and speech are temporal and sequential. That is a profound difference, and its consequences for representation and communication are now beginning to emerge in this semiotic revolution.

Multimedia design, however, is far more complex now than indicated then. Worth commentary is that Kress is on record (Kress and van Leeuwen 1996) describing the evolution of alphabetic writing from images—the linkage being very much still evident in logographic systems, such as Mandarin Chinese. So how can the logic of writing be so polarized from the logic of image, either in process—which in both cases requires temporal logic: you have to write or create the product—or in product—which, depending on the kind of image and kind of writing, may be spatial in orientation but is likely to be more complexly intertwined with temporal and perhaps even tactile perception? In terms of writing, consider subtitles at an opera, electronic highway signs, moving advertising on rotating e-bulletin boards. These require timed spatial perception. In terms of image, consider sculpture, which requires spatial perception but also requires movement around the sculpture, and thus is temporal in

perception, too. Also sculpture may invite tactile perception as part of the sensory experience.

Text composition today engages and remixes graphic resources, sound files, and moving images from gifs to videos into alphabetic text in a key-stroke. Elemental semiotic resources, such as pictures and letters, are essentially mashable by virtue of their shared materiality: the pixel. Mobile smart devices embed a portable digital toolkit for designing, producing, and sharing multimedia texts. Making sophisticated multimedia texts on a smartphone is easy-as-pie.

Language education requires principled conceptualization of the semiotic elements being functionally captured and remixed in contemporary digital communication so they are included pedagogically. Given the immense changes occurring in digital language form and function, gaining a full understanding of how multimedia textual production varies from linear alphabetic writing is a vast project. In the following section, two radically novel digital twists to language today are examined through Elleström's intermediality lens: the inclusion of *emoji*, for example, 🙄, 🌸, 🍏, in digital writing and the incorporation of AI interlocutors, or conversational digital agents, in spoken communication in digital environments such as mobile phones.

7.5.1 *Emoji*

Language learners routinely expect vocabulary and grammar to constitute basic language learning. However, the question “What is a word?” must be asked in an environment where novel word-like formations that have no pre-digital era precedent, including hashtags (e.g., #BlackLivesMatter), and emoji (e.g., 🍏), have taken on a life of their own. This section examines emoji as a novel element in digital R/W vocabulary.

Pardes (2018) describes *emoji* as, “tiny, emotive characters—from 🙄 to 🍏—[which] represent the first language born of the digital world, designed to add emotional nuance to otherwise flat text.” Emoji developed in Japan following early digital chat play with emoticons using the ASCII keyboard, for example, :-). Japanese emoji were seen as a valuable contribution to digital platforms and adopted by tech giants in the early 2000s, who then petitioned for their inclusion in the Unicode Consortium (Pardes 2018). This stabilized emoji for universal keyboard use in terms of the binary code computers use.

Emoji are pictograms, as are many public signs, for example, , indexing “recyclable.” Pictograms differ slightly from logograms, which are words encoded in a script; for example, 你好 in traditional Chinese, pronounced *nǐ hǎo* (literally “you good”), encodes the greeting “hello,” which in English requires five letters. You can speak a logogram but not a pictogram. As such, emoji exist only in literate form; they do not directly encode a spoken form, though they can be translated, for example,  is a *smiley face*, though just one variant of smileys, also including , .

Emoji utilize an elemental iconic keyboard, which makes an emoji as easy to insert into written text as a letter (e.g., *a* or *b* or *c*). The emoji keyboard is analogous to the *qwerty* keyboard as *technical* interface (used with a screen as *technical medium of display*). The *qwerty* keyboard has been a primary technical interface between writer and text since its mechanically driven design in the nineteenth-century manual typewriter (Noyes 1983), used with paper. There are parallel keyboards for languages, for example, accented Roman (French, Swedish); non-alphabetic scripts (Chinese, in traditional or simplified logograms); and emoji, which can be imported into text-making at the touch of a button on a smartphone (see, e.g., Fig. 7.2).

Given that all basic media are characterized by four modalities—material, spatiotemporal, sensorial, and semiotic—and that a medium may demonstrate multimodal characteristics within a single modality, how might emoji be modally described? As emoji exist only in literate form, they are visually perceived, spatially interpretable symbols that are encoded materially in pixels, using a keyboard as technical interface and screen as technical medium of display. Consider the attempt to parse these complementary modalities in Table 7.1. It is critical to state up front that this untangling of the modalities of emoji is intended to illustrate the complexity and relative complementarity of media, not to deconstruct emoji into isolatable component parts that do not blend into each other. Parsing is intended to highlight more and less prominent modalities in interpretation.

Media and mediation are exceedingly complex in today’s communication landscape. A media product or medium exists in historical-social-cultural space as well as physical-sensorial-cognitive space. According to Elleström, “qualified media” are “media types, which depend on history, culture and communicative purposes” (2020: 57). A more delineated

Fig. 7.2 Parallel keyboards available on smartphone

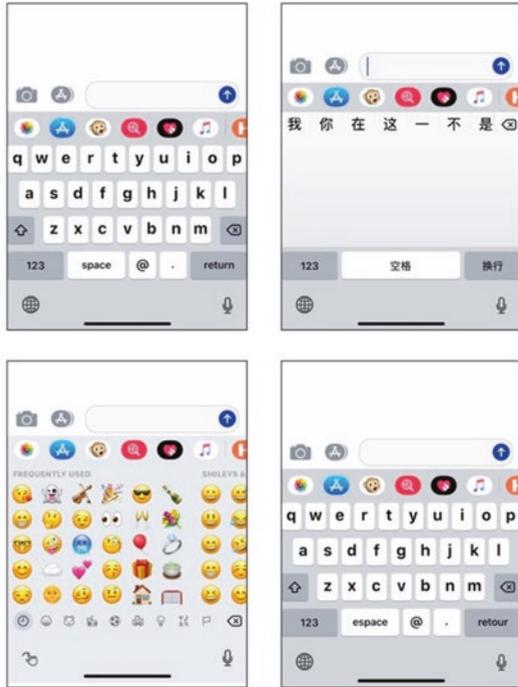


Table 7.1 Basic and technical media of emoji

<i>Medium product</i>	Emoji	
<i>Basic modalities</i>	Material	Pixel
	Spatiotemporal	Spatial
	Sensorial	Visual
	Semiotic	Textual emotional nuance
<i>Technical media</i>	Digital screen as display	Informal register: texting, IM chat
	Emoji keyboard	Mobile phone, computer, etc.
		Alternate parallel emoji keyboard

understanding of emoji in contemporary communication comes to light in the qualified aspects of the medium.

Emoji are media products; they are also a media type that is endemic to informal texting and digital chat environments. Despite their immense cross-cultural popularity and widespread use, they have not migrated to more formal writing environments. The qualified aspects of emoji help to

Table 7.2 Basic, qualified, and technical media of emoji

<i>Medium product/type</i>	Emoji	
<i>Basic modalities</i>	Material	Pixel
	Spatiotemporal	Spatial
	Sensorial	Visual
	Semiotic	Textual emotional nuance
<i>Qualified media</i>		Informal: texting, IM chat
	Social acceptability	Established quotidian
	Social availability	In-built in smart devices
	Educational approval	Not recognized in academic learning
	Historical evolution	Japanese texts to global big tech
<i>Technical media</i>	Cultural uptake	Unknown degree of universality
	Digital screen as display	Mobile phone, computer, etc.
	Emoji keyboard	Alternate parallel emoji keyboard

flesh out a broader understanding of emoji as semiotic resource, going some distance in explaining the reluctance of formal educational infrastructures to absorb this element of contemporary communication into language study (see Table 7.2).

Intermediality is, of course, not neatly packaged in a grid, as the prefix *inter-* suggests: each aspect of media seeps into and colours complementary aspects. Hence, the unreality of separating technical media of display from social and historical context despite the practicality of knowing the intended or most suitable device. Nonetheless, creating a constituent media analysis of emoji is a valuable exercise for understanding how and where this literate innovation works in digital composition. Let us now turn to a more nebulous digital phenomenon: the surreptitious permeation of AI into digital conversation.

7.5.2 *Conversational AI*

A keen interest in understanding communication in mobile digital context is the emergence of the conversational digital agent. The expansion of voice recognition software into the sophisticated conversational digital agent historically coincides with the release of smart mobile devices, which, because of their limited screen size, benefitted from a voice to provide assistance rather than relying on tiny written instructions (Pinola 2011). The integration of global positioning system (GPS) receptors in smart devices enabled digitally voiced navigation. AI voices are also incorporated incognito in language teaching apps.

How might a disembodied digitized computer voice be described in terms of the four modalities of basic media? To hear a voiced message from a disembodied computer-activated conversational agent, switch on the voiceover accessibility feature on a computer; ask a question to your in-phone conversational digital assistant, e.g., Siri, or plug in a destination on a GPS device and enable voiceover directions. This will result in a media product that could be basically parsed as in Table 7.3.

Digital voices are, of course, a vehicle for a linguistic message, which would require elegant semantic and structural delineation within the semiotic category. This analysis can rely on the extensive attention paid to how language conveys meaning to whom, when, where, why, how, and so on that is entailed in linguistic theories. However, the basic media product of the voice of the conversational digital agent (as well as less complex AI *chatbots*) is also a media type. The digital conversational agent has permeated not simply auditory media but also robotic shapes. In the spoof Amazon advertisement about the integration of the Amazon conversational agent into the Internet of things, as shown during the 2019 Super Bowl (American football), Alexa, the voice-activated digital agent, is capable of, among other things, interpreting dog bark commands ordering dogfood.² Though very silly, the advertisement showcases the voices of male and female humans (speaking English), a dog, and a digital agent with attendant media properties, semantic comprehensibility, and so forth.

Quite apart from the exceedingly complex technical embedding of AI in media products that are more multifaceted—a smart mobile device, a refrigerator, a thermostat, and so forth—the emergence of the

Table 7.3 Basic media modalities of conversational digital agent voice

<i>Medium product</i>	Digitized voice	
	<i>Basic modalities</i>	
	Material	Sound waves
	Spatiotemporal	Temporal transmission
	Sensorial	Auditory
		Haptic activation
	Semiotic	Spoken message
		Characteristics of voice
		Humanness
		Gender
		Age
		Authority
		Accent
		Regional location
		Social information

conversational digital agent in language learning is a qualified media question mark. In the messy and dispute-ridden world of education, formal and nonformal, the contextualizing aspects of media may be a total deal-breaker. Whereas novel digital genres and discourses have reshaped communication practices, normalizing activities such as social media posting (Facebook, Instagram), blogging, vlogging (YouTube), microblogging (Twitter), and texting (instant messaging), to mention just a few, the uptake of such discourses and their conventions in formal language and literacy teaching contexts has been spotty. Some schools encourage a bring-your-own-device (BYOD) approach to students' using their personal mobile phones for learning; others expressly forbid mobile devices in class. In some classroom contexts, designing new media texts is encouraged, whether using institutional or personal devices; in others, curricula default to the conventions of static print media, stuck in *speaking-listening-reading-writing* skills. An analysis of the digital conversational agent in terms of qualified media in addition to the complex morass of the technical display in our parsing exercise might look as in Table 7.4.

Table 7.4 Basic, qualified, and technical media of conversational digital agent voice

<i>Medium product</i>	Digitized voice	
	<i>Basic modalities</i>	
	Material	Sound waves
	Spatiotemporal	Temporal transmission
	Sensorial	Auditory
		Haptic activation
	Semiotic	Spoken message
		Characteristics of voice
		Humanness
		Gender
		Age
		Authority
		Accent
		Regional location
		Social information
<i>Qualified media</i>	Social acceptability	Limited social awareness
	Social availability	Unregulated digital marketplace
	Educational approval	Unregulated; unrecognized
	Historical evolution	AI development
<i>Technical display</i>	Cultural uptake	Dependent on users
	Mobile digital devices	Button or voice command
	GPS device	Button or voice command
	Appliances	

This simplified grid is intended to spotlight modalities of media in terms of their prominence in the media product, not to cleanly disambiguate modalities from each other, as previously disclaimed. Qualified aspects of digitized voices re: legitimation for an educational activity, for instance, may colour acceptability of modalities such as voice quality as well as semi-otic aspects of accent programmed into the digitized voice, and vice versa. The following discussion on materiality indicates the mammoth complexity of a disembodied digitized voice.

Language from Whose [sic] Perspective: The Complex Materiality of AI Communication

Given the infusion of conversational digital agents in daily communication practices, from following voiced GPS navigation in the car to asking Siri for help on the iPhone to including the robotic Amazon digital assistant, Alexa, in family conversations at home, we are justified in asking: Is language still human?

Interestingly, humans using voice-activated digital assistants often assign human genders to them (e.g., “Thank you, lady in the computer!”; “She’ll tell you when to turn”). These voices, though, are programs: circuits, not people, despite the fact that the original mediating material was produced by humans prior to being subjected to complex technological remediation. Peña and James (2016, 2018), writing about glitch art and pedagogy, problematize materiality in digital transmediation across sensory domains, making the case that computers can interpret and create sounds that are only partially interpretable to humans.

Human talk and text do not travel through computer circuits in humanly recognizable form. Computers function on a binary 0-1 code. The materiality of human–computer–human communication loops with digital conversational agents, such as Siri or Alexa, has complex technological layers, tapping a *material* modality that is a veritable iceberg reaching into depths that are only partially interpretable to human interlocutors.

So the answer to “Is language still human?” is convoluted: computer programs were originally programmed by humans—though whether this continues to be true in Web 3.0 environments is another question.

7.6 CONCLUSION: FROM ABCS TO INTERMEDIALITY

The speed of development in literate conventions, genres, texts, and discourses as they mutually develop with socio-technical advancement has not been matched in formal language and literacy instruction, which still tends to prioritize socio-historically and politically sedimented literate forms. One problem in moving conceptions of writing into multimedia composition is that rapid changes in previously understood constancies, such as spelling, are *user-driven* in the digital age.

At the beginning of the school year, children go off to school to learn their ABCs. At least that is what used to happen and indeed is still the idea parents have in their heads. It is not so different with adult (and school) learners of second languages. But times have changed. Everyday digital communication practices—social media posting (e.g., Facebook, LinkedIn), microblogging (Twitter), videologging (e.g., YouTube), and photography-centred posting (e.g., Instagram, Pinterest), for example—are co-encoded with emojis (e.g., 😊), and other graphic and sound media, such as photography, film, music, random sounds, gifs, and on and on. These forms are co-encoded with alphabetic symbols that can also be voiced—by software programs.

Conventions evolving in digital forums have revolutionized concepts as basic as the word: novel word-like forms, such as the hashtag, for example, #TimesUp; #foodporn that do not follow historical semantic or structural word formation patterns have crossed from digital environments into print media. Similarly, AI, which is commonly used for narrating GPS directions or giving help instructions on mobile device screens, is also being programmed into language learning apps.

Given this reality, language professionals and educational policy makers need to be aware of how media resources work to create texts and textual meaning from an arts-based as well as a language-based perspective. Elleström's (2020) intermediality paradigm offers a much-needed vector for analysing multimedia communication from an arts perspective that can be integrated with linguistic theories to understand truly digital language as it is evolving.

This chapter undertook a comparative look at *modality*, *mode*, and *media* through a social semiotic and an intermediality lens to clarify the complex task of understanding the contribution of different semiotic resources in multimedia textual products. The raw data on new conventions in digital language use are from an extensive literature review of how

mobile digital access has affected linguistic communication. This review informs current exploratory research to build appropriate mobile production pedagogies for language learners, using, rather than ignoring, today's communicative potential.

Two innovative communicative features that evolved with communications technologies advancements were selected for analysis using an intermediality lens: the use of emoji in textual products and spoken communication with conversational digital agents. Though it must be said that emoji are unlikely to make a breakthrough into academic writing any time soon, other visual-spatial lexical innovations in digital communication have already begun to cross from specific digital forums into print manifestations, such as #hashtag topics: forms that do not cohere with lexical borders or alphabetic principles. The AI voice anonymously joins us in everyday communication. Many questions ensue.

NOTES

1. Heather Lotherington and Kurt Thumlert, *Production Pedagogies for Language Learning in Mobile Digital Environments*. Grateful acknowledgment to the Social Sciences and Humanities Research Council of Canada for funding our Insight grant.
2. See https://youtu.be/8y-1h_C8ad8.

REFERENCES

- Bezemer, Jeff, and Gunther Kress. 2008. Writing in Multimodal Texts: A Social Semiotic Account of Designs for Learning. *Written Communication* 25: 166–195.
- . 2016. *Multimodality, Learning and Communication: A Social Semiotic Frame*. London: Routledge.
- Bijker, Wiebe E. 2010. How Is Technology Made?—That Is the Question! *Cambridge Journal of Economics* 34: 63–76.
- Cope, Bill, and Mary Kalantzis. 2004. Text-Made Text. *E-Learning and Digital Media* 1: 198–282.
- Eide, Øyvind, and Zoe Schubert. 2020. Seeing the Landscape Through Textual and Graphical Media Products. In *Beyond Media Borders: Intermedial Relations among Multimodal Media, Volume 2*, ed. Lars Elleström, 175–209. Basingstoke: Palgrave Macmillan.

- Elleström, Lars. 2010. The Modalities of Media: A Model for Understanding Intermedial Relations. In *Media Borders, Multimodality and Intermediality*, ed. Lars Elleström, 11–48. Basingstoke: Palgrave Macmillan.
- . 2020. The Modalities of Media II: An Expanded Model for Understanding Intermedial Relations. In *Beyond Media Borders: Intermedial Relations among Multimodal Media, Volume 1*, ed. Lars Elleström, 3–91. Basingstoke: Palgrave Macmillan.
- Gershon, Ilana, and Paul Manning. 2014. Language and Media. In *The Cambridge Handbook of Linguistic Anthropology*, ed. N.J. Enfield, Paul Kockelman, and Jack Sidnell, 559–576. Cambridge: Cambridge University Press.
- Halliday, Michael A.K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Have, Iben, and Birgitte Stougaard Pedersen. 2020. Reading Audiobooks. In *Beyond Media Borders: Intermedial Relations among Multimodal Media, Volume 1*, ed. Lars Elleström, 197–216. Basingstoke: Palgrave Macmillan.
- Henson, John. 1744. *A New Latin Grammar, Adapted to the Capacities of Young Scholars; Comprising Every Thing in the Art Necessary for Grammar-Schools. With Short, Easy, and Proper Examples to All, and Every Part of, the Rules in Syntax. By the Rev. Mr. Henson, Master of the Free-School in Nottingham*. Nottingham: J. Rivington.
- Hodge, Robert, and Gunther Kress. 1988. *Social Semiotics*. Cambridge: Polity Press.
- Jewitt, Carey. 2004. Multimodality and New Communication Technologies. In *Discourse and Technology: Multimodal Discourse Analysis*, ed. Philip LeVine and Ron Scollon, 184–195. Washington, DC: Georgetown University Press.
- . 2008. Multimodality and Literacy in School Classrooms. *Review of Research in Education* 32: 241–267.
- Kress, Gunther. 2000. Multimodality: Challenges to Thinking About Language. *TESOL Quarterly* 34: 337–340.
- . 2003. *Literacy in the New Media Age*. London: Routledge.
- . 2005. Gains and Losses: New Forms of Texts, Knowledge, and Learning. *Computers and Composition* 22: 5–22.
- . 2009. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- . 2011. Multimodal Discourse Analysis. In *The Routledge Handbook of Discourse Analysis*, ed. James Paul Gee and Michael Handford, 61–76. London: Routledge.
- Kress, Gunther, and Theo van Leeuwen. 1996. *Reading Images: The Grammar of Visual Design*. London: Routledge.
- Lawson, Clive. 2008. An Ontology of Technology: Artefacts, Relations and Functions. *Techné: Research in Philosophy and Technology* 12 (1): 48–64.

- Lotherington, Heather. 2018. Mobile Language Learning: The Medium Is Not the Message. *L2 Journal* 10: 198–214.
- Noyes, Jan. 1983. The QWERTY Keyboard: A Review. *International Journal of Man-Machine Studies* 18 (3): 265–281.
- Ong, Walter J. 1982. *Literacy and Orality: The Technologizing of the Word*. New York: Methuen & Co.
- Pardes, Arielle. 2018. The Wired Guide to Emoji. *Wired*. <https://www.wired.com/story/guide-emoji/>. Accessed 3 February 2020.
- Peña, Ernesto, and Kedrick James. 2016. A Glitch Pedagogy: Exquisite Error and the Appeal of the Accidental. *Journal of the Canadian Association for Curriculum Studies* 14 (1): 108–127.
- Peña, Ernesto, Kedrick James, and Digital Literacy Centre. 2018. Raw Harmonies: Transmediation Through Raw Data. *Leonardo* 51: 1–12.
- Pinola, Melanie. 2011. History of Voice Recognition: From Audrey to Siri. *itbusiness.ca*. <https://www.itbusiness.ca/news/history-of-voice-recognition-from-audrey-to-siri/15008>. Accessed 3 February 2020.
- Satrapi, Marjane. 2003. *Persepolis*. New York: Pantheon Books.
- Thumlert, Kurt, Suzanne de Castell, and Jennifer Jenson. 2015. Short Cuts and Extended Techniques: Rethinking Relations Between Technology and Educational Theory. *Educational Philosophy and Theory* 47 (8): 786–803.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copy-right holder.

