



HAL
open science

Phraseology and Lexicogrammatical Patterns in Two Emergent Paragame Genres Videogame Tutorials and Walkthroughs

Christopher Gledhill

► **To cite this version:**

Christopher Gledhill. Phraseology and Lexicogrammatical Patterns in Two Emergent Paragame Genres Videogame Tutorials and Walkthroughs. Astrid Ensslin; Isabel Balteiro. Approaches to Videogame Discourse: Lexis, Interaction, Textuality. Bloomsbury Academic: Oxford, pp58-86., Bloomsbury Academic, 2019. hal-03274243

HAL Id: hal-03274243

<https://u-paris.hal.science/hal-03274243>

Submitted on 6 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



3

Phraseology and Lexico-grammatical Patterns in Two Emergent Paragame Genres

Videogame Tutorials and Walkthroughs

Christopher Gledhill

1 Introduction

When looking at the discourse of videogames, it is possible to focus on highly visible features of language such as novel terminology (*combo*, *to plink*, *whiffed*, etc., see Álvarez-Bolado and Álvarez de Mon, this volume), original combinations of existing forms (*cutscene*, *sidequest*, etc., see Balteiro, this volume), or markers of oral interaction and emotion (*ouch!* see Ensslin and Finnegan, this volume). However, in this chapter, I concentrate on “phraseology,” which I define here very informally as “the preferred way of expressing meaning in a particular discourse.” Whereas many linguists consider phraseology in terms of idiomatic expressions, proverbs, fixed phrases, and the like, in this chapter I adopt the “contextualist” approach, first proposed by J.R. Firth and then developed during the early days of corpus linguistics by J. Sinclair and others (Firth 1957; Sinclair 1991; Stubbs 1993; Hunston and Francis 2000; Hoey 2005; Sinclair and Mauranen

2006). Following this approach, I suggest that videogame tutorials and videogame walkthroughs are not only well-defined varieties of language, but they also have their own particular phraseology, that is to say, their own particular configuration of lexico-grammatical patterns (as defined by Hunston and Francis 2000; Gledhill 2000a, b). In the following sections, I set out a methodology for establishing phraseological patterns, which starts off by examining “key” (statistically significant) grammatical items, and then proceeds to analyze the most typical ways that these items are used in characteristic phrases on the basis of corpus evidence (an approach first set out in Gledhill 1995 and developed in other studies, e.g., Groom 2007, 2010).

The advent of corpus linguistics has changed the ways that analysts think and talk about routine patterns of expression. Corpus linguists have developed a variety of terms to talk about the units of phraseology, such as *collocational frameworks*, *lexical patterns*, *collostructions*, *discourse routines*, and so on (Renouf and Sinclair 1991; Hunston and Francis 2000; Stefanowitsch and Gries 2003; Tran et al. 2016). While phraseology is often discussed in terms of the general language, analysts have more recently explored how routine patterns are used in different registers, in particular in specialized areas such as scientific writing (Gledhill 2000a, b), academic discourse (Groom 2007, 2010), technical instructions (Couterut 2016), business communication (Née et al. 2017), and so on. However, while there is now a sizable literature on the phraseology of specific genres in English and other languages, only a few studies (Ensslin 2011) have examined routine phrases in videogame discourse.

In this contribution, I claim that the basic unit of phraseology is the “lexico-grammatical pattern” (LG pattern, for short). A typical LG pattern can be defined as a recurrent sequence of lexical items which extends beyond the syntactic group (i.e., it can be longer than a nominal group, verbal group, etc.). In addition, each LG pattern has an identifiable semantic or rhetorical function, which is specific to the particular discourse in which it is observed. In the following sections, I set out a replicable methodology (as set out below) for identifying the most typical LG patterns in videogame tutorials (henceforth VGT) and videogame walkthroughs (VGW). While I suggest that this methodology is systematic, I do not claim that it is very sophisticated. Indeed, corpus-based methods such as multifactorial analysis (Biber et al. 2004, 2010), the analysis of n-grams and tag-grams (Née et al. 2017), and more recent approaches (such as textometrics) are now widely used by corpus linguists in order to identify regularities of expression. However, I find that such advanced methods pose problems for those analysts who want to look at the behavior of a specific type of discourse without prior training in statistics or programming.

For this reason, I have previously suggested (Gledhill 2000a, b, 2015) that the analysis of grammatical keywords provides an efficient way of

conducting a preliminary analysis of the main LG patterns in a particular corpus. For example, in a corpus of VGTs, the pronoun *it* is found to be a statistically significant key word when this specific corpus is compared with a corpus of general English. This observation is not significant on its own, but in the VGT corpus, it can be seen that the word *it* is associated with longer patterns of expression such as *<it is good for VV+ing NN>*¹ in which the embedded verb (here marked VV) refers regularly to a specific type of attack (*dodging fireballs, punishing whiffed air attempts, starting combos*, etc.). In other words, we have moved from the observation of a highly frequent grammatical pronoun to the observation of game-related terminology (and the way it is evaluated, in the phrase *<it is good for VV+ing>*). It can be shown that micropatterns such as these may vary according to context but often express an abstract meaning as a whole which goes beyond the local meaning of its constituent units and their frame of reference (Adam 2011 calls these “macro-propositions” [131]; in Gledhill 2000a, b, I call these “discourse functions”). Thus, in the context of VGTs, the pattern *<it is good for VV+ing>* has a rhetorical function which can be paraphrased as “summarizing the main advantage of a previously mentioned fighting ability.” It is in this respect that lexico-grammatical patterns resemble more traditional phraseological phenomena, such as idioms, proverbs, routine formulae, and similar multiple-word units.

I would argue that by identifying lexico-grammatical patterns and associating them with rhetorical functions in this way, it is possible to arrive at a description of the most characteristic features of a particular type of text. In addition, I suggest that if it is possible to show that a particular type of text or discourse has a predictable and productive repertoire of lexico-grammatical patterns, then it is likely that this discourse has evolved into a mature LSP, that is to say, a variety of language that serves the purposes of a self-defining group or community, which participates in the adaptation of its own conventionalized lexico-grammatical patterns of language (i.e., phraseology), as well as its own codified channels of communication (i.e., genres, as described by Swales 1990; Gee 2005; and others). In the concluding section of this chapter, I return to this notion in the light of my analysis of the VGT and VGW corpora.

2 Data selection: General characteristics of the VGT and VGW corpora

Before looking in detail at corpus data, it is worth pointing out some of the general features of VGTs and VGWs. The two corpora analyzed in this chapter were collected by my students as part of a course titled “Technical Discourse Analysis” (TDA).² This course is part of a two-year master’s

course at the Université Paris Diderot, France (*Master ILTS—Industries de la langue et traduction spécialisée*). The aim of the master's is to train technical communicators and translators, with particular emphasis on the acquisition of terminology and phraseology in different technical and specialized domains. The specific aim of the TDA course is to raise awareness about the different types of technical genres in English, as well as to promote the systematic use of corpus analysis as a transferable research skill.

The TDA course requires students to build and then analyze a corpus which is representative of a particular genre of written English. The students can explore any genre, as long as it belongs to a technical register.³ The exercise therefore excludes literature, fiction, and journalism but includes expert-to-expert genres (such as dissenting opinions, oil refinery operating manuals, scientific research articles, etc.) as well as expert-to-nonexpert genres (organ donation brochures, political manifestos, social network privacy policies, etc.). Over the years, my students have worked on all of these genres (the characteristics of some of these are set out here for the purposes of comparison). However, more recently, some students have also asked to study genres which do not easily correspond to the canonical notion of a technical text, and they have shown increasing interest in texts related to gaming and videogames. In 2016, two groups of students chose to study VGTs and VGWs. Although these texts present a number of contextual differences in relation to more traditional technical genres, my students argued convincingly that VGTs and VGWs constitute legitimate topics for the TDA project.

The VGT corpus (111,695 tokens) is described by my students as “character guides, general tutorials and glossaries from various sources, such as videogame websites (*IGN*, *Gamefaqs*, *Supercheats*), websites [specializing] in competitive fighting games, and online versions of paperback guides.” The two students working on this project decided to concentrate on the tutorials available for one game, *Streetfighter 4*,⁴ published by two well-known websites on fighting games: *Eventhubs* and *ShoRyuKen*. Each of these sites was mined to obtain 97 texts of around 1000 words each. Each text describes in detail the fighting abilities of the characters encountered in *Streetfighter 4*. Although these texts are essentially instructional, they can also be seen as recreational: the authors are assumed to be experts in the game, and they take pleasure in exploring the different abilities and tactics employed by the player characters and their different adversaries, as well as passing on their experience and overt evaluation about these specific fighting styles.

The VGW corpus (568,998 tokens) is made up of eight “Japanese” RPG walkthroughs. These texts are available from various sources (the one used by my students was *Gamefaqs*). As the wordlist data suggest, these texts are extremely long: their average length is comparable to a short novel or PhD thesis (over 70,000 words in this corpus). Walkthroughs originated

as strategy guides in videogame magazines (they are related to, but not the same as video “longplay” and “playthrough”), although the term *walkthrough* itself is related to the development of computer software (*software walkthrough*, *software technical review*). While tutorials focus on combat techniques, RPG walkthroughs present a comprehensive description of an entire game world, thus allowing players to explore every location and to succeed (or make informed choices) in every encounter in the game. Yet while the settings and the types of activities are very different, walkthroughs seem to share a number of rhetorical aims that are similar to fighting game tutorials: the author talks the reader through a world that is mutually recognized and enjoys recounting the hazards and solutions to specific tasks that have to be overcome before further exploration.

As mentioned above, VGTs and VGWs are different but related text types. Both are examples of what Ensslin (2011: 8) calls “paratextual genres,” since they represent fan literature about the game rather than discourse which emerges within the game. Yet they also appear to belong simultaneously to two subcategories (“language about games by gamers,” and “language used in instruction manuals,” Ensslin 2011: 6). I would suggest that many of the linguistic differences between these texts and the general language corpus (in this case, my reference is the British National Corpus [BNC]) can be explained in terms of *technicity* (the extent to which these texts engage with the specificities of the fictional game world, or the mechanics of the game) and *interactivity* (the extent to which these texts provide a space for gamers to interact). The following two features (labeled F1 and F2 for “Feature 1” and “Feature 2”) are evidence of technicity:

F1) *Impersonal expression*. Both VGTs and VGWs use a variety of ergative, passive, and other impersonal structures. Such structures focus on inanimate objects (by topicalizing or “thematicizing” entities in sentence-initial position), thus allowing the omission of animate participants. As has often been observed, these are typical of scientific or technical registers:

- (1) **Crafts work** in the same fashion as arts, some are **target-fixed**...
[*Trails in the Sky*, walkthrough]⁵
- (2) To use talent arts, the talent gauge must be full—**gauge fills up**
every time **an auto-attack hit connects**. [*Xenoblade*, walkthrough]⁶

F2) *Terminological networks*. VGTs and VGWs use a wide range of technical uses of general language items (*juggle*, *meaty*, *poke*), as well as specialized collocations (*lag tactics*, *negative edge*, *tick throw*) and abbreviations (*c.mk* “crouching medium kick,” *SJC* “super jump cancel”) that are often mentioned without comment or definition (it is assumed that users have access to online glossaries such as “Terminology and glossary guide for fighting games”⁷). In addition, many terms involve novel forms of amalgamation such as nominal-verbal conversion: *cutscene*, *rushdown*,

sidequesting, *super cancel*, and so on (in the VGT corpus, many of these examples involve particles, as discussed in Section 4.1 below).

In the following discussion, I suggest the addition of two further features (F3 and F4) on the basis of corpus analysis:

F3) *Multiple embedding*. As shown in sections 4.1 and 4.2, below, some particles and prepositions are highly statistically salient in both the VGT and VGW corpora. Observation in the corpus suggests that these items are involved in technical nouns and verbs which are made up of embedded sequences of particles and prepositions (often associated with particle verbs/prepositional verbs), as in:

- (3) Spin Drive Smasher—Fairly easy to combo [short for “combination”] **into off of** a HK Spiral Arrow or any version of the Quick Spin Knuckle. [*StreetFighter 4*: Cammy, tutorial]⁸
- (4) During the middle of the game, Junpei won’t be as available, so you can miss **out of** maxing **out** this SL. [*Persona 3*, walkthrough]⁹

A further example of embedding can be seen in premodification and postmodification of nominal groups, as in the following examples (here the complex nominal is enclosed in single brackets [...]):

- (5) Starting your punish combos with [**the first hit of close Heavy Kick canceled into Heavy Punch Whip of Love**] can be a little easier to land than attempting to start the same combo with Heavy Punch... [*Xenoblade*, tutorial]
- (6) The Heavy Kick version of Lynx Tail is active almost twice as long, creating a bit of an unsafe mix-up for foes expecting [**the less active Medium Kick Lynx Tail**]. [*StreetFighter 4*: Elena, tutorial]
- (7) It’s a 3-frame start up and is probably [**the furthest teaching 3 frame normal**] in the cast. [*StreetFighter 4*: Cody, tutorial]

F4) *Complex subordination*. As discussed in Section 4.3, both corpora (but especially VGWs) use a variety of complex conditional clauses, as well as other forms of subordination, a characteristic which they share with other technical genres such as procedural manuals (Couterut 2016). This can be seen in the following examples (following the conventions of Systemic Functional Grammar, embedded clauses are signaled by square brackets [...]) and bound/subordinate clauses by slash symbols: // A = main clause // B = bound clause):

- (8) (A) [[Starting your punish combos with the first hit of close Heavy Kick [[canceled into Heavy Punch Whip of Love]]]] can be a little easier [[to land]] // (B) than attempting to start the same combo with Heavy Punch, // (C) though this easier combo does lose a bit of damage./// [*Xenoblade*, walkthrough]

- (9) (A) The Heavy Kick version of Lynx Tail is active almost twice as long, // (B) creating a bit of an unsafe mix-up for foes [[expecting the less active Medium Kick Lynx Tail.]] /// [*StreetFighter 4*: Elena, tutorial]

Thus far, we have seen that features F(1–4) are typical of formal, technical registers. We now turn to a further set of features (labeled F5–8 here), which show that both VGTs and VGWs also make use of resources that are thought of as typically interactive and oral (many of these characteristics have also examined elsewhere, for example in Balteiro, this volume):

F5) *Marked appraisal and evaluation*. VGTs and VGWs both involve numerous asides, comments, judgments, and other markers of authorial stance. These functions can be realized by various structures ranging from vague quantifiers to idiomatic expressions:

- (10) This boss is a **bit tricky** to hit, for he is located at the very edge of the battlefield [*Trails in the Sky*, walkthrough]
 (11) Battle Basics 1. **Piece of cake**, just gang up on the ghosts. [*Trails in the Sky*, walkthrough]

F6) *Lexical reduction*. Although VGTs and VGWs involve a number of technical terms, they also make use of highly informal terms, including vague nominals (*pantonyms*):

- (12) Don't buy any weapons, though, you can get better **stuff** soon. [*Tales of Symphonia*, walkthrough]¹⁰
 (13) Hit and run won't work, this **thing** has too great a range. [*Tales of Symphonia*, walkthrough]

F7) *Lexical expansion*. Since both VGTs and VGWs often deal with combat, they expand the lexical repertoire for this area, thus introducing a rich set of quasi-synonyms. While some items are euphemisms or attenuating expressions (*clean up, deal with, get rid of, finish off, make quick work out of, pick off*, etc.), others refer to death and destruction more directly, often using slang or taboo language to express these meanings with more force (*destroy, hack away, kill, pound the crap out of, do serious damage, happy slaughtering! take out, whack, wipe out*, etc.).

F8) *Interaction markers*. These items also contribute to the high degree of engagement (authorial stance, subjectivity, oral style) in VGTs and VGWs. This category includes oral discourse markers and typographic and other features below the level of the word (speech marks, exclamation marks, contractions, etc.):

- (14) Fight, then you'll be at the exit. **Ouch**. [*Radiant Historia*, walkthrough]¹¹

- (15) A lone knight. **Easy, right? Wrong**, this knight will wipe the floor with you if you're not careful. [*Disgaea*, walkthrough]

Again, on the basis of corpus analysis (below), I suggest two further features of interactivity:

F9) *Directed imperatives*. As discussed in sections 4.1, 4.2, and 4.4, a particularly characteristic feature of the VGW corpus is the widespread use of imperative instructions (including widespread ellipsis) to direct to the player/reader through their adventure. Here are some typical examples of this:

- (16) Ask Murray about “Sound,” **then** [ELLIPSIS] about “Bell toll.” [*Trails in the Sky*, walkthrough]
- (17) Also **note that you can use** a LK Scarlet Terror (Vf5 downleft.gif charge Vf5 right.gif + LK) kick instead of the HK **and then juggle** with an EX Scarlet Terror afterwards. [*StreetFighter 4: Vega*, tutorial]
- (18) **Then** [ELLIPSIS] back to the first room and [ELLIPSIS] west. [*Devil Survivor*, walkthrough]¹²

F10) *Deontic modality*. As discussed briefly in Section 4.4, VGTs and VGWs use a rich variety of modal forms (which are more typically encountered in oral registers) to express advice, directions, and evaluation. Here are just two examples from a wide variety of potential forms in the corpus:

- (19) If you kill one enemy with a crit/weakness and get “1 More,” you **gotta** kill another enemy and get ANOTHER “1 More”... [*Persona 3*, walkthrough]
- (20) He'll start by using Vampire's Mist, which is a Mystic-type skill that targets all groups in the field and heals him up. WTF [NB **‘what the fuck’** are we supposed to do?! Mmm... [*Devil Survivor*, walkthrough]

It should be clear that not all of the characteristics listed here can be identified using the methods I set out below. The corpus-based analysis of LG patterns I set out below only reveals indirect evidence for evaluation, terminology, lexical expansion (or reduction), and markers of interaction. On the other hand, other features emerge from the corpus analysis: there is more emphasis on clause structure (complex subordination and embedding) as well as the construction of verbal groups (transitivity, direct imperatives, deontic modality). In the following sections, I examine these features in more detail and gradually make the case for identifying extended patterns of expression (LG patterns) as an important first step in the analysis of these texts.

3 Data collection: The identification of grammatical keywords

In this section, I describe the statistical methods used to identify the first ten¹³ salient grammatical keywords in the VGT and VGW corpora. As a first step in the characterization of a particular corpus, it is useful to identify the key lexical and grammatical items of that corpus using a tool such as *Keywords* (Anthony 2014). The AntConc Keywords program compares the lexical frequency lists from two comparable corpora; when a word has a significantly higher than average frequency in one corpus compared to the other corpus, that word is placed toward the top of the Keywords list. The position of each keyword depends on its “keyness” score. The keyness score is based on a comparison between the probability of encountering a particular word in the corpus under study and its probability of occurrence in the reference corpus. Thus, for example, the item *opponent* has a frequency of 878 per 111,695 words in the VGT corpus (or a relative probability of occurrence of 7.8 per 1000 words). In contrast, the same item has a frequency of 1,428 per 100,000,000 in the BNC (and thus a probability of occurrence of 0.01428 per 1000 words in that corpus). In this instance, the difference in probability is so great that the item *opponent* achieves a very high keyness score, and as can be seen in Appendix 1, AntConc consequently places *opponent* at rank 5 in the table of Keywords for the VGT corpus.

To give an idea of the kind of data that the Keywords tool reveals, Appendices 1 and 2 show the first ten items which emerge as the highest scoring Keywords for the VGT and VGW corpora.¹⁴ In a project on technical terminology, many of the lexical items on these lists (such as *combo*, *enemy*, *opponent*, *quest*) could be considered as candidate terms for a glossary in this domain. However, in this study, I am interested in the key use of particular grammatical items. I contend that if function words such as *you* or *up* obtain a relatively high score as Keywords in a particular corpus (even if they are not at the top of the Keywords list), then these items nevertheless have greater significance in that corpus, because—when one compares different registers of English—grammatical items are not usually expected to have as high a degree of variability as their lexical counterparts. In other words, most observers would agree that lexical items (such as *opponent* or *combo*) are expected to vary quite markedly in their frequency of occurrence. By contrast, it is less obvious that the relative frequency of grammatical words such as *you* and *up* can also vary, according to different registers (e.g., *you* will occur much more in spoken conversations than in scientific research articles, a fact that is behind the multifactorial analysis of linguists such as Biber et al. 2004, 2006). However, it is also the case that grammatical items are so ubiquitous and frequent in the English language as a whole that their degree of relative frequency does not vary as much as

lexical items. Thus, even if we observe a relatively small movement in the overall frequency of an item such as *you* (say an increase by 5%), it is likely to be more significant than an equivalent movement in the frequency of a typically less frequent lexical item such as *opponent*.

In Appendices 3 and 4, I have set out the first ten grammatical keywords found in the VGT and VGW corpora. These items are identified as “grammatical” because they belong to the closed lexical classes of

- (1) Adverbials and particle-like items such as *up* (AV),
- (2) Coordinating and subordinating conjunctions such as *if* (CJ),
- (3) Pronouns and deictic items such as *here* (PN),
- (4) Prepositions and grammaticalized items such as *right* (PR),
- (5) Auxiliaries, modals, and grammaticalized verbs items such as *get* or *let* (VX).

Of course, it is not useful to analyze grammatical items in isolation. It is, however, important to have some idea of the extent to which different types of grammatical items are salient in different types of texts. I have therefore set out in Table 3.1 the distribution of the first ten key grammatical items in the VGT and VGW corpora in comparison with the first ten key grammatical items in three other major registers: legal, informative, and specialized (these being the main registers that are studied in the TDA course, as mentioned in Section 2).

Explanatory Note 1. Some items are in parentheses because they belong to more than one part of speech.

Explanatory Note 2. All the corpora mentioned here were compiled by students on the TDA course, with the exception of scientific research articles, which are reported in Gledhill (2000, 2015).

Table 3.1 gives a general idea of some the broad similarities and differences that can be observed in a sample of major technical registers in English. For example, it can be seen that coordinating conjunctions (*and*, *or*) are salient in legal, informative, and specialized registers, while subordinators (*that*, *until*, *when*) are salient in legal and informative texts. In the data analysis below, I explore some of the reasons why a different set of conjunctions is preferred to these in the VGT and VGW corpora (namely *if* in the VGT corpus, and *if/after/once* in the VGW corpus).

A similar observation can be made with regard to prepositions (PR). Table 3.1 shows that legal, informative, and specialized registers share a preference for items which typically introduce circumstantials (*for*, *in*, *with*), or the item *of*, which is used in complex nominals. In contrast, VGTs and VGWs have a marked preference for AVs of direction (*down*, *up*), AVs

TABLE 3.1 Key grammatical items in a sample of four technical registers

Register	Genre	Texts	Words	key AV	key CJ	key PN	key PR	key VX
Legal	Social network privacy policies	30	105,321		or (that)	(any), (that), (to), you, we	For, (to), with	May, is
	U.S. dissenting opinions	489	1,881,336	not	or; (that)	it, (that)	At, by, for, in, of	is
Informative	Organ donation leaflets	58	112,526		and, or (until) when	(this), you,	For, of, (to),	will
	U.K. political manifestos	53	313,180		and	you, we	for, of, on, (to)	Are, have, will
Specialized	Oil refinery operating manuals	39	162,366		and		by, from, in, of, (to), with	are, be, is
	Scientific research articles	150	515,073	not	and (that)	(that)	For, in, of	Did, have, is, was
Instructional	Videogame tutorials (VGTs)	97	111,695	(down), forward, (into), (off) right, (up)	it, you	(down), (Intro), (off), (up)	Can, lets	
	Videogame walkthroughs (VGW)	8	568,998	(down) now then, (Up)	if once	here, you	(down), (up)	Can, get

of time (*now, then*) and PRs and other particles expressing a direction (*forward, into, off*). None of these items is significant on its own; rather, the fact that these items vary from one register to another points to significant differences in the phraseological patterns that are typical of these genres.

Finally, looking at Table 3.1, it is also interesting to note that there are certain affinities between genres or registers. For example, privacy policies, donation leaflets, and political manifestos all share a preference for the PN *you*, an item which is also a key word in VGTs and VGWs. One explanation for this is that all these genres are “directive”; they aim to make their readership respond to (or be responsible for) the content of the text. This is not a rhetorical function found in dissenting opinions or research articles (although it is perhaps surprising to find that *you* is not a key item in operating manuals). Having said this, as discussed below, I suggest that it is longer stretches of expression such as <*you can get VV/AJ*>¹⁵ which account for the particular distribution of items such as *you* in VGTs and VGWs.

4 Data description

In this section, I explore the first ten grammatical items that rank as salient keywords in the VGT and VGW corpora, with a particular focus on the phraseological patterns that are associated with these items. Rather than looking at each corpus in turn, in the following sections I divide the analysis into five “semantic zones,” each corresponding to a different set of grammatical items which are salient in VGTs and VGWs. These categories are set out in Table 3.2.

Although Table 3.2 sets out five categories of analysis, there is only space in Section 4 to discuss categories 1–4. Category 5, “causation and transitivity,” is partially discussed in the other sections (the attributive use of *get*, which is its most frequent use, is discussed in Section 4.3 in relation to the pattern <*if your opponent gets AJ*>, while the modal *can* is discussed in Section 4.4 in relation to the pronoun *you*).

Finally, the following conventions are observed in the data analysis:

- (1) A simplified version of the BNC tagset is used to signal abbreviated parts of speech: AT (article, determiner), AJ (adjective), AV (adverb), CJ (conjunction), NN (Noun), PN (pronoun), PR (preposition), VX (verb auxiliary or modal), VV (verb).
- (2) Authentic corpus examples are presented in a different font (or in italics if quoted in the text). Each bullet point represents the start of a new example.
- (3) Lexico-grammatical patterns are presented in triangular brackets, with generic items labeled by part of speech, as in <CJ (*and, but, so*) it's AJ (*good, great*) for VV (*catch, dodge, knock*)+ing>.

TABLE 3.2 *Semantic zones and grammatical keywords in the VGT and VGW corpora*

Semantic zone	Grammatical keyword
1. AV/PR direction and spatial extent	down, forward, into, off, right, up
2. AV/PR deictics and transitional location	here, now, then
3. CJ conditional advice and choice	if, once, after
4. PN reformulation and evaluation	it, this, you
5. VX causation and transitivity	can, does, get/gets, lets

4.1 AV/PR directions and spatial extent

In grammatical terms (Halliday and Matthiessen 2014), the semantics of adverbial and prepositional phrases can be analyzed in terms of either *location* (expressing a static point in space or time) or *extent* (expressing dynamic movement across space or time). Generally speaking, the items found in the VGT and VGW corpora tend to have a usage that is closer to extent. This notion is expressed in terms of relative space (*forward, into*), cardinal directions (*north*), and somatic relations (i.e., body-oriented items such as *right*).

In the general language, each of these items has many different potential contexts of use. But in technical corpora such as VGTs and VGWs, the relative frequency of each of these items can often be explained in terms of one or two very regular, recurrent patterns of use. For example, a number of these words, notably *forward, left, right, up, and down*, refer specifically to inputs on a joystick or remote control, and this usage accounts for the majority of their occurrences. Other items, especially *off, up, and down*, are predominantly used with particle verbs, such as *hold up* and *knock down*. Another very frequent use of *up* in both the VGTs and VGWs is as a post-modifying particle in a variety of converted (deverbal) nouns such as *cross-up, follow-up, jump-in, mix-up, start-up, wake-up* (these being the most frequently encountered occurrences). This usage is particularly prevalent in the VGT corpus. As can be seen in the following sample, the particle *up* provides a productive way of creating neologisms in this domain, with each verb + participle compound premodifying another noun which expresses a specific type of attack or maneuver (or more generally a fight, as in *mix-up*):

- (1) Best use is the EX version which will knock the opponent into the air for a **follow-up** juggle.

- (2) She's got solid pokes, a decent **mix-up** game, and fairly simple combos into her super and ultras.
- (3) Guacamole Leg Throw—In EX form, this is Fuerte's best anti-air attack. Use it especially as a **wake-up** counter to meaty **jump-ins**.

These examples are significant, but they have more to do with the formation of new terminology than the phraseology of this particular register. Perhaps the clearest example of an extended LG pattern which I can identify in this category involves *into*. In general English, *into* is often used with transitive verbs to form a causative construction of the type <VV NN *into* NN>. In contrast, in the VGT corpus, *into* typically occurs in combination with an intransitive verb, as in the following structure: <VV (*buffer, cancel, combo, land, tick*) *into* AT NN (*combo, move, [specific attack]*>. Semantically, this pattern refers to a transitional maneuver from one type of attack to another:

- (4) Pressure with low attacks and **cancel into the Hazanshu** to both maintain pressure and go for the mix-up.
- (5) Gouken can also **combo into his Shin Shoryuken** off of a backward throw.
- (6) As a meaty cross-up, if it gets blocked you can **tick into a grab** attempt.

Although the notion of extent is typically expressed by adverbial and prepositional phrases, many other constructions convey direction and movement (including verbal ellipsis—signaled below as [Ø], cohesive markers, and temporal AVs such as *again, then...*). This is especially the case in the VGW corpus, which employs an impressive repertoire of constructions expressing directionality. The following example shows how directed movement can be expressed by the same extended pattern within the same stretch of text (I suggest that the pattern occurs several times in the following extract, and has the form: <CJ (*and, but*)/AV (*again, then*) VV (*go, [Ø]*) AV (*back, east, left, north, right*)/PR (*down, to*) NN>):

- (7) <Go north> for a Panacea Plus. <Then [Ø] back to the first room> <and [Ø] west>. On the north side is a hidden Freikugel Mercy. <Again [Ø] to the first room>, <but [Ø] east> this time. <[Ø] Down the vine>, grab the Nirvana Plate, <and go east> for another vine."

Versions of this "directed movement" pattern will also be discussed in the following sections.

4.2 AV/PR deictics and transitional location

The adverbials (AV) *then, here, and now* all occur within the first ten key grammatical items in the VGW corpus. In grammatical terms, these items

express *location*, that is, a static point of reference, whether spatial or temporal, or a point in the development of the text itself. In functional terms, as we see below, each of these items is regularly associated with a distinct set of constructions which express the management of space and time, or distinct stages within the text itself (all of these being functions that are more typical of VGWs).

Looking at *then*, one common pattern with this item takes the form <*then* VV [*imperative*]>. This pattern introduces the next step in a sequence of instructions and generally involves a verb of action or movement (as mentioned above, with a potential ellipsis of the VV) or, less frequently, a verb expressing a communicative or mental process:

- (8) Turn right immediately after exiting, **then** [Ø] **right and left**.
- (9) Head up to his office on 2F. To the item shop! **Then** [Ø] **to the bar**.
- (10) Have 100 Kills by the time you finish the stage. **Then choose** the option “to kill”

The adverbial *now* has a similar directional function to *then*. However, while *then* instructs the player to select a direction or to engage in the next action in a sequence, *now* appears to operate at a higher level of in-game activity: the player is instructed to travel to the next scene or to a separate location. As with *then*, sentence-initial *now* is sometimes accompanied by ellipsis of the following verb, hence the commonly encountered pattern <*now* VV to NN [*location*]>:

- (11) **Now return to 5F to Bridge to Apocrypha** by retracing your steps.
- (12) **Now** [Ø] **south to the next screen**. Across the bridge, then north.
- (13) **Now zoink over to Central Seal Island, SAVE YOUR GAME**, ascend the stairs and...

Sentence-initial *now* also has a function to play in expressions which refer to the metagame; in this pattern, the AV introduces a summative comment on the current state of play:

- (14) **Now comes the fun bit**, pick up the Defense penalty symbol, but don't throw it just yet,
- (15) **Now comes the hard part** in dealing with the Crusaders.
- (16) **NOW, this is important**. If you try to rush it, it will just run away...

We have seen that the AV *then* appears to have a “staging” function in terms of giving immediate directions, while *now* appears to signal a transition between scenes or a commentary on the gameplay. In contrast, *here* (used as a pronominal) is typically found in contexts in which it presents new items

(these are metagame items about the game world, such as *lists*, *locations*, *rewards*, etc.):

- (17) **Here's a list** of those tags.
- (18) **Here are the locations** of all of the frogs in the city:
- (19) **Here are your rewards** for winning: ...

Finally, another significant use of *here* is as a deictic adverbial to briefly describe an encounter, especially in the recurrent pattern *<here you VX ('ll, will) find NN>* (to some extent, this usage brings us back to the sequential use of *then*):

- (20) 1F: **Here you find** 3rd Lift Engine Room,
- (21) **Here you'll find** the final pedestal [*sic*], and the final change for your ring.
- (22) **Here you will find** some houses with strange computers that offer information on the final dungeon.

4.3 CJ conditional advice and choice

In this section, I examine three key subordinating conjunctions (CJ). It is not surprising that items such as *if* and *once* (and related items such as *after*) are salient in the VGT and VGW corpora, as they are associated with the expression of conditional instructions (which as pointed out by Hawreliak [this volume], are part of the underlying computational “source code” of this type of discourse). It is also interesting to note that this pattern is also found in other procedural genres such as boardgame rules, laboratory protocols, recipes, and so on (Coucherut 2016).

The subordinator *if* has a high Keyness score not only in the VGW corpus (rank 8) but also in VGTs (albeit slightly lower down, at rank 15). Although *if* is used in many different contexts, I suggest that it is associated with three general patterns in these corpora. In pattern IF-1, *<if you VV (cancel, cause, connect, land, hit) a NN (attack, poke, knockdown, etc.)>*, the author offers advice to the player in the main clause, although this is conditional upon a move or attack described in the subordinate clause with a technical VV or NN:

- (23) Lightning Kicks are very safe when blocked and so give you some free chip damage **if you can cancel a poke into them.**
- (24) And **if you cause a knockdown**, try using Zangief's s.HK to purposely whiff over the opponent...
- (25) **If you land a Tenshin throw**, this is a good way to capitalize and use up your super meter.

In contrast, pattern IF-2 (<*if an opponent get/s AJ / AV [evaluative]>*) tends to occur in the VGT corpus and is semantically more restricted. This construction involves a main clause offering advice modified by a subordinate clause, in which an *opponent* is described in terms of his/her behavior, movement, or some other quality expressed by an attributive use of the VV *get*. In many cases, the main clause offering advice involves a serial VV such as *try*, as we can see in these examples:

- (26) **If an opponent gets comfy** behind a low guard, *try* jumping backward and tagging the opponent with an instant overheadj.HP¹⁶
- (27) **If your opponent gets wise** to your cross-up attempts, *try* mixing it up with this diving kick.
- (28) **If your opponent somehow get [sic] out of the corner**, don't panic and *try* to switch...

Finally, pattern IF-3, <*if you VV (expect, know, sense)>*, formulates advice to the player in terms of a choice expressed in the main clause, with the conditional clause qualifying this choice as an aspect of the player's affect or mental state. This rather more subtle pattern of advice is typical of the VGW corpus:

- (29) This is an optional battle, so do it only **if you feel like it**.
- (30) It also hits twice, so throw it out **if you sense** a Focus Attack coming.
- (31) **If you want** your arts to wear down all that HP faster, you may consider...

The other key conjunctions to be found in the VGW list are *once* (rank 6 in the VGW) and *after* (rank 12 in the VGW, although this item may be an adverb, preposition, or conjunction). While *if* expresses a clause relation which affects whether the propositional content of the main clause is realized or not, items such as *once* and *after* express circumstantial meanings which affect the manner or means by which the main clause is realized (Halliday and Matthiessen 2014). In general, clauses introduced by *once* refer to in-game situations that have to be accomplished before the next action or event can occur:

- (32) **Once you successfully make the ramp jump**, go NW someways and find Zain [...]
- (33) Have Hero and Atsuro do the same and attack him. **Once he's gone**, gang up on the one at E07 (white dude).
- (34) You have to do the Priestess Door first. **Once that's done**, you can do whichever door you want.

In contrast, *after* as a subordinator is typically used with meta-comments, that is to say, instructions that relate to game controls, a particular stage of the game, or more general gameplay:

- (35) **After this stage is cleared** you cannot go back to any of the maps in Episode 13
- (36) Asgard, **after you've completed disc 1**. This will be your final encounter.
- (37) Her second unique cancel is triggered by tapping PP immediately **after inputting the command** for her Seismic Hammer

Generally speaking, the patterns discussed above suggest that VGWs exploit a much wider range of clause expressions than VGTs, an observation that is also confirmed when looking at the PN *you* and auxiliary verbs, as will be described here.

4.4 PN reformulation and evaluation

There has been much previous research on the key role of pronouns such as *you* in gaming discourse and in other procedural genres (Lassen 2003; Ensslin and Bell 2012; Couterut 2016, among others). It is therefore not surprising to find that there are several pronouns (PN) toward the top of the Keywords list of salient grammatical items in both the VGT and VGW corpora. For space reasons, I cannot here provide a full analysis of these items, but instead I will focus on a small handful of patterns that are associated with *you*, *it*, and *this* (the latter two both ranking highly in the VGT).

Broadly speaking, it is possible to identify three main patterns for *you* in the VGT and VGW corpora: YOU-1: definitions of the player's specific abilities using *can*, as in <*you can VV NN*>; YOU-2: advice expressed by deontic modals, such as <*you VX (better, had better, should,) VV*>; and YOU-3: conditions and comments expressed in subordinate clauses, such as <*CJ (if, when, whenever) you are ready*>.

I suggest that pattern YOU-1 accounts for many uses of both *you* and *can* in both corpora. This usage corresponds to the "radical" (enabling) use of the modal VX *can*. In many contexts, *you can* defines a player's special ability, generally in terms of a technical verb. The following examples come from the VGT:

- (38) Here's a quick list of effective moves during which **you can buffer** the inputs: ...
- (39) **You can easily juggle** the ultra for most hits off of a High Step Kick, an FADC'd Tiger Uppercut, ...
- (40) **You can plink lk** and mp together (which the system will use the mp) and tap out EX legs for a quick confirm into a juggle.

Pattern YOU-2 is similar, but here *you* is the subject of a variety of modal expressions expressing advice, such as <you VX (*be better off, gotta, got to, have better, wanna, want to, will need to*) VV (*learn, rely on, try, use*)>. Here, the lexical verbs are not as specific as in pattern YOU-1; rather they tend to be “conatives” (such as *to learn to, to try to*), that is to say, items that express the relative success or failure of a process rather than the process itself, or that express the subject’s attempts to accomplish a process:

- (41) Really, **you’re better off** *using* the meter for EX attacks and FADCs.
- (42) A good general combo into Gen’s solid super, which **you should learn** well.
- (43) As a Honda player, **you want to try and make** the fight an up-close battle...

Pattern YOU-3 involves a variety of subordinate clauses, in which the player’s state of readiness is presented as a condition on which the next piece of advice or the rest of the adventure is dependent. One extended version of this pattern that is typical of VGWs has the following wording: <CJ (*once, when, whenever*) you are AJ (*ready, set*)>:

- (44) **Once you are set**, focus on taking them down by one.
- (45) **When you’re ready** to proceed, go to the lower level.
- (46) **Whenever you are ready** to leave, speak with Mitsuru.

We now turn to the pronouns *it* and *this*. Both are especially salient in VGTs. Unsurprisingly, *it* has a wide range of uses, most notably as an anaphoric item referring to a specific object or skill. In the VGT corpus, *it* as subject is often used to introduce an evaluating clause. One regular example of this involves the extended LG pattern <CJ (*and, but, so*) it VV AJ (*good, great*) for VV (*catch, dodge, knock, pressure, throw+ing*)>. This pattern provides a positive judgment of a skill or attack that has been defined in the antecedent context, while at the same time specifying what type of attack *it* can be used for:

- (47) Dash Low Straight—Hits low, **so it’s good for catching** opponents as they try to back away from Balrog’s mean face.
- (48) The invincibility on the teleport starts instantly, **so it’s good for dodging** full screen fireballs...
- (49) It’ll pass over the heads of crouching opponents... **But it’s good for knocking** opponents out of the air.

As complement, *it* is also used in another clear example of an extended LG pattern <*use it as a NN (counter, poke)*>, in which the VGT author describes an alternative way of using a previously defined attack:

- (50) The EX version will even go through projectiles so **use it as a counter** to fireballs from mid-range.
- (51) Instead, Gouken's c.MP—go figure—acts a lot like most characters' c.MK, hitting low with decent range and cancelability, so **use it as a poke**.
- (52) If you get knocked down, **use it as a wake-up counter**.

The general grammatical difference between *it* and *this* is that *it* typically reiterates a specific antecedent referent, while *this* potentially introduces an element of reformulation or evaluation, with the possibility of broadening or narrowing the frame of reference (Huddleston and Pullum 2002). One “narrow” use of *this* in the VGT corpus is to introduce the definition of an immediately adjacent term (a type of attack) while introducing an evaluation (clearly a fundamental function of this kind of instructional discourse, as pointed out in Ensslin 2012). As these contexts are typically definitions, the subsequent reformulation renames the specific term for an attack using a hypernym: <*this is AT AJ (good, solid) NN (attack, move, overhead, super, way, etc.)*>:

- (53) Shienkyaku—**This is a good anti-air attack** with good priority.
- (54) Fuhajin—**This is a very interesting projectile move**.
- (55) **This is a pretty good super** that's very easy to combo—for example.

In the VGW corpus, *this* is typically used in a pattern that sums up an entire situation or reformulates an in-game item or event in order to evaluate it. Since this usage involves various different patterns, the following sample gives an idea of the different kinds of structure which share this “summarizing” function (as mentioned above, a similar set of expressions are introduced by the AV *now*):¹⁷

- (56) However, they're very powerful, and not easily beaten **at this point in the game**.
- (57) **The key to this battle** is to exploit the knock back effect of the enemies' normal attack. How? Simple.
- (58) If you want to polish off an opponent **this isn't a bad way to do it**.
- (59) **For this reason**, it's best to use the teleport defensively or to cover distance after knocking down an opponent.

5 Data summary and discussion

In this chapter, I have shown that certain particles (such as *up*, *off*), adverbs (*now*, *here*), conjunctions (*if*, *once*), and pronouns (*you*, *it*) are statistically more likely to occur in VGTs and VGWs in comparison with the general language (and in comparison with other major technical genres). Rather than analyzing these items in isolation, I have argued that function words provide the framework for longer stretches of expression—LG patterns, which I claim are a useful focus of analysis when we are looking initially for the main linguistic characteristics of a particular register or genre.

The analysis of LG patterns shows a number of interesting similarities and divergencies between VGTs and VGWs. Generally speaking, there are three overall sets of patterns which emerge. First, a dominant phraseology¹⁸ that emerges in both corpora involves the *management of the player's moves within the imaginary space of the game*. In VGTs, many LG patterns (formed around particles and prepositions) deal with the manipulation of the joystick or provide a precise definition of fighting moves as in examples such as *Fairly easy to combo into off of a HK Spiral Arrow*. Similarly, in VGWs, many LG patterns (often built around adverbials and textual adjuncts) serve to stage events or tell the player to transition from one scene to the next (Now, you can do two things.). A second set of phraseological patterns involves the *framing of advice*: in VGTs (as in many other instructional genres), conjunctions such as *if* set out the conditions for concrete actions, especially types of maneuver (*If you land the cross-up, combo into a LK Tatsumaki*), while in VGWs, we see a rich diversity of clause relations, which serve to summarize necessary actions or set the scene for events (*Once the first group of Demons is defeated, an enemy raid group will appear at A01*). A related set of phrases involves “directed imperatives,” often structured around sequences of verbs, ellipses, and adverbials such as *then* (*Down back charge then half circle forward then dash*). This usage is also related to the verbal construction of instructions which involves a variety of deontic modals (*If this still doesn't work, you might want to try over compensation.*). Finally, a third set of phraseological patterns is concerned with *tracing the relationships between discourse referents*. Thus, in VGTs, pronouns such as *it* are used to introduce definitional and evaluative constructions (*it's a great anti-air and easily juggles*), while in VGWs, items such as *this* are used to comment on the ongoing state of play (*This is the final event in the 7th day*). Other patterns exist, but the above-mentioned examples represent some of the most recurrent and regular sequences of expression that can be observed in these corpora. It is notable that in each case, we are not only concerned with the behavior of a single grammatical item, but also with its role within an extended pattern of use.

It is important to note that the LG patterns found in VGTs and VGWs are not unique. Even if some patterns have the appearance of specific turns of

phrase, such as (<*if an opponent get/s* AJ/AV [*evaluative*]->) (a pattern more typical of VGTs), or an even longer stretch such as <CJ (*and, but*) / AV (*again, then*) VV (*go, [Ø]*) AV (*back, east, left, north, right*)/PR (*down, to*) NN> (a prototypical pattern in VGWs), it is likely that these constructions can be found in other, similar, contexts elsewhere in the language, especially in related types of texts, especially procedural genres **such as such as** (written) games rules, (oral) street directions, and so on. In addition, it is worth noting that the patterns observed in this study are only the most routine expressions to be found in these corpora. Thus the LG patterns identified above should really only be seen as a core linguistic “background,” against which the more specific features of these texts can be brought into focus.

This point leads me back to the notion of LSP. As I mentioned in the first part of this chapter, the view of language I have adopted here is the systemic functional model (Bloor and Bloor 1986; Stubbs 1993; Halliday and Matthiessen 2014, among others). From this perspective, it is considered that all discourse is necessarily adapted to a specific context of situation (i.e., a *genre*), and as such involves distinct rhetorical functions (such as reporting, recommending, exploring, expounding, etc.). According to this model, specific discourse functions are realized by specific lexical and grammatical phrases, and these are in turn derived from the repertoire of potential LG resources (thus, a functional genre is realized by a *register*, a specific set of linguistic choices taken from the overall system). On the basis of this approach, I would suggest that it is not useful to characterize VGTs or VGWs as genres which can exist somewhere on a continuum between LSP and LGP. Rather, I suggest that there is no such thing as “LGP”; all instances of language can be seen as belonging to one specific register or another, and thus are all forms of LSP. If this hypothesis can be entertained (following Bloor and Bloor 1986), then the important question is not whether VGTs and VGWs are closer to LSP or to LGP, but rather the whether these genres represent highly recognizable, codified, conventionalized genres or whether they are hybrid, emergent, indeterminate text types. On the basis of the above study, and having looked at the prototypical forms of expression to be found in these texts, I am tempted to say that VGTs and VGWs belong to the rather more codified end of the spectrum. Indeed, I would suggest that the phraseological patterns that can be observed in VGTs and VGWs display a very high degree of regularity as well as expressing functions that have been adapted very specifically for the particular purposes of the gamers who have produced them. This is perhaps surprising, because—as we have seen—these texts are produced by amateur fans who are not necessarily professional writers or even proficient speakers of English. It is this degree of stability or “maturity” that I would claim is the most important factor in determining whether these texts represent a recognizable genre or register, since it can be shown that they have developed not only a series of

conventions relating to terminology and other local features of language, but also much broader features such as phraseology.

However, although I claim here to have identified some of the core regularities of expression in VGTs and VGWs, this is only a partial picture. The analysis of LG patterns I have set out above still does not tell us much about patterns of regularity at higher levels of analysis, such as “rhetorical moves” (Swales 1990). In addition, the analysis of phraseology has little to say about the social functions of these texts within a broader “ecology of genres” (Spinuzzi and Zachry 2000), although some corpus analysts, such as Groom (2010), have claimed that the analysis of “semantic sequences” shows the underlying ideology of certain forms of academic writing. In this regard, it would be interesting to explore the relationship between VGTs, VGWs, and other paratexts such as the “post-match report” (a genre associated with sports such as chess, cricket, and tennis in English, etc.), as well as related “alternative” genres such as games-related webseries, game reviews, let’s-play commentaries, and so on. In addition, it would be interesting to see to what extent the language forms appropriated in VGTs and VGWs are “fed back” into other genres of the language, so that texts such as VGTs and VGWs may be seen as nexus points in the development of new forms of English. As Gee (2005) has pointed out, both in-game discourse and paragame genres provide an “interaction space” in which authors and players enact (or re-enact) the social contract of gameplay in verbal form. Thus rather than seeing VGTs and VGWs as overly codified discursive “cul-de-sacs,” it may be possible instead to see them rather as “discourse sandboxes,” in which a significant subset of language users (including presumably many young gamers and speakers of English as a lingua franca, as defined by Seidlhofer 2001) enjoy a relatively safe communicative environment, and thus a high degree of freedom to expand their language use, at the same time as exposing themselves and others to very advanced forms of the English language.

Notes

- 1 In this paper, I use a simplified version of the BNC tagset, where NN = noun, VV = verb, PN = pronoun, PR = preposition, and so on. See Section 4 for the full list.
- 2 I am grateful to my past students Camille Croz, Walter Goguillon, Barbara Paul, and Arnold Savary for their part in designing and building the VGT and VGW corpora.
- 3 Here I assume Halliday’s distinction between “genre” as a social/functional label and “register” as a linguistic/formal label (Halliday and Matthiessen 2014).
- 4 *Streetfighter 4* (2008), fighting videogame, Capcom. Available at: <https://www.playstation.com/en-us/games/street-fighter-iv-ps3/>.

- 5 *The Legend of Heroes: Trails in the Sky* (2004), Japanese role-playing videogame, Nihon Falcom. Available at: <http://trailsinthesky.com/fc/>.
- 6 *Xenoblade Chronicles* (2012), Japanese role-playing videogame, Monolith Soft. Available at: <http://xenobladechroniclesx.nintendo.com/>.
- 7 Available at: <https://www.eventhubs.com/guides/2007/oct/21/street-fighter-terminology-acronyms-lexicon-and-glossary-guide/>.
- 8 *Streetfighter 4* (2008), fighting videogame, Capcom. Available at: <https://www.playstation.com/en-us/games/street-fighter-iv-ps3/>.
- 9 *Persona 3* (2006), Japanese role-playing videogame, Atlus. Available at: <https://www.atlus.com/persona3/>.
- 10 *Tales of Symphonia* (2004), Japanese role-playing videogame, Namco Tales Studio. Available at: http://www.bn-ent.net/cs/list/talesofsymphonia_ps2/.
- 11 *Radiant Historia* (2011), Japanese role-playing videogame, Atlus. Available at: <http://www.atlus.com/radianthistoria/>.
- 12 *Devil Survivor* (2009), tactical role-playing videogame, Atlus. Available at: <https://www.atlus.com/devilsurvivor/>.
- 13 This methodology was first discussed in Gledhill (2000a, b). The reason why ten grammatical items are selected as a cut-off point is that it is difficult to find more than a handful of grammatical items within the first few hundred salient Keywords, as discussed in Section 3 (and as demonstrated in Appendices 3 and 4). If the corpus is large enough and representative enough of a specific genre, then I find that the analysis of up to ten grammatical items gives a good overall picture of the main *n-grams* (repeated expressions) that are particular to this type of text.
- 14 There are still many items which constitute “noise” in Appendices 1 and 2. For example, the keywords list for VGW includes symbols such as *o* and *x* which are not lexical items but are used to decorate the text.
- 15 Here the symbols *VV/AJ* refer to a past participle verb used as a predicative adjective.
- 16 This is one of many abbreviations in the VGT corpus (here *j.HP* = “jump High Punch”).
- 17 As reformulating items, some of these examples of *this* are not PNs but rather adjectival determiners (AT).
- 18 As mentioned in the introduction, ever since the development of large-scale corpus analysis, lexicographers and descriptive grammarians have been interested in examining the routine expressions of not only the general language (Renouf and Sinclair 1991; Stubbs 1993; Hunston and Francis 2000) but also in specific types of discourse (Biber et al. 2004, 2010; Groom 2010, inter alia). Regardless of methodological differences (which are many), many of these linguists refer to such regularities as *phraseology*. It is worth stating here that such a broad use of the word *phraseology* happens to coincide with the way the term is used in the general language, as noted in the Collins COBUILD dictionary: “Phraseology: A set of phrases used by a particular group of people” (Cobuild 2018).

References

Primary sources

- Collins COBUILD (2018), *Collins Online English Dictionary*. Available online: <https://www.collinsdictionary.com/dictionary/english/phraseology> (accessed March 3, 2018).
- Demon King* (2005), "Fighting videogame," *Nintendo*. Available online: <https://www.nintendo.com/games/detail/demon-king-box-3ds> (accessed March 3, 2018).
- Devil Survivor* (2009), "Tactical role-playing videogame," *Atlus*. Available online: <https://www.atlus.com/devilsurvivor/> (accessed March 3, 2018).
- Eventhubs, "Videogame tutorial archive." Available online: <https://www.eventhubs.com/moves/sf4/> (accessed December 5, 2018).
- Gamefaqs, "Videogame walkthrough archive." Available online: <https://gamefaqs.gamespot.com/psp/933329-the-legend-of-heroes-trails-in-the-sky/faqs> (accessed March 3, 2015).
- Persona 3* (2006), "Japanese role-playing videogame," *Atlus*. Available online: <https://www.atlus.com/persona3/> (accessed March 3, 2018).
- Radiant Historia* (2011), "Japanese role-playing videogame," *Atlus*. Available online: <http://www.atlus.com/radianthistoria/> (accessed March 3, 2018).
- Shoryuken (2011), "Tutorials." Available online: <http://shoryuken.com/category/tutorials-2/> (accessed March 3, 2018).
- Streetfighter 4* (2008), "Fighting videogame," *Capcom*. Available online: <https://www.playstation.com/en-us/games/street-fighter-iv-ps3/> (accessed March 3, 2018).
- Tales of Symphonia* (2004), "Japanese role-playing videogame," Namco Tales Studio. Available online: http://www.bn-ent.net/cs/list/talesofsymphonia_ps2/ (accessed March 3, 2018).
- The Legend of Heroes: Trails in the Sky* (2004), "Japanese role-playing videogame," Nihon Falcom. Available online: <http://trailsinthesky.com/fc/> (accessed March 3, 2018).

Secondary sources

- Adam, J.M. (2011), "Les consécutives intensives: Un schéma syntaxique commun à plusieurs genres de discours," *Linx*, 64–65: 115–31.
- Anthony, L. (2014), "AntConc (Version 3.4.3)" [Computer Software], Tokyo, Japan: Waseda University. Available online: <http://www.laurenceanthony.net/> (accessed April 3, 2018).
- Biber, D., S. Conrad, and V. Cortes (2004), "If you look at...: Lexical bundles in university teaching and textbooks," *Applied Linguistics*, 25 (3): 371–405.
- Biber, D., R. Reppen, and E. Friginal (2010), "Research in corpus linguistics," in R. Kaplan (ed.), *The Oxford Handbook of Applied Linguistics*, 548–67, Oxford: Oxford University Press.
- Bloor, M. and T. Bloor (1986), "Languages for specific purposes: Practice and theory," *Centre for Language and Communication Studies Occasional Papers* 19, Dublin: Trinity College CELTS.

- Coutherut, M. (2016), “Les textes procéduraux en Anglais: Création d’une échelle de prototypicalité,” PhD thesis, Université Paris Diderot.
- Ensslin, A. (2011), *The Language of Gaming*, Basingstoke: Palgrave Macmillan.
- Ensslin, A. and A. Bell (2012) “‘Click = Kill’: Textual you in ludic digital fiction,” *Storyworlds* 4: 49–74.
- Firth, J.R. (1957), *Modes of Meaning: Papers in Linguistics, 1934–1951*, Oxford: Oxford University Press.
- Gee, J.P. (2005), “Semiotic social spaces and affinity spaces: From The Age of Mythology to today’s schools,” in D. Barton and K. Tusting (eds.), *Beyond Communities of Practice: Language, Power, and Social Context*, 214–33, Cambridge: Cambridge University Press.
- Gledhill, C. (1995), “Collocation and genre analysis: The phraseology of grammatical items in cancer research abstracts and articles,” *Zeitschrift für Anglistik und Amerikanistik*, 43 (1/1): 11–36.
- Gledhill, C. (2000a), *Collocations in Science Writing*, Tübingen: Gunter Narr Verlag.
- Gledhill, C. (2000b), “The discourse function of collocation in research article introductions,” *English for Specific Purposes*, 19 (2): 115–35.
- Gledhill, C. (2015), “On the phraseology of grammatical items in lexicogrammatical patterns and science writing,” in P. Thompson and G. Diani (eds.), *English for Academic Purposes: Approaches and Implications*, 11–42, Newcastle upon Tyne: Cambridge Scholars Publishing.
- Groom, N. (2007), “Phraseology and epistemology in humanities writing,” PhD thesis, University of Birmingham.
- Groom, N. (2010), “Closed-class keywords and corpus-driven discourse analysis,” in M. Bondi and M. Scott (eds.), *Keyness in Texts*, 59–78, Amsterdam: John Benjamins.
- Halliday, M.A.K. and C.M.M. Matthiessen (2014), *Halliday’s Introduction to Functional Analysis*, 4th edn., London: Arnold.
- Hoey, M. (2005), *Lexical Priming: A New Theory of Words and Language*, London: Routledge.
- Huddleston, R. and G. Pullum (2002), *The Cambridge Grammar of the English Language*, Cambridge: Cambridge University Press.
- Hunston, S. and G. Francis (2000), *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*, Amsterdam: John Benjamins.
- Lassen, E. (2003), *Accessibility and Acceptability in Technical Manuals*, Amsterdam: John Benjamins.
- Née, E., F. Sitri, M. Veniard, and S. Fleury (2017), “Routines discursives et séquentialité dans des écrits professionnels: La mise au jour d’une séquence évaluative,” *Corpus*, 17: 145–68.
- Renouf, A. and J. Sinclair (1991), “Collocational frameworks in English,” in K. Aijmer and B. Altenberg (eds.), *English Corpus Linguistics*, 128–43, London: Longman.
- Seidlhofer, B. (2001), “Closing a conceptual gap: The case for a description of English as a lingua franca,” *International Journal of Applied Linguistics*, 11: 133–58.
- Sinclair, J. McH. (1991), *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

- Sinclair, J. McH. and A. Mauranen (2006), *Linear Unit Grammar: Integrating Speech and Writing*, Amsterdam: John Benjamins.
- Spinuzzi, C. and M. Zachry (2000), "Genre ecologies: An open-system approach to understanding and constructing documentation," *Journal of Computer Documentation*, 24 (3): 169–81.
- Stefanowitsch, A. and S. Gries (2003), "Collostructions: Investigating the interaction between words and constructions," *International Journal of Corpus Linguistics*, 8 (2): 209–43.
- Stubbs, M. (1993), "British traditions in text analysis. From Firth to Sinclair," in M. Baker, G. Francis, and E. Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair*, 1–33, Amsterdam: John Benjamins.
- Swales, J. (1990), *Genre Analysis: English in Academic and Research Settings*, Cambridge: Cambridge University Press.
- Tran, T., A. Tutin, and C. Cavalla (2016), "Typologie des séquences lexicalisées à fonction discursive dans la perspective de la rédaction scientifique," *Cahiers de Lexicologie*, 108: 161–79.

Appendices

APPENDIX 1 *The first ten keywords from the VGT corpus (568,998 tokens)*

Rank	Freq.	Keyness	Item
1	4247	57788.557	Vf
2	4249	57594.019	gif
3	946	11684.395	combo
4	900	10988.616	downright
5	878	8885.698	opponent
6	438	5934.267	jpg
7	551	5678.751	ultra
8	451	5671.799	MK
9	444	5473.516	EX
10	362	4925.702	streetf

*APPENDIX 2 The first ten keywords
from the VGW corpus (111,695 tokens)*

Rank	Freq.	Keyness	Item
1	54588	606116.703	
2	2305	25593.519	Lv
3	1346	14400.420	xx
4	1060	11069.711	HL
5	1076	10531.019	location
6	1272	8291.952	x
7	1384	7750.775	o
8	758	7618.154	enemy
9	1491	7077.486	E
10	681	6878.193	quest

*APPENDIX 3 The first ten grammatical
keywords from the VGT corpus*

Rank	Freq.	Keyness	Item
13	1300	4193.935	down
15	1200	4105.332	right
39	1157	1675.753	can
45	1866	1542.512	you
58	927	1113.553	up
102	210	681.547	forward
103	611	656.297	into
164	1555	373.658	it
192	52	311.637	lets
211	272	273.407	off

APPENDIX 4 *The first ten grammatical keywords from the VGW corpus*

Rank	Freq.	Keyness	Item
21	6351	4983.754	you
30	608	3755.339	up
158	1120	1010.787	get
257	1230	712.373	then
308	1741	608.446	can
394	220	501.225	once
690	91	272.549	down
821	563	219.931	if
1049	445	162.058	here
1109	245	149.288	now