French Language Studies **14** (2004), 25–40, © Cambridge University Press
DOI: 10.1017/S0959269504001395

Code and Norm Revisited

MICHAEL D. KLIFFER

McMaster University

(Received December 2003; revised January 2004)

ABSTRACT

This paper deals with the relationship between rules and conventionality, as reflected in a case study of the two structures N_1 *de* N_2 and N_1 *du* N_2. Most occurrences of these can be explained via two rules which evoke ease of referent identification and are a subset of the principles governing the ± definiteness opposition in French. Via dictionaries and *Google* searches, however, we detect variation that our rules cannot predict, typically when N_2 is non-countable and abstract. Local semantic contrasts, i.e. dependent on the lexical content of N_1 or N_2, further complicate the picture. We conclude that Coseriu's *Norm*, alias conventionality, must be recognized as a factor coexisting with the rules.

1 INTRODUCTION

This paper will argue a point often ignored or downplayed by linguists: grammatical rules have their limits. As Bolinger (1976) observed, language may be a system where things hold together, but it has as much patching and gluing as it does architectonics. This message is crucial for language learners, who tend to view rules as a panacea. The more advanced learners, as well as insightful beginners, realize that morpho-syntax in the sense of *servitudes grammaticales* (best represented today by generative theory, notably in the preponderance it assumes for the 'computational component' of linguistic competence) can help only to a point. Sooner or later the learner confronts data which involve either a localized semantic opposition, not at all predicted by the morpho-syntax, or a structure which simply must be learned as is, which no amount of linguistic or social ratiocination can explain.

Granted, not all linguists adhere to Meillet's view of language as a system where everything holds together. Guillaume (1919/1975) treated the article in French with a fine-grained analysis, never since equalled, of heterogeneous semantic/psychological distinctions. Coseriu (1952/1978) proposed the now classic distinction between system (consisting of functional oppositions) and norm, based on social convention. Our analysis in this paper is essentially a reconciliation of these two components. In a similar vein, Fillmore et al. (1988) advocates an approach which integrates into a continuum 'knowing' (i.e. conventional, stored knowledge) and 'figuring out' (using productive morpho-syntactic rules). Furthermore, recent work in Optimality Theory attempts to present a unified treatment of 'hard'

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

grammatical constraints and 'soft' usage preferences via a stochastic – that is, probabilistic – model, as seen in Bresnan et al. (2001).

We will illustrate this need for harmonization of rules and conventionality by examining what at first looks like a very straightforward opposition: the two French structures N_1 *de* N_2 and N_1 *du* N_2, where the portemanteau form *du* is to be read as including *de la* and *des* as well. I will first give a synthesis of the two major syntactic/semantic rules which favour one or the other. Next we will demonstrate the applications of these rules to data where first N_1 and then N_2 is kept constant, after which I will discuss the results of a questionnaire regarding a minimal pair. Finally, we will look at data that appears beyond the scope of general rules and consider the theoretical implications of such apparent collocations.

2 THE METHODOLOGY

Supplementing the dictionary-based data are selections from Internet searches using *Google*, which provides access to approximately 8 million French-language sites that constitute a veritable mega-corpus. With *Google*'s advanced search feature, it is possible to set up quickly and easily a concordance for each targeted expression. While I have not rigorously verified *Google*'s reliability, searches for collocations like *grève de la faim* 'hunger strike' give frequencies matching what we would expect, i.e. almost zero for this expression without *la* and thousands of hits for *de la*. Moreover, the variation manifested in the search for *soif du/de pouvoir* 'thirst for power' generally matched the variability in native speakers' responses to the choice of linker for this same expression. Of course, use of *Google* as a corpus necessitates certain caveats, such as being wary of syntactically irrelevant hits and, more specific to this study, unwanted qualifiers for N_2. Such problems, though, presumably arise with other concordances too.

3 THE RULES

I will propose two rules which handle selection of *du* or *de* to link N_1 and N_2. They reflect a synthesis of, first, the few allusions to *du/de* alternation in normative grammars, such as Byrne and Churchill (1986: 24–25) and Hooke and Miller (1962: 192–3), and second, of numerous attemps to discern the pragmatic and cognitive facets of definiteness and referentiality, such as Hawkins (1978), Corblin (1987: 8–9), Flaux (1992/1993), Lyons (1999: 2–4), and Croft (2003: 4–7).

1. Use *du* if: (*du* is shorthand for any of the *de* + definite article combinations *du, de la, de l'* and *des*)
   a) N_2 is generic:
      (1) ministère des *finances* 'Finance Ministry'
      (2) un goût du *débat* 'a taste for debate'
   b) N_2 is particularized, whether through context or co-text:
      (3) la fin du *mois* 'the end of the month' (i.e. this particular month)

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

(4) un soupçon des *ennuis* qui nous attendent 'a taste of the troubles awaiting us'

At first glance, this distribution seems strange: 'generic' would appear to be the polar opposite of 'specified'. Specialists in cognition and pragmatics, though, have suggested that these two categories have in common the property that neither poses a problem of referent-identification. Givón (1984: 406–35), quoted in Popescu-Belis (2001) proposed that definiteness and referentiality distinctions be viewed as a hierarchy defined in terms of specificity (S) and presupposition (P): 'definite' (+S + P) > 'referential indefinite' (+S − P) > 'non referential' (−S − P) > 'generic' (− S + P). He situates these points on a helix, in which the two extreme cases are contiguous, albeit on a different level (individuals vs. categories). He also defends a cognitive model of reference understanding that gives a central role to the activation of the referents in the working memory of the speakers. This is also one of the conclusions arrived at by Hopper and Thompson (1984): in a wide variety of cases, over several languages, the articles and other nominal markers are assigned on the basis of discourse salience or preeminence rather than following strictly the logical status of the referent. Since generics include the totality of members of the set denoted by N_2, pinpointing a referent for them is simply not an issue. The referent for particularized nouns is obvious, thanks to their individuated status.

2. Use *de* if N_2 is non-specific:

(5) une histoire de *guerre* 'a war story'
(6) Permet-on l'emploi de *caractères phonétiques*? 'Is the use of phonetic characters permitted?'[1]

This is simply the corollary to rule 1. N_2 is often used merely to qualify N_1, as if it were an adjective, so that N_2 in such cases is devoid of individuation, i.e. its reference is problematic in that it points to neither a specific instance of N_2 nor the totality of referents.

These two rules do not exist in a vacuum. They are in fact localized versions, applied to the N_1 *de/du* N_2 context, of the general principles underlying the distinction between definiteness and indefiniteness in French. Such rules are indispensable for explaining the great majority of *de/du* alternations. Our study will focus on one semantic type of N_2 that is the most problematic for these ground rules: the abstract noun. Such nouns are a challenge because the criteria of definiteness and referentiality do not apply to them in a straightforward fashion, i.e. unlike concrete nouns, it is not obvious how to measure the genericness or specificity of an abstract noun.

These two rules are not the whole story either. With the names of countries, *de* alone is used if N_2 is feminine, but *du* is required with masculine N_2s. While the requirement for *du* with masculine N_2s is indeed categorical, feminine countries

[1] This example illustrates the 'règle de cacophonie' proposed in Gross (1967), a deletion rule that applies more generally than just to our N_1 *de* N_2 structure, i.e. whenever prepositional *de* and partitive determiner *du* are underlyingly juxtaposed.

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

manifest an alternation between *de* and *de la*, which I will attempt to show is a consequence of the two major rules proposed.

Moreover, there can be a syntactic domino effect: if the indefinite article qualifies N_1, there is a likelihood that N_2 also will be non-specific, i.e. occur with *de* and conversely, the definite article before N_1 may favour *du* before N_2:

(7) une histoire d'amour 'a love story' vs. l'histoire de l'amour 'the history of love'

Just like the oft-claimed domino effect in politics, this correlation is far from being inevitable. It is not difficult to find counter-examples like:

(8) un accident du travail 'a work-related accident', le coût de production 'production cost'

Moreover, even if we restrict ourselves to the quasi-collocations in (7), it is possible to switch the *du/de* alternation. *L'histoire d'amour* could be a love story already mentioned in the discourse, while *une histoire de l'amour* could mean a particular historical account of love through the ages. The data culled from the Internet show an undeniable tendency to make N_1 and N_2 carry the same value for ± definiteness, but the semantic independence of the *du/de* contrast can usually override this *servitude grammaticale*, as we will see.

4 PRODUCTIVITY OF THE TWO RULES OF REFERENTIALITY

Returning to the two rules initially given, we will now demonstrate how productive they are by examining the logical possibilities of alternation, i.e. holding N_1 constant, holding N_2 constant and minimal pairs where neither N changes.

5 HOLDING N_1 CONSTANT

Let us first consider the following contrast:

(9) l'âge d'or 'the golden age' vs. l'âge de la pierre, du fer, du bronze 'the stone/iron/bronze age' (Guillaume, 1919/1975: 136)

As in English, *l'âge d'or* is a metaphor: gold is used as an epithet to suggest a flourishing, usually cultural. The other metals take *du* because they are to be interpreted literally and generically. How can we reconcile these interpretations with our two rules? For *l'âge d'or*, N_2 is clearly non-referential, devoid of specificity and genericness (in the sense of applying to all instances of the metal). In fact, if the genericness of the metal gold were intended, the result would be the ahistoric but linguistically impeccable *l'âge de l'or* 'the gold age', structurally and semantically matching the other ages.

Another revealing alternation occurs with *sentiment* 'feeling' as N_1. If its relationship with N_2 is superordinate, i.e. when, as Guillaume puts it, by 'permanent definition' N_2 is a type of feeling, *de* alone arises:

(10) sentiment de pitié 'feeling of pity', de peur 'of fear', de solitude 'of loneliness'

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

I would argue that the superordinate relationship evinced here is a corollary of rule 2, since N_2 is simply a qualifier of N_1 and therefore lacks the pinpointable referent or generic status which would require *du*. When on the other hand N_2 comes across as salient, as notionally independent of N_1, *du* occurs:

(11) sentiment de la musique, des arts 'feeling for music, the arts'

Neither music nor the arts are types of feeling; they are domains which exist on their own, with no necessary link to *sentiment*. Consequently, they are interpreted here as true generics, for which referent determination is not a problem since the entire set of N_2 is referred to. Hence rule 1 applies.

If N_1 is an institution, the data reveal a skewing based on whether N_2 is singular or plural. If singular, *de* is the usual result: (the figures are the number of *Google* hits)

(12) club – patinage artistique (de: 1 200; du: 2) 'figure skating club'
(13) département – histoire (d': 13 700; de l': 0 (where *histoire* is unqualified)) 'department of history'
(14) école – danse (de: 12 600; de la: 29 (nearly all the latter have *danse* qualified)) 'dance school'
(15) institut – psychiatrie (de: 314; de la: 3) 'psychiatric institute'

If N_2 is plural, *des* is the norm:

(16) école – beaux-arts (des: 20 700; de: 282) 'school of fine arts'
(17) école – langues vivantes (des: 1 250; de: 28) 'school of modern languages'
(18) faculté – lettres (des: 24 300; de: 3 360) 'faculty of arts'
(19) département – littératures (des: 1 880; de: 179)
(20) club – sports (des: 5 250; de: 719)

We hypothesize that with the singular, N_2 constitutes an expected sub-division within N_1, i.e. N_2 is notionally subordinated, analogous to the previously given superordinate cases, and therefore lacks the salience that would justify adding the article. For the plural, Guillaume proposes that the article arises as a way of unifying the domain of N_2, of linking all its sub-realms. Unification is not an issue with the singular, hence no article. In terms of our two referentiality rules, the singular's lack of salience entails a non-obvious referent and hence rule 1 applies, while the need to 'unify' the plural leads to the totalizing, generic result required by rule 2.[2]

[2] If a grammatically plural N_2 is notionally singular, *de* predominates:

(i) département – mathématiques (des: 455; de: 13 300)
and vice-versa:
(ii) Faculté – sport (du: 221; de: 98)

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

6 HOLDING N_2 CONSTANT

The same contrast is apparent if we hold N_2 constant. With *jour* 'day' as N_2, *de* renders this noun indistinct as to the actual day, denoting simply daytime as opposed to nighttime:

(21) service de jour 'day duty', hôpital de jour 'outpatient clinic', infirmier de jour 'day nurse', équipe de jour 'day shift'

If *du* arises, the referent of *jour* is a specific day, normally 'today':

(22) plat du jour 'today's special', évangile du jour 'today's Gospel', oeuf du jour 'freshly laid egg', carte du jour 'today's menu'

De jour thus functions as a quasi-adjective qualifying N_1, while *du jour* has an easily identifiable, deictic referent. *du jour* may also signal a generic, as in:

(23) lumière du jour 'daylight'

where the sense is light whose source comes from daytime, i.e. any *jour*.

So far we have given an either/or impression, namely that the two rules make *du* OR *de* inevitable with a given N_1/N_2 combination. The data are more complex than this complementary distribution suggests. With feminine countries, especially *la France*, there is considerable variation. With a unique referent like a country, the explanation cannot be solely in the ease of referent determination; if this were the main or only factor, we should expect *de la* all the time. In fact, pedagogical grammars stipulate *de* before a feminine country, but *du* before a masculine one, a *servitude grammaticale* that ignores semantic properties like referentiality. Our *Google* searches bear out the obligatory use of *du* with masculine countries, but give only statistical backing to *de* + feminine country. With N_1s like *ambassade* 'embassy', *histoire* 'history', and *catholiques*, *de France* is the overwhelming preference. For instance *histoire de France* had about 18 300 hits, versus 3 400 for *histoire de la France*, though the latter figure is reduced by about half if we exclude hits where *France* is qualified by adjectives like *libre* 'free', *contemporaine*, and *coloniale*. Yet, with an N_1 like *maux* 'troubles', the figures are reversed: only 4 hits for *maux de France* versus 46 for *maux de la France*. I believe that our two rules once again are consistent with such variation. With N_1s of the *ambassade* and *histoire* type, we are dealing with expected categories: any classification of embassies and histories is likely to have 'country' as a, if not the, major criterion, so that *France* in such contexts is simply a pigeon-hole for the purposes of classification, taxonomically subordinate to N_1. With *maux*, on the other hand, there is no routine categorization, as this noun lends itself to classification on virtually any scale. *France* in this case stands out because it represents a far less expected category than is the case with embassies or histories.

7 HOLDING BOTH NS CONSTANT

We will now examine a minimal pair, *soif du/de pouvoir* 'thirst for power'. For a questionnaire, I selected from a *Google* search 20 in-context tokens of this pair, with half manifesting each possibility, *du* or *de*. I then submitted the 20 tokens,

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

each in its original context but with *de/du* replaced by a blank, to nine native speaker colleagues, and asked them to indicate their preference for *de*, *du* or either one. The data and results are reproduced in Appendix 1. I reasoned that French instructors' judgments would be as reliable as those from linguistically naive informants because the point in question carried no socio-linguistic baggage, such as a written/oral divergence, liable to skew a purist's evaluation.

A majority of the informants agreed with 10 of the 20 tokens, a majority disagreed with 6 and the remaining 4 tokens showed a 50/50 split. Clearly, we are not dealing with a categorical rule.

The one overt, textual property linking the tokens where *du* predominated is mention of a specific figure with political or military power, namely, Jean Chrétien (former prime minister of Canada), Simon de Montfort (a leader of the medieval crusades), Slobodan Milosevic, and Jacques Chirac. *Pouvoir* here has a sense of hegemony, authority or even omnipotence, i.e. it is a question of having it or not having it. Moreover, in contexts without *soif*, *pouvoir* in this all-or-nothing sense occurs with the definite article:

(24) détenir le pouvoir 'to hold power', le parti au pouvoir 'the party in power'

Where *de* was preferred, the context suggests that we are dealing with a scalar property, a matter of more or less power, in the sense of partial control over one's economic, political or social destiny. For example:

(25) Les charnières sont la peur, les frustrations, les tourments, les ambitions, la soif *de* pouvoir et la fascination de l'argent. 'The crucial elements are fear, frustration, suffering, ambition, thirst for power and obsession with money'.

Looking at contexts without *soif* where *pouvoir* has this non-absolute sense, the determiner we find is the partitive:

(26) acquérir du pouvoir 'acquire power', perdre du pouvoir 'lose power', garder du pouvoir 'retain power'.

Let us turn again to our two rules. The sense of *pouvoir* as an absolute, something associated with a political, military, or religious leader, reflects a totalizing, generic status, and hence rule 1 applies to produce *du*. If *pouvoir* manifests a partitive sense, a matter of degree, its non-definite status fits with rule 2, with *de* resulting.

Of course, the pigeon-holing of *pouvoir* into one or the other category is a matter of perception. The mention of an authority did not guarantee *du*, nor did the lack of one ensure *de*, as seen in the differences in subjects' preferences. One insightful subject remarked that the context did not suffice to clarify if it was a question of 'political power, as in the case of a head of state (*soif du pouvoir*)' or 'any form of power in any domain' (*soif de pouvoir*). I suspect that even if I had included the full discourse for each token, there would still have been variation in subjects' responses, because we are dealing with a semantic opposition whose choice is ultimately up to the speaker, depending on which sense of *pouvoir* is intended.

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

8 COLLOCATIONS

As useful as the two rules are, the learner inevitably encounters data that resist them. If we hold N_1 constant, we find alternations like

(27) mal de mer 'seasickness'

vs.

(28) mal de la route 'carsickness'

Here the semantic parallel is obvious: in both cases N_1 has the same sense ('sickness') and both N_2s force us to infer the same type of indisposition, namely motion sickness. One would be hard pressed to claim that *route* here is any more specific than *mer*: in neither is it a case of any particular road or sea. Likewise, neither seems more generic than the other and they seem evenly matched in terms of the ease with which we could pinpoint their referent, i.e. not an easy task with either. And yet, both dictionaries and the Internet searches which I attempted show that these distinct structures are indeed the fixed way of encoding what seems to be an identical N_1/N_2 relation.

Another challenge is the *de/du* alternation between *théorie de la pertinence* and *principe de pertinence*, both occuring in Sperber and Wilson (1989). One might first suspect that N_1 holds the key, but a *Google* search confirms that *théorie* and *principe* occur in substantial numbers with both structures (see Appendix 2). In fact, the alternation can usually be explained by the two referential rules. For instance, in

(29) théorie du chaos 'chaos theory' du: 3 570; de: 27

N_2 is a typical generic, so its referentiality is not a problem, hence rule 2 applies, while in:

(30) théorie de base 'basic theory' de: 740; de la: 77

N_2 is clearly a modifier of N_1 and lacks referentiality; if *base* is part of an area of knowledge lending itself to theorizing, *du* is the norm, e.g.:

(31) théorie des bases de données 'data-base theory' des: 206; de: 5

Clearly, then, N_1 alone cannot explain the skewing *théorie de la pertinence* vs *principe de pertinence*.

If we look at the chapter 'Pertinence' in Sperber and Wilson (1989), we find that *principe de pertinence* is introduced more than two-thirds of the way through this chapter. Although *théorie de la pertinence* does not arise, there are numerous occurrences with 'N_1 de la pertinence' near the beginning of the chapter, e.g. *les conditions de la pertinence, la définition de la pertinence*. These facts suggest the following: for *théorie de la pertinence*, N_2 is by its nature more salient than N_1, i.e. in the context of Sperber and Wilson's research and their background as philosophers and linguists, we can take for granted that their goal is theory-building; the element deserving the limelight is the content of the theory, relevance. With *principe de pertinence*, N_1 has more salience because at this point in the chapter, they have developed a cognitive definition of 'relevance' and their goal is now to show how it can be explained as

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

a principle underlying communication. Indeed, the *Google* data suggest that where *principe de la pertinence* occurs, it is typically the first mention of *pertinence*.[3]

Moreover, with an abstract noun as N_2, there is a predilection for *du* if N_1 is a nominalization or paraphraseable as a verb:

(32) spécialiste du sport 'sports specialist' (du: 316; de: 7) cf. se spécialiser dans le sport 'specialize in sport'
(33) goût du bonheur 'taste for happiness' (du: 1 140; de: 92) cf. goûter le bonheur 'taste happiness'
(34) animateur du patrimoine 'heritage coordinator' (du: 592; de: 24) cf. animer le patrimoine artistique 'promote one's artistic heritage'
(35) prévention de l'alcoolisme 'alcoholism prevention' (de l': 5 060; d': 19) cf. prévenir l'alcoolisme
(36) refus du plaisir 'refusal of pleasure' (du: 43; de: 7) cf. refuser le plaisir
(37) amour du risque 'love of risk' (du: 957; de: 2) cf. aimer le risque

At first glance, the key appears to be the de-verbal nature of N_1, as seen when we contrast the above with:

(38) professeur de sport (de: 1 880; du: 3)

where, except as an archaism, there is no *professer une matière* 'teach a subject'. The institutional data already given supports this correlation, e.g. in:

(39) département de chimie 'chemistry department'

de la would be excluded because *département* lacks a verbal analogue.
However, there are numerous counter-examples:

(40) étudiant de français 'student of French' (de: 767; du: 30) cf. étudier le français
(41) arrêt de travail 'work stoppage' (de: 27 000; du: 541 (without 'travail du/des/de')) cf. arrêter le travail[4]

Moreover among non de-verbal nouns, either possibility exists:

(42) ingénieur du son 'sound engineer' (du: 15 700; de: 560)
(43) ingénieur du génie civil 'civil engineer' (du: 150; de: 66)
(44) ingénieur de la connaissance 'knowledge engineer (artificial intelligence, computer science)' (de la: 72; de: 4)

[3] For example, the Web citations: . . . *le CMILT permet le respect des principes suivants: – principe de la pertinence. – principe de l'efficacité. – principe de l'optimum économique* . . . Source: www.itu.org.eg/coe/interconnection/08-F%20ZOUAKIA%20ANRT-Morocco%20New%20Challenges.ppt –

> *La référence dans la fiction repose sur deux principes qui, tout en se retrouvant ailleurs, ont depuis toujours formé le noyau de l'ordre fictionnel – le [114] principe de la distance et le principe de la pertinence.*
> Source: www.litterature-reunionnaise.org/marimoutou/partie2/chapitre3.htm

[4] With *préférence de langue* (de: 668; de la: 4), the correspondence is with 'préférer quelle/une langue' rather than '..la langue', i.e. the low referentiality of N_2 is clearly what favours *de*.

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

(45) accident du travail 'work accident' (du: 31 600; de: 11 400)
(46) accident de la circulation 'traffic accident' (de la: 11 200; de: 2 430)

vs.

(47) ingénieur de fabrication 'manufacturing engineer' (de: 295; de la: 3)
(48) ingénieur de vol 'flight engineer' (de: 526; du: 5)
(49) ingénieur d'études 'design engineer' (d': 15 700; des: 306)
(50) ingénieur de recherche 'research engineer' (de: 12 000; de la: 28)
(51) accident de santé 'sudden health problem' (de: 480; de la: 0)

So, can we really talk about 'collocation', i.e. fixed combinations? The *Google* figures suggest a cline: at one extreme, e.g. *accident de santé*, we indeed have a rigid structure; at the other, e.g. *accident du travail*, there is a high (over 33%) tolerance of variation.

If N_2 has a non-specific, non-generic sense, the second referentiality rule routinely applies and there is a straightforward correspondence with verb phrases:

(52) vendeur de montres (de: 58; des: 0) 'wristwatch salesman'
(53) faiseur de miracles (de: 637; des: 5) 'miracle worker'

Both examples have VP correlates with the partitive before the object, i.e. *vendre des montres* 'sell watches' and *faire des miracles* 'work miracles'. The problem cases, then, involve as N_2 non-countable abstract nouns, which may not present a clear referential status, i.e. they could in a given context be taken as partitive or generic.

9 LOCALIZED SEMANTIC CONTRASTS

Semantically unmotivated contrasts like *mal de mer/mal de la route* are few. In some cases, an isolated semantic contrast correlates with the structural distinction:

(54) Le Havre a *le mal de la mer*, comme d'autres ont *le mal du pays*. 'Le Havre has a longing for the sea, just as others long for their country (i.e. are homesick)'. (http://www.humanite.presse.fr/journal/1999-07-31/1999-07-31–293978)

Adding the article before *mer* changes the sense from 'sickness' to 'longing', parallel to collocations like *mal du pays* and

(55) mal du pouvoir 'longing for power', mal du voyage 'longing to travel', mal du nord 'nostalgia for the north'

However, this is a highly localized semantic contrast: unlike *mal de la mer*, *mal du voyage* is ambiguous between 'longing to travel' and 'travel sickness', so that the mere inclusion of the article does not force the reading of *mal* to be 'longing'.

In other cases, the semantic role of N_2 will correlate with the structural difference. With *grève* 'strike', *du* occurs with N_2s which constitute the means, as in:

(56) grève de la faim 'hunger strike', grève du zèle 'work-to-rule', grève de l'impôt 'tax strike'

while *de* alone occurs when N_2 is the motive, e.g.:

(57) grève de protestation 'protest strike', grève de solidarité 'solidarity strike'.

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

This 'local' contrast between means and motive is analogous to the many others which Guillaume (1919/1975) painstakingly examines. Such lexically-specific contrasts undoubtedly make for challenging learning, even for advanced students, but they testify to the efficiency with which the linguistic code exploits this minimal *de/du* opposition.

These data prompt us to take a deeper look at Coseriu's *norm*, alias conventionality. An operational definition for considering a given combination 'conventional' or fixed would be a very low incidence in our concordance for its competitor, e.g. ingénieur de vol 'flight engineer' (de: 526; du: 5). Low incidence is necessary but insufficient because it can reflect an unequivocally high or low referentiality status of N_2 so that the choice of rule 1 or 2 is obvious, e.g. *vendeur de montres* (de: 58; des: 0) 'wristwatch salesman'. The vast majority of candidates for conventionality involve abstract, usually non-countable N_2s whose referentiality is uncertain, i.e. they could be seen as generic or partitive even if the context is held constant. A comparative piece of evidence supports this: in Spanish and Portuguese, the equivalents of *grève de la faim* 'hunger strike' both occur without the definite determiner (Sp.: *huelga de hambre*; Port.: *greve de fome*), even though these languages have close analogues of our two referentiality rules and present no language- or culture-specific peculiarities for the notion of 'hunger strike'.

Moreover, conventionality does not necessarily equate with arbitrariness: as we have seen, there is usually some semantic or pragmatic motivation underlying a fixed combination. Furthermore, in cases where both structures have significant incidence in nearly identical contexts, e.g. *accident du travail* (du: 31 600; de: 11 400) 'work-related accident', the choice arguably reflects competition between equally plausible treatments of N_2 as a salient generic or as a non-referential appendage to N_1.

10 CONCLUSION

The first contribution this paper has attempted to present is methodological: via our *Google* searches, we have brought out the constant variation between *de* and *du*, showing that the alternation reflects other factors than just the complementary distribution evident in dictionary treatments. Any other large corpus would have likely given a similar picture, but *Google*'s advantages are its accessibility, its minimal cost, and the speed with which one can arrive at coherent patterns of usage. Of course, there are necessary caveats, notably in regards to the lexical and syntactic environment which follows N_1 *de* N_2, but the data in this paper has been culled with such potential skewings in mind.

In accord with a venerable linguistic tradition, I have given much attention to demonstrating the explanatory power of two rules of referentiality, which are furthermore but a subset of the general principles handling the ± definite opposition in French. However, should language students have only these two rules to help them correctly produce the two competing structures, their competence would be less than native. Most of their utterances involving the *de/du* contrast

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

would, admittedly, have correlates in the speech of *some* native speakers. This was demonstrated in the *Google* data, where a zero count is rare even given an overwhelming preference for the other structure. Yet, the non-natives' production would still come off as marked, in relation to the option that native speakers massively prefer. To attribute this preference to 'usage' is, of course, not very enlightening, but such is likely the major difference between a student who has thoroughly mastered an L_2 grammar but has been unexposed or under-exposed to everyday L_2 speech, and one who benefits from both metalinguistic training and immersion in the L_2 community.

We should also stress that true collocations do not actually contradict the rules. Rather, the most recalcitrant cases involve abstract nouns whose degree of referentiality is often under-determined for the purposes of the two rules proposed. It is not then surprising that the language would opt now for *du*, now for *de*, parallel to sister languages following different diachronic paths for a given lexical choice.

So, as linguists, we should concede that certain combinations reflect simply the happenstance of usage, and that no amount of semantic or syntactic theorizing has sufficed to explain fully such collocations. Of course, as with any science, the jury is still out on whether a new set of rules will eventually provide a full explanation.

Author's address:
Michael D. Kliffer
Department of French
McMaster University
Hamilton, Ontario
Canada, L8S 4M2

REFERENCES

Bolinger, D. (1976). Meaning and Memory. *Forum linguisticum*, 1: 1–14.
Byrne, L. and Churchill, E. (1986). *A Comprehensive French Grammar*. Oxford: Blackwell.
Bresnan, J., Dingare, S. and Manning, C. D. (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In: *Proceedings of the LFG '01 Conference*. Stanford: CSLI Publications, pp. 13–32.
Corblin, F. (1987). *Indéfini, défini et démonstratif. Constructions linguistiques de la référence*. Geneva: Droz.
Coseriu, E. (1978). Sistema, norma y habla. In *Teoría del lenguaje y lingüística general*. Madrid: Gredos, 11–113. (Originally published in 1952 by the Facultad de Humanidades y Ciencias, Montevideo.)
Croft, W. (2003). *Typology and Universals*, 2nd ed. Cambridge: Cambridge University Press.
Fillmore, C. J., Kay, P. and O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of Let Alone. *Language*, 64: 501–38.
Flaux, N. (1992). Les syntagmes nominaux du type 'le fils d'un paysan': Référence définie ou indéfinie? *Le Français moderne*, 60: 23–45.
Flaux, N. (1993). Les syntagmes nominaux: Référence définie ou indéfinie? (deuxième partie). *Le Français moderne*, 61: 113–40.

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

Givón, T. (1984). *Syntax. A Functional Typological Introduction*. Vol 1. Amsterdam: John Benjamins.
Gross, M. (1967). Sur une règle de cacophonie. *Langages*, 7: 105–19.
Guillaume, G. (1919/1975). *Le problème de l'article et sa solution dans la langue française*. Paris: Hachette, 1919. Republished by R.Valin, Paris: Nizet and Québec: Presses de l'Université Laval, 1975.
Hawkins, J. A. (1978). *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. London: Croom Helm.
Hooke, M. K. and Miller, M. H. (1962). *French Review Grammar*. New York: Holt, Rinehart and Winston.
Hopper P. J. and Thompson, S. A. (1984). The discourse basis for lexical categories in universal grammar. *Language*, 60: 703–52.
Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
Popescu-Belis, A. (2001). The use of articles to express referring cases: towards a mentalist explanation. Paper delivered at the International Workshop on Reference and Coherence in Discourse: Formal, Functional and Cognitive Approaches, Utrecht, The Netherlands. (available at http: //andreipb.free.fr/textes/apb-refcoh-01.html)
Sperber, D. and Wilson, D. (1989). *La pertinence. Communication et cognition*. Paris: Minuit.

APPENDIX 1: *SOIF DE/DU POUVOIR* SURVEY

Veuillez remplir les tirets soit avec **de**, soit avec **de** + **article** (*du, de l'*), soit avec les deux.

1. Et la lecture que nous avons entendue du Livre de la Sagesse, écrit durant le deuxième siècle avant J.C., au moment où les Juifs rêvaient encore d'une restauration de leur pouvoir politique, nous donne un exemple des effets moraux d'une telle soif _____ pouvoir . . .
2. Dirigés par la gloire et la soif _____ pouvoir, certains médias (car il y encore de bons journalistes) perdent de plus en plus de leur crédibilité.
3. Les conflits entre les hommes proviennent toujours des conflits qui sont dans le coeur des individus, et finalement de la soif _____ pouvoir qui nous est en quelque sorte innée.
4. Cependant, et c'est là sa force, Berthier analyse avec clairvoyance les deux camps. Il dépasse les considérations purement idéologiques pour montrer l'importance des facteurs économiques et la soif _____ pouvoir de certains.
5. Dans l'insouciance, on met le feu aux ressources du pays pour assouvir la soif _____ pouvoir personnel de Bédié.
6. En rompant avec la traditionnelle discrétion des vampires, Lestat s'attire la haine de Marius, son ancien maître, et réveille la soif _____ pouvoir de la reine Akasha, qui décide d'en faire son compagnon.
7. On aimerait quand même sentir chez M. Chrétien autre chose que la soif _____ pouvoir, autre chose que le souci de sa longévité politique, quelque chose qu'on puisse apparenter au respect de l'électorat.

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

8. [Simon de Montfort] Il devient le grand chef de la croisade, il en fait une affaire personnelle, bien plus dans la soif _____ pouvoir que par conviction philosophique ou religieuse.
9. Rabah s'est intéressé au destin présidentiel de François Mitterrand. Comment est-il parvenu à la fonction suprême après avoir parcouru un marathon politique peu commun? La soif _____ pouvoir a-t-elle suffi? Ou bien n'y a-t-il pas aussi eu adéquation entre les propositions de changements faites par François Mitterrand et la volonté des Français?
10. La mondialisation capitaliste, c'est cette marche chaotique de ce système fondé sur la concurrence, la soif _____ pouvoir, l'avidité à l'argent.
11. Avant que, dans un avenir lointain et plutôt incertain, notre 'siècle des idéologies' rende son âme souillée, avant que s'installe – si jamais elle s'installe – l'habitude d'écrire le mot 'idéologie' entre guillemets et que la pathologie _____ pouvoir et la soif _____ pouvoir soient appelées par leur nom, on devrait introduire dans les programmes scolaires et partout où cela est encore possible des cours quotidiens illustrant ce qu'a été et ce qu'est toujours idéologie moderne – sans guillemets.
12. D'autres cultivent l'insécurité comme une sorte d'épouvantail afin de détourner l'attention des citoyens du vrai problème de notre société qui est un mépris total de l'être humain au profit de la soif _____ pouvoir de quelques-uns, qu'il soit politique, économique ou religieux.
13. La soif _____ pouvoir, _____ domination et _____ conquête engendre des guerres meurtrières et, pour la première fois dans le conte, Micromégas ne contrôle pas sa colère.
14. Toutes les formes sociales démodées doivent être abandonnées et la soif _____ pouvoir doit se transformer en coopération constructive . . .
15. Animé par une seule motivation, la soif _____ pouvoir, Slobodan Milosevic a sciemment orchestré trois guerres qui ont occasionné des atrocités «quasi médiévales» et des «souffrances indicibles», pour lesquelles il n'éprouve aucun regret.
16. La soif _____ pouvoir l'emporte sur les grands principes qui peuvent toujours attendre leur heure de gloire.
17. La soif _____ pouvoir commande tout. La vie familiale d'abord, où Chirac asservit les siens à son ambition, et mobilise, jour et nuit, sa fille Claude, avec laquelle il entretient, on le sait, un rapport quasi exclusif.
18. Les charnières sont la peur, les frustrations, les tourments, les ambitions, la soif _____ pouvoir et la fascination pour l'argent.
19. Les dérives actuelles sont la conséquence d'une démocratisation mal négociée, où le détenteur du pouvoir veut s'arroger tous les droits, réduisant l'opposition à sa portion congrue. Pire, la soif _____ pouvoir pousse à ouvrir des boîtes de pandore d'où surgissent des démons incontrôlables comme l'ivoirité.
20. Comment ne pas reconnaître l'intemporalité de l'auteur Jean Racine, surtout après la semaine que nous venons de vivre à la suite des événements à New York. La soif _____ pouvoir, les meurtres, le bien, le mal n'ont pas d'époque.

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

RESULTS

	Orig.	#de	#du/de l'	Either
1.	de	8	1	
2.	du	4	4	1
3.	de	6	2	1
4.	de	7	2	
5.	du	6	3	
6.	de	9		
7.	du	4	5	
8.	du	3	4	2
9.	de	5	4	
10.	de	7	2	
11a.	du		9	
11b.	du	8	1	
12.	de	9		
13.	du	6	3	
14.	de	6	2	1
15.	de	2	7	
16.	du	5	3	1
17.	du	3	5	1
18.	de	6	2	1
19.	du	7	2	
20.	du	6	2	1

APPENDIX 2: DATA FOR *PRINCIPE DE/DU* AND *THÉORIE DE LA/DE*:

principe de la thermodynamique: 1690; de: 96
principe de la conservation: 708; de: 1240
principe de la mesure: 909; de: 1570
principe de la résonance: 85; de: 43
principe de la construction: 408; de: 1310
principe de la durabilité: 208; de: 298
principe de la dignité humaine: 135; de: 56
principe de la fidélité: 56; de: 85
principe de la transparence: 1 070; de: 1 980
principe de la confiance: 651; de: 368
principe de la territorialité: 625; de: 1230
principe de la reconnaissance: 974 (without *de/du/des* after *reconnaissance*); de: 1260
principe de la réversibilité: 62; de: 191
principe de la commutation: 111; de: 75
principe de la participation: 1 090; de: 972
principe de la séparation: 3 340; de: 2 370
principe de la séparation des biens: 6; de: 2
principe de l'offre: 337; de: 25

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press

principe de l'offre et de la demande: 176; de: 4 010
(N.B. with 'de', there is often a qualifying phrase after N_2.)

théorie des noeuds: 591; de: 7
théorie de la gestion: 466; de: 151
théorie de la diffusion: 580; de: 38
théorie de la femme: 9; de: 5
théorie du circuit: 124; de: 5;
théorie de l'information: 9 940; d': 129
théorie des systèmes: 6 630; de: 58
théorie de la statistique: 22; de: 18
théorie de l'engagement: 438; de: 7
théorie de la communication: 2 700; de: 340
théorie de l'architecture: 1 380; d': 184
théorie de l'évolution: 10 200; d': 216
théorie de la restauration: 55; de: 6
théorie du tissage: 11; de: 10
théorie de la sélection: 1 140; de: 31
théorie de la consommation: 288; de: 2
théorie de la sexualité: 970; de: 6
théorie des risques: 326; de: 2

https://doi.org/10.1017/S0959269504001395 Published online by Cambridge University Press