

Joëlle POPINEAU

Université de Tours, France
joelle.popineau@univ-tours.fr

CHARACTERIZING LEGAL STYLISTICS WITH A TXM TEXTOMETRIC TOOL

Abstract: Textometry tools can be used to characterize specialized varieties of English, such as Legal English. Our goal is to use the TXM textometry tool (Heiden et al., 2010) on Sales and Purchase Agreements (SPAs) to show agreements are drafted following a rigid structure, either in their form and shape (macrostructure) or their content (microstructure). We assume that the choice of words (Weisberg, 2014), and the construction of sentences or paragraphs [in agreements] shape how agreements are written and comprise the deal, that is, what is being sold. We hypothesize that there exists legal stylistics based on structural features. We built a corpus comprising Sales and Purchase Agreements of different sizes and topics to which the TXM tool was applied: parts of speech tagging help single out structure, linking words, word variety, frequencies, and writing routines. All descriptive details resulting from our implementation build legal stylistics that thus helps legal professionals, translators, teachers or clients better understand and draft agreements.

Key words: Textometry tool, Legal Stylistics, Sales and Purchase Agreements, legal English, POS tagging.

1. Introduction

This paper¹ proposes a linguistic study of Sales and Purchase Agreements (SPAs), which are highly specialized legal English documents and represent the outcome of key commercial and pricing negotiations. Our goal is to examine legal English wording and structure. We hypothesize that frequent parts of speech and words, syntactical frames, and structural routines characterize legal stylistics for SPAs. Our computer-aided demonstration was implemented on a textometric tool, called TXM (Heiden et al., 2010) which uses POS tagging (parts of speech

¹ Thank you to Elizabeth Sheppard Sellam for proofreading this paper.

tagging) on texts. We built a corpus of SPAs on which TXM extracted statistics on wording and structuring divisions. Our results give lexical and structural patterns (on 3 levels) for SPAs and will help students and professionals in academic or professional contexts better understand and draft SPAs.

2. Stylistics

The first part of our article² deals with theoretical frameworks and definitions in order to reach a definition for legal stylistics. Let's start by a first definition of stylistics (in general).

Stylistics is defined as “the linguistic study of style in language” (Jeffries & McIntyre, 2010: 61). Moreover, “the goal of most stylistics studies is not simply to describe the formal feature of texts for their own sake, but to show their functional significance for the interpretation of the text” (Wales, 2001: 373). Indeed, stylistics lies at the crossroads between linguistics and literature and “reçoit des deux disciplines et tente, en retour, d'enrichir les deux champs” [receives from both domains and, tries to enrich them both in return]³ (Sorlin, 2014: 12); stylistics is a purposeful communication tool in a specific context:

[la stylistique vise] à saisir la façon dont un discours (écrit ou oral) utilise les potentialités de la langue à des fins spécifiques dans un contexte particulier de production et de réception (Sorlin, 2014: 12) [stylistics aims to understand how a written or oral discourse uses all linguistic potentials for a specific purpose in a particular event where discourse is produced and received].

Furthermore, stylistics “aims to account for how texts project meaning, how readers construct meaning and why readers respond to texts in the way they do” (Jeffries & McIntyre, 2010: 61-64); so, stylistics has three goals: the first goal is to use words and sentences to express a message; the second goal is to help people interpret the meaning of the texts based on their knowledge, experiences, and cultural background; and the third goal is to make readers react to the text according to factors such as motivations and expectations. Stylistics does not only apply to literature, because “stylistics or general stylistics can be used as a cover term to cover the analyses of non-literary varieties of languages or registers” (Wales, 2001: 373): as this paper deals with agreements, stylistics applies to commercial and legal varieties of English. Therefore, speaking of stylistics for agreements makes sense.

² The author would like to thank the reviewers for their precious recommendations to improve the article.

³ All translations are made by the author.

3. Legal English

Legal English is a non-literary variety of specialized English related to the law used by legal professionals. This specialized variety of a language is defined according to its linguistic, discursive and cultural relationships with a targeted specialized domain:

Une langue de spécialité « émerge lorsqu'un domaine spécialisé tisse des relations avec la langue afin de la mettre au service de sa finalité. L'expression du spécialisé dans la langue apparaît selon des modalités très diverses [pouvant] être réparties en trois grands ensembles : linguistique, discursif et culturel. » (Van der Yeught, 2016: 2). [A specialized language “is born when a specialized domain establishes relationships with language in order to make it serve its purpose. The “specialized side” of the language appears in very diverse ways [which] can be divided into three main sets: linguistic, discursive and cultural.”]

Terms used in legal English are called *jurilects* and comprise lexical and discursive aspects of legal English:

un ensemble d'usages lexicaux et discursifs, propres au monde juridique ; les productions écrites et orales juridiques, englobant la terminologie savante juridique, les textes à fort degré de juridicité, mais aussi le vocabulaire traditionnel juridique et la terminologie juridique populaire viendront se ranger dans le jurilecte (Popineau, 2021: 429) [a set of lexical and discursive phrases and terms, particular to the legal world; written and oral legal works, including legal scholarly terminology, texts with a high degree of legal complexity, but also common legal vocabulary and everyday legal terminology can be included in the term jurilect]

Sales and Purchase Agreements are written in a specialized variety of language using *jurilects* and their linguistic, discursive, and cultural interaction with legal matters.

4. Legal English as a genre

Legal English is also defined as a genre that is “a category of artistic, musical, or treaty composition characterized by a particular style, form or content” (Merriam-Webster, n.d.). And Swales (1990: 58) gives the following definition:

A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community and thereby constitute the rationale for the genre.

In this respect, legal English comprises communicative events (e.g., buying a car) followed by communicative documents (e.g., a Sales and Purchase Agreement). The “expert members of the parent discourse community” are legal professionals, clients, buyers and sellers (companies or private persons), translators and translation

teachers, to name a few (Swales, 1990). In this respect, the schematic structure of each legal document is influenced and constrained by the purpose achieved, e.g., reaching an agreement on a price.

To sum up, when we combine and apply definitions (Sorlin 2014, Jeffries & McIntyre 2010) to legal stylistics, we can say that legal stylistics aims to understand how a legal written or oral discourse uses all linguistic potentials for a specific legal purpose in a legal particular event where legal discourse is produced and received. Moreover, legal stylistics has three goals: the first goal is to use words and sentences to express a legal message; the second goal is to help people interpret the meaning of the legal texts based on their knowledge, experiences, and cultural background; and the third goal is to make readers react to the legal text according to factors such as motivations and expectations.

As far as legal agreements are concerned, legal stylistics is a purposeful linguistic and semantic tool to better understand and interact with legally binding documents.

5. Sales and Purchase Agreements (SPAs) are legal acts

Let's continue our demonstration with our linguistic study based on a corpus of Sales and Purchase Agreements. Our next question will deal with the definition of a Sales and Purchase Agreement.

A Sales and Purchase agreement is “a decision or arrangement, often formal and written, between two or more groups of people” (Cambridge University Press, 2023a). A Sales and Purchase Agreement is a “binding legal contract that obligates a buyer to buy and seller to sell a product or service for a sum of money” (Chen, 2023). Once the agreement is made and entered (or agreed) by all parties, if a party fails to perform their commitment or obligations, there is a breach of contract, leading to penalties and damages. To be legally binding, oral or written agreements require “consideration” which is “a favour or advantage granted in return for something” (Oxford University Press, 2023a). The word “agreement” comes from the old French *agrément* (to please), or “mutual understanding” (Harper, 2001-2023a).

The first occurrences of written agreements date back to the early seventeenth century; they were basic agreements, made up of a single sentence (less than 40 words). Mandatory details were mentioned: the object of the sale, buyer's and seller's names, sum to be paid, delivery date and time of payment:

Sold to N.N. a quarter of Witte Kroonen for the sum of 525 gld. when the delivery takes place; and four cows at once, which may be now taken from the stable and led to the seller's house (Poitras, 2009: 488).

In this 1636 agreement (Poitras, 2009: 490), more details were incorporated to prevent a faulty payment or a breach of the agreement, as trade was increasing:

I, the undersigned, acknowledge to have bought from N.N., on conditions hereunder mentioned, one Gouda of 48 acres standing planted in N.N.'s garden, for the sum of 520 gld. in sterling. But in case 8 days after the notifying, the buyer were not to come to take the bulb, the seller may take it out of the ground, in the presence of two praiseworthy persons, and seal it in a box. And if a fortnight after this, the bulb has not been fetched by the buyer, the seller may sell it anew. If he gets more for it, the first buyer will not profit by it, and, when less, has to pay the difference. In case of any obscurity or misunderstanding or dispute arising out of this transaction, it will remain with two praiseworthy people, who know these things and who live in the place or town, where this transaction has taken place. And by default of payment of the aforesaid sum, I hereby engage all my goods, movable and immovable, submitting same in the power of all rights and magistrates; all this without arch 13 or cunning. Have signed this. Act in Haarlem on December 12th, 1636.

5.1. Legal doublets in SPAs

In our study, we have analysed Sales and Purchase Agreements. This name contains a legal doublet: sales and purchase. Doublets are a frequent feature in legal English⁴ and are described as *synonymic strings* (Mellinkoff 1963) or *merisms* (Garner 1995) as an analogy to a biology term;⁵ doublets are often considered as archaisms or figures of speech but they remain a particular feature of legal texts: “In law, a merism is a figure of speech by which a single thing is referred to by a conventional phrase that enumerates several parts or lists several synonyms for the same thing” (Standards, 2001-2023).

There are three types of legal doublets (Popineau, forthcoming):

etymology doublets (e.g., *goods and chattels*),
hypernym/hyponym doublets (e.g., *alter and change*),
meronymy doublets (e.g., *sales and purchase*).

The first class or etymology doublets such as *goods and chattels* refers to a string of words combining a Latin or Old French word and its Anglo-Saxon counterpart or translation: *goods* is the Anglo-Saxon equivalent for *chattels* from Old French *chatel* meaning *goods, wealth possessions, property, profit, cattle*. Etymology doublets were created by lawyers or judges in eyre whose job was to write and dispatch (and maybe translate) French-trying judgments to English-speaking people after England was invaded by Norman William the Conqueror in 1066; French was the language of England's courts and tribunals in the eleventh century.

⁴ More than 150 legal doublets have been analysed in our database.

⁵ “A merism is the repetition of similar parts, units, or segments within the structure of an organism” (Oxford University Press, 2023b).

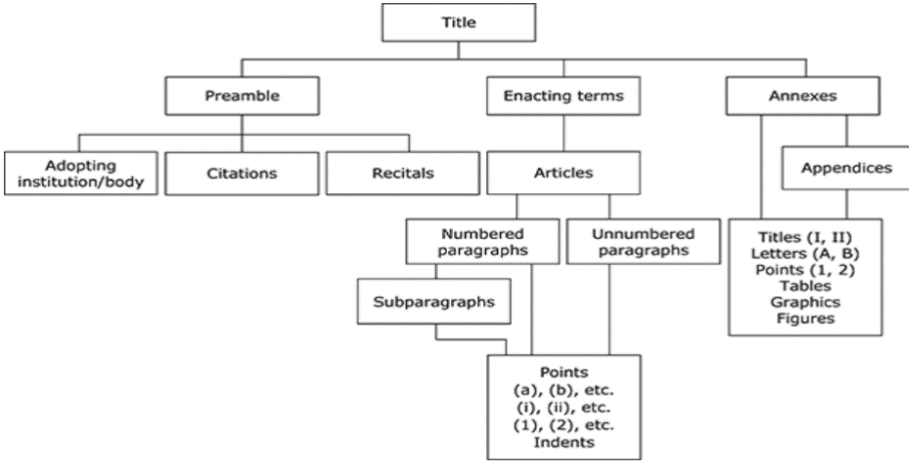
The second class is hypernym doublets which refer to a string of words with super-subordinate relations with each other: “alter and modify” means modify, and the hypernym modify includes alter.

The third class of meronym doublets expresses a part-whole relation: sales and purchase are the two necessary parts of a commercial transaction which is the holonym. Sales and purchase is thus a meronymy doublet and its mandatory name: contrary to synallagmatic *contrats français* in the French Napoleonic Code legal system, common law agreements are not reciprocal: the two parties, Seller and Buyer, need to be mentioned in full in the contract.

5.2. Structure of SPAs

Generally speaking, the purpose of a Sales and Purchase agreement is to outline relevant terms and conditions to be used for transactions (Chen, 2023). Modern-day agreements may contain hundreds of pages with dozens of supporting exhibits to prevent loopholes (breach or fraud) in transactions. Despite this, Sales and Purchase Agreements all share predictable structures detailed by legal advisors on both private and institutional websites; for example, a style guide was drafted by the European Union, with a diagram showing the basic elements of a legal act:

Figure 1: Predictable structure of a legal act (Publications Office of the European Union, 2021)⁶.



Their general frame has been standardized for many centuries: Table 1 shows their structures in the Middle-ages where parts and articles are similar to their modern-day counterparts:

⁶ The source is in public domain.

Table 1: Structure of a Post-Mediaeval agreement (“England Land Holders”, 2023).

THIS INDENTURE	The introduction
BETWEEN	Names of the parties. There can be any number of parties, and one party can consist of many individuals. Residences and occupations are usually given
WHEREAS	Recitals of previous transactions which place the present deed in context, and allow the researcher to identify former deeds.
NOW THIS INDENTURE WITNESSETH	This is the <i>testatum</i> or terms of the present contract
TO HAVE AND TO HOLD	This is the <i>habendum</i> which gives the length of the contract and any conditions and restrictions.
IN WITNESS THEREOF	This is the <i>testimonium</i> where the representatives of the parties sign and impress their seals.

6. Computational study on SPAs with TXM

We have used a textometric platform, called TXM, for our research on language patterns in Sales and Purchase agreements. TXM was developed in France in the 1980s and provides “powerful techniques for the analysis of a large body of texts. Following textometry and text statistical analysis, it offers tools and methods tested in multiple branches of the humanities” (Decorde et al., 2023). We applied TXM to our corpus made of six recent Sales and Purchase Agreements with different topics and sizes.

Finding authentic Sales and Purchase agreements may sometimes be complex because they contain details that are protected by copyrights. We have selected agreements and built a corpus: some agreements have been found on the U.S. Securities and exchange commission (SEC) website, an independent agency of the U.S. Federal government with a “mission of protecting investors, maintaining fair, orderly, and efficient markets, and facilitating capital formation”; we have anonymized some agreements for privacy reasons (and marked them with an asterisk *):

Table 2: List of Sales and Purchase Agreements we selected in our corpus.

Name of the sales agreement	Description of the agreement	sales on
[Car 2020]*	vehicle sales agreement	a car
[Cement 2009]	cement sales agreement	cement in bags
[Collection 2020]*	agreement between an institution purchasing a collection and a private seller	a collection of masterpieces or artifacts
[Newmark 2012]	A 3-party share purchase agreement between Newmark Investment Limited, Carbon Strategic Ltd and Oak Ridged Micro Energy Inc.	shares in a company
[Real Estate 2020]*	real estate agreement	a flat
[SoftDigit 2010]	software sales agreement	software products

Each agreement has been converted into text format and imported into TXM. We launched a first command and obtained descriptive statistics [propriétés] for all agreements (Fig. 2): an overall view (or statistiques générales) shows the total number of lexical units (the agreement [Car 2020] has 1,874 words), of tagging for each word (word, lemma (or canonical form), parts of speech, and line number) and the number of structure units of the text (2, sentence and text); *en* stands for English language. A tagged example of the first sentence of the agreement is shown under “propriétés des unités lexicales”; “propriétés des structures” gives the name of the input file (PRESSE-PAPIER6 or clipboard#6) and the number of lines in the text (84 lines).

Figure 2: Descriptive details for a sales and purchase agreement given by TXM for [Car 2020].

Propriétés de car2020 (CQP ID=PRESSE-PAPIER6)

Statistiques Générales

- Nombre de mots : 1 874
- Nombre de propriétés de mot : 4 (word, enlemma, enpos, lbn)
- Nombre d'unités de structure : 2 (s, text)

Propriétés des unités lexicales (max 10 valeurs)

- enlemma : this, sale, agreement, (, the, ",), date,
- enpos : DT, NNS, NN, (, `",), VVD,
- lbn : 4,
- word : THIS, SALES, AGREEMENT, (, the, ", Agreement,), dated,

Propriétés des structures (max 10 valeurs)

- text
 - id (1) = PRESSE-PAPIER6.
- s
 - n (84) = 44, 45, 46, 47, 48, 49, 50, 51, 52, 53.

We have conducted our study on three stylistic levels macro, mezzo, and micro, and each level is described with its stylistic features below.

6.1. Macrostylistic level of SPAs

The first level, or macro-stylistics level, deals with the outer structure of the agreements. TXM analysis gives quantitative details:

Table 3: Number of articles and sections in our corpus of Sales and Purchase Agreements

Name	description	date	Number of words	Number of sentences	Number of articles	Number of sections
[Car 2020]	vehicle sales agreement	2020	1,874	84	16	down to 13 sections
[Cement 2009]	cement sales agreement	2009	3,580	193	27	down to 7 sections
[Collection 2020]	agreement between an institution purchasing a collection and a private seller	2020	1,115	60	8	down to 8
[Newmark 2012]	a 3-party share purchase agreement between Newmark Investment Limited, Carbon Strategic Ltd and Oak Ridged Micro Energy Inc.	2012	13,913	456	10	down to 27 and subsections
[Real Estate 2020]	real estate agreement	2020	13,772	482	32	down to 20
[SoftDigit 2010]	software sales agreement	2010	3,516	170	23	down to 5

The total number of words, articles, and sections is given. All articles have many sections, proving all cases and mishaps need to be contemplated to prevent loopholes and money loss or remedies for the seller and penalties for the buyer.

TXM gives further quantitative details on the length of sentences in Table 4. Sentences are between 18 and 30 words long, which corresponds to a highly specialized text. A common plain English guideline says there should be an average of 15–20 words in a sentence (Cutts, 2009; Plain English Campaign, 2015⁷):

Table 4: Number of words in a sentence in our corpus

Name	Number of words	Number of sentences	Number of words per sentence
[Car 2020]	1,874	84	22
[Cement 2009]	3,580	193	18.5
[Collection 2020]	1,115	60	18.5
[Newmark 2012]	13,913	456	30.5
[Real Estate 2020]	13,772	482	28.5
[SoftDigit 2010]	3,516	170	20.7

Agreements are all structurally articulated similarly:

- an introduction sentence (and a title) followed by,
- recitals;

⁷ Plain English campaign is an independent British group fighting for plain English in public communication and releasing reports among other things.

- definitions;
- the body text of the agreement describes the transactions (up to 40 articles);
- the signature page ends the agreement;
- some exhibits can be attached.

The name of each subdivision is sometimes in capital letters.

6.2. Mezzostylistic level of SPAs

Let’s now explore the mezzostylistic level to find how articles are semantically articulated. The introduction sentence is syntactically correct and semantically important: it says the agreement was signed by the two parties (full name and address head office) on a specific date. Legal wording is used: the agreement was “made and entered” (a doublet), by and between (another doublet) the parties (referred to as abbreviated names between brackets with initial capital letters) on an “effective date” [Collection 2020]:

AGREEMENT OF PURCHASE

This agreement is made and entered by and between (“Institution”), and (“Seller”) on (date).

RECITALS⁸ contain the background and motivations of an agreement, that is “the main details about a contract including whom it involves and why they are making the contract” and start with “WHEREAS” in capital letters:

WHEREAS, the Seller desires to sell to the Institution a comprehensive collection of (“Collection”), which is more particularly described in the attached inventory, Attachment A, which is incorporated herein by reference [Collection 2020]

DEFINITIONS in an agreement only refer to restricted or special meanings: in [SoftDigit 2010], “market” means a geographical area jointly agreed on by Buyer and Seller:

DEFINITIONS. In this Agreement, the terms listed below have the following meanings:

“Products” mean those retail computer software products developed by SoftDigit listed in the price list attached as Exhibit A.

“Market” means the geographical area and product market set forth in Exhibit B. If the parties agree to include additional geographical areas or product markets in the Market during the term of this Agreement, they shall be added to Exhibit B.

“Price” means the price to Distributor for the software products set forth in Exhibit A. SoftDigit may, in its sole discretion, increase or decrease prices for the products upon giving the notice required by section 6E.

⁸ <https://dictionary.cambridge.org/dictionary/english/recital?q=recital+> (Accessed on January 21st, 2023)

As a teacher in legal translation,⁹ we designed a SPA template for didactic purposes: there are three semantically articulated parts in agreements, marked by legal link words (in capital letters in agreements): WHEREAS, NOW THEREFORE, IN WITNESS WHEREOF:

Figure 3: Template of a Sales and Purchase Agreement we created for our legal English lessons

(Title) – Sales and Purchase Agreement (SPA)
This/the agreement is made and entered on (date) day of 2022 by and Between (Name) (hereinafter referred to as “Seller”), On the one part And (Name) (hereinafter referred to as “Buyer”), On the other part
[Recitals] WHEREAS, WHEREAS, WHEREAS
NOW, THEREFORE, in consideration of the promises hereinafter made by the parties hereto, it is agreed as follows:
Seller shall sell and Buyer shall buy, free and clear of all liens, encumbrances, and liabilities, those assets of Seller’s business, commonly known as _____ consisting of _____ and equipment, all of which are more fully described and enumerated in Schedule A which is attached and by this reference made a part hereof. [...]
Articles (body of the agreement): what is sold; what price; Delivery time and terms (and Incoterms); Duties of the purchaser; Duties of the seller; Warranties; Infringements; Trademark and protection; termination; force majeure; governing laws; miscellaneous.
IN WITNESS WHEREOF, the parties hereto have set their hands and seals, the date and place first above written.
Signature and capacities Exhibit

WHEREAS, NOW THEREFORE, IN WITNESS WHEREOF are groups of legal phrases or *jurilects*; we assume they play the role of markers and help the reading of articles in three steps:

motivations to the agreement (WHEREAS),
its description (NOW THEREFORE, or what, how much, who, when, why, where) and
its execution (IN WITNESS WHEREOF).

These legal markers are often referred to as archaisms:

Whereas is used to introduce contractual recitals and the like, but modern drafters increasingly prefer a simple heading, such as “Recitals” or “Preamble”, and in that way avoid the legalistic *whereas* (Garner 2009: 1733).

⁹ The author has been teaching legal translation for 20 years.

Five out of six agreements in our corpus use these markers, though partially, in our corpus.

Table 5: WHEREAS, NOW THEREFORE, IN WITNESS WHEREOF in our corpus

SPAs	WHEREAS	NOW THEREFORE	IN WITNESS WHEREOF
[Car 2020]	no	no	yes
[Cement 2009]	yes	no	yes
[Collection 2020]	yes	yes	no
[Newmark 2012]	no	yes	no
[Real Estate 2020]	no	no	no
[SoftDigit 2010]	no	no	yes

6.3. Microstylistic level of SPAs

The third level, or microstylistic level, focuses on parts of speech (POS) which is the most used criterion in computer-based analyses. Grammar traditionally gives eight parts of speech (POS) in English (verb, noun, pronoun, adjective, adverb, preposition, conjunction, and interjection). There are 58 POS tags under TXM (Table 6)¹⁰:

Table 6: 58 POS Tags used in TXM

Parts of speech		Examples
CC	coordinating conjunction	and, but, or, &
CD	cardinal number	1, three
DT	determiner	the
EX	existential <i>there</i>	there is
FW	foreign word	hors d"oeuvre
IN	preposition or subordination conjunction	in, of, like, after whether
+IN/that	complementizer	that
JJ	adjective	green
JJR	adjective comparative	greener
JJS	adjective superlative	greenest
LS	list item marker	(1)
MD	modal	could, will
NN	noun, singular or mass	table
NNPS	proper noun, plural	Vikings
NNS	noun, plural	tables
NP	proper noun singular	john
PDT	prodeterminer	both
POS	possessive ending	friend's

¹⁰ DT (determiner), SENT (end punctuation) and : (general joiner) are excluded from our study.

PRP PP	personal pronoun	I he it
PRP\$ PP\$	possessive pronoun	my his
RB	adverb	however, usually,
RBR	adverb, comparative	better
RBS	adverb, superlative	best
RP	particle	give up
SENT	end punctuation	? ! .
SYM	symbol	
TO	to	to go, to him
UH	interjection	uhhuhhuhh
VB	verb be infinitive	be
VBP	verb be plural	are
VBZ	verb, be 3 rd person singular present	is
VV	verb, base form	take
VVD	verb, past tense	did
VVG	verb, gerund or present participle	doing
VVN	verb, past participle	taken
VVZ	verb, 3rd person singular present	takes
WDT	wh- determiner	which
WP	wh- pronoun	who, what
WP\$	possessive wh-pronoun	whose
WRB	wh-adverb	where, when
\$	currency symbol	\$£
:	general joiner	;,...

POS tagging has developed accordingly to the analyses led and the congregated Verb class includes all forms a verb can have in a text:

Table 7: 58 POS Tags for verbs

V classes	Details	Examples
VB	verb be infinitive	be
VBP	verb be plural	are
VBZ	verb, be 3 rd person singular present	is
VV	verb, the base form	take
VVD	verb, past tense	did
VVG	verb, gerund or present participle	doing
VVN	verb, past participle	taken
VVZ	verb, 3rd person singular present	takes

Our TXM analysis shows that lexical variety is low in agreements. NN (noun singular or mass) and NP (proper noun singular) have the highest frequencies; adjectives (JJ) are between twice and three times less frequent than nouns:

Table 8: Frequencies for some POS

Name	Number of ITEMS	NN	IN	NP	JJ	CC	MD
[Car 2020]	1,874	292	252	156	97	93	43
[Cement 2009]	3,580	723	385	129	201	167	76
[Collection 2020]	1,115	199	148	77	79	72	25
[Newmark 2012]	13,913	1,788	1,356	1,583	964	892	173
[Real Estate 2020]	13,772	2,337	1,349	1,063	768	613	224
[SoftDigit 2010]	3,516	669	444	191	192	194	80

We took a closer look at Table 7 and Table 8. Congregating distinctive Verbs forms and tenses under a unique Verb label helps to confirm further features. Table 9 shows congregated classes of nouns (N), verbs (V), wh-determiners (W), adjectives (JJ), coordinating conjunction (CC), and modals (MD):

Table 9: Frequencies for congregated POS classes.

Name	Number of ITEMS	N	V	W	JJ	CC	MD
[Car 2020]	1,874	525	195	28	100	93	43
[Cement 2009]	3,580	1,129	281	14	207	167	76
[Collection 2020]	1,115	243	152	11	80	72	25
[Newmark 2012]	13,913	4,185	1,232	86	988	892	173
[Real Estate 2020]	13,772	4,147	1,569	91	785	613	224
[SoftDigit 2010]	3,516	1,122	349	14	198	194	80

The agreements we studied show a high percentage of Nouns (with an average of 30% of all items in all agreements) and almost 10% of Verbs: nouns are three times higher than verbs.

Verbs have a low frequency (1,232 forms) whatever their forms (past participle (VVN), VBZ (verb, be 3rd person singular present), VV (verb, base form), VVG (gerund) in [Newmark 2012]:

Table 10: Congregated POS for V [Newmark 2012]

Class	Frequency	Class	Frequency
VB	112	VHN	2
VBD	9	VHP	27
VBG	7	VHZ	44
VCN	40	VV	195
VBP	44	VVD	55
VBZ	68	VVG	131
VH	28	VVN	393
VHD	3	VVP	22
VHG	2	VVZ	50

This feature (30% of N and 10% of V) is a key feature of agreements and legal English in particular; in general English, verbs (or congregated class of verbs) are

more frequently used than nouns (Biber & al, 1999). With 30% of N and 10% of V, agreements belong to academic writing.

Shall is the only modal (MD) and is frequent in all agreements studied: Figure 4 shows concordances for [Cement 2009]; *shall* means a strong binding obligation in a contract.

Figure 4: Concordances for *shall* [Cement 2009]

Contexte gauche	Pivot	Contexte droit
authority, certifies represents and warrants that each therefore agreed as follows: 1. PRODUCT	can	fulfill the requirements of this agreement and respectively provide the pr...
/ PRODUCT DISCHARGE: 3. 1 Destination	SHA...	CONFORM TO INTERNATIONAL STANDARD ACCORDING TO, BRITISH ST...
The basis of delivery for all other shipments	shall	be CIF.. _ (non USA sanctioned port) (
metric ton shipments. 3. 3 Delivery	shall	be minimum twelve thousand five hundred (12, 500) metric
) of bill of lading (CIF)	shall	begin thirty to forty (30 -40) days after receipt and
delivery. 4. 2 The first delivery	shall	be considered the date (s) of delivery. 4.
the seller. 4. 3 The seller	shall	begin thirty to forty (30 - 40) days after receipt
the buyer. 4. 4 The product	shall	have the right to deliver earlier than agreed in this contract giving
-5 %) 4. 5 Total shipments	shall	be shipped to the relevant provisions of this contract with the shipment
the customary and usual exigencies. Consecutive shipm...	shall	be according to shipping schedule. (See " Appendix 2 "
CONTRACTED QUANTITY The total quantity of the cont...	shall	be shipped by delivering and receiving schedule of buyer, sent to
that each shipment of cement 42. 5	shall	be determined by the certifications of weight issued by the inspection au...
quality at time of loading and such certificate	shall	be provided with an inspection certificate of weight and quality at time
similar recognized authority at seller's expense, and	shall	be provided by Société or similar recognized authority at seller's expense,
be final. 6. 2 The seller	shall	be deemed to be final. 6. 2 The seller shall
) rules. 6. 3 The buyer	shall	instruct said authority to carry out the inspection in strict accordance with
	shall	, if desired, and at his own expense provide additional inspection

Linking words show cross-referencing: *hereto*, *hereby*, *hereunder*, *herewith*, *hereinafter* and *this agreement* (30 occurrences in [SoftDigit 2010]) are frequent; such cross-referencing connectors show the consistency and reflexivity of agreements.

Further analyses carried out on lexical categories such as those given by Sorlin (2014: 54) show that:

words are simple words (short words with few syllabi; no derived nouns (no suffixes) and no compound nouns);

nouns are concrete words, semantically linked to products, damage, or risks;

adjectives, mainly epithets, are not very frequent; they mostly refer to the agreement, the payment, or the commitment of each party;

verbs are mostly transitive verbs explaining the commitment of each party, the product to be sold, and the payment to be done;

shall is the only modal used in agreements;

abbreviations in agreements only refer to Incoterms (INternational COMMERCIAL TERMS which make up a list of 11 words)

In addition, there are many words in legal English with Latin root, due to the history of the legal English language: « Le domaine de la justice (comme l'administration et l'armée) est un des domaines qui gardent l'empreinte de cette transformation lexicale (propagation du normand) d'où l'origine française de nombreux mots dans le texte ayant une fonction juridique » [Law (as well as

administration and military) is one of the domains where French origin words have strongly influenced the vocabulary (Norman language expanded in Law], hence the French origin of many words in texts having a legal function] (Sorlin, 2014: 76). As an example:

*purchase*¹¹ is from Anglo-French *purchase*, Old French *porchaz* “acquisition, gain, profit; seizing, plunder; search pursuit, effort”, from Anglo-French *purchaser*, Old French *porchacier*.

7. Some results of TXM approach to SPAs

The study we have conducted shows that agreements are structured rigidly on the three levels we have given. Here are some results.

Firstly, on the macro-stylistic level, our study shows a regular and highly structured layout in agreements, which are long documents (up to 40 pages and almost 14,000 words in our corpus). Some markers (title or legal phrases) in capital letters help distinguish parts: each agreement has an orderly outer layout with up to 10 sections containing mandatory details: an introductory line with the name of the parties and the effective date, definitions of each term used in the agreement, description of the goods to be sold, purchase price and payment, delivery of goods and shipping and insurance, litigation and remedies, cancellation, *force majeure*, governing law, and the signature page). A missing section may lead to either breach of agreement or money loss. Moreover, sentences are between 18 and 30 words long.

Secondly, the mezzo-stylistic level shows how the general layout and semantics interconnect in articles of all agreements. Legal archaisms (WHEREAS, IN WITNESS WHEREOF) play the role of markers in agreements, explicating semantic links on how the agreement unfolds. Other legal linking words show semantic connections in agreements (hereby, hereto, herewith).

Thirdly, the micro-stylistic level shows a low lexical variety in agreements. A full list of keywords can easily be built, as nouns (NN, NNPS, NNS, NP) are frequent (30% of all occurrences): this can be explained either by the pleonastic feature of legal vocabulary (legal doublets) or a missing -confusing- synonymy in legal words. Verbs represent 10% of all occurrences, among which 2% are modals (MD) and 8% are verb base forms (VV). Then, there are less than 10% of adjectives (JJ) and 5% of adverbs (RB) in agreements. Unsurprisingly, key sales nouns (*Buyer, Seller, Goods, Agreement*) are high in nouns (almost 9%). And last, but not least, sentences are long, but there are few subordination conjunctions (1%), which may lead to misunderstanding; personal pronouns or anaphors are not represented as they are legally and semantically misleading as well.

¹¹ <https://www.etymonline.com/search?q=purchase>. (Accessed on March, 3rd, 2023)

The initial hypothesis of our study was to show that the legal stylistics of agreements could be characterized with the use of the textometric tool called TXM. The TXM tool was implemented on a corpus of agreements built from real and anonymized agreements. We conducted a structural, syntactical, and lexical study (on 3 levels) with TXM for each agreement. Our results give a list of lexical, structural, and semantic characteristics of legal stylistics which proves to be a purposeful tool for drafting and understanding agreements in academic or professional contexts. The list can be seen as the general outline and layout for any agreement, each division being then completed with adequate details. Thus, the outer visible complexity of Sales and Purchase Agreements appears less and less complex, as Garner says, “Contract drafting is a specialized form of expository prose – although it is not as specialized as many would make it” (2019: introduction).

The TXM analysis we carried out proposes an innovative approach to analysing and implementing research in legal English and corpus linguistics. It helps to teach Legal English in ESP classes as well; such an approach can be further refined and developed with bigger corpora of agreements or other legal texts.

Bibliography

- Biber, D., Johansson S., Leech G., Conrad S. & Finegan E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
- Cambridge University Press. (2023a). Recital. In *Cambridge Dictionary*. Retrieved January 21, 2023, from <https://dictionary.cambridge.org/dictionary/english/recital?q=recital+>
- Cambridge University Press. (2023b). Agreement. In *Cambridge Dictionary*. Retrieved January 3, 2023, from <https://dictionary.cambridge.org/fr/dictionnaire/anglais/agreement>
- Chen, J. (2023). Sales and Purchase Agreement (SPA): What It Is, With Examples. In *Investopedia*. Retrieved January 3, 2023, from <https://www.investopedia.com/terms/s/salesandpurchase.asp>
- Cutts, M. (2009). *English language*. Oxford: Oxford University Press.
- Decorde, M., Heiden, S., Jacquot, S., Lavrentiev, A. & Pincemin, B. (2023). Textometry. Retrieved January 3, 2023, from <https://txm.gitpages.huma-num.fr/textometrie/en/>
- England Land Holders, Title Deeds, Leases - International Institute. (2023). In *FamilySearch*. Retrieved January 3, 2023, from https://www.familysearch.org/en/wiki/index.php?title=England_Land_Holders,_Title_Deeds,_Leases_-_International_Institute&oldid=5043832
- Garner B. A. (1995). *A Dictionary of Modern Legal Usage*. Oxford: Oxford University Press.
- Garner B. A. (2009). *Black's Law Dictionary*. Ninth edition. St Paul: West Publishing Co.
- Garner B. A. (2019). *Guidelines for Drafting and Editing Contracts*. West Academic Publishing.
- Harper, D. (2001-2023a). Agreement. In *Online Etymology Dictionary*. Retrieved January 3, 2023, from <https://www.etymonline.com/search?q=agreement>
- Harper, D. (2001-2023b). Purchase. In *Online Etymology Dictionary*. Retrieved March 3, 2023, from <https://www.etymonline.com/search?q=purchase>
- Heiden, S., Mague J. & Pincemin B. (2010). *TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement* [Conference paper]. JADT 2010: 10th International Conference on the Statistical Analysis of Textual Data, Rome. http://halshs.archives-ouvertes.fr/docs/00/54/97/79/PDF/Heiden_al_jadt2010.pdf

- Jeffries, L. & McIntyre D. (2010). *Stylistics*. Cambridge: Cambridge University Press.
- Mellinkoff, D. (1963). *The Language of the Law*, Boston: Little, Brown & Co.
- Merriam-Webster. (n.d.). Genre. In *Merriam-Webster.com dictionary*. Retrieved January 3, 2023, from <https://www.merriam-webster.com/dictionary/genre>
- Oxford University Press. (2023a). Consideration. In *Oxford Learner's Dictionaries*. Retrieved January 17, 2023, from <https://www.oxfordlearnersdictionaries.com>
- Oxford University Press. (2023b). Merism. In *Oxford Learner's Dictionaries*. Retrieved January 17, 2023, from <https://www.oxfordlearnersdictionaries.com>
- Poitras, G. (2009). The early history of options contracts. In W. Hafner & H. Zimmermann (Eds.), *Vincenz Bronzin's Option Pricing Models: Exposition and Appraisal*. Berlin & Heidelberg (pp. 487-518), Springer-Verlag.
- Popineau, J. (2021). Traduire des technoclectes juridiques ou jurilectes dans la presse française et anglaise: approche juritraductologie et terminologique, *Meta* 66(2), 427-450. <https://doi.org/10.7202/1083186ar>
- Popineau, J. (forthcoming). Imprudence et prudence en traduction juridique : le cas des doublets juridiques ou *legal pairs*. In N. Khalfallah & H. Moukannas (Eds.), *Imprudence et prudence en traduction juridique*. Paris: Classiques Garnier.
- Publications Office of the European Union. (2021). Structure of a legal act [Diagram]. In *Europa: Interinstitutional Style Guide*. Retrieved January 10, 2023, from <https://publications.europa.eu/code/en/en-120000.htm>
- Sorlin, S. (2014). *La stylistique anglaise. Théories et pratiques*. Presses Universitaires de Rennes.
- Stand4. (2001-2023). *Definitions.net*. Retrieved January, 5, 2023, from <https://www.definitions.net>
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Van der Yeught, M. (2016). Protocole de description des langues de spécialité, *Recherche et pratiques pédagogiques en langues de spécialité*, 35(1). <https://doi.org/10.4000/apliut.5549>
- Wales, K. (2001). *A Dictionary of Stylistics*. London and New York: Longman.
- Weisberg, R. (2014). Rue d'Ulm: Comparative Judicial Stylistics. *Clio@Themis*, 7. <http://journals.openedition.org/cliothemis/1649>.