

**Capitalizing on genre-based corpora with  
the use of the AI-powered research tool Notebook LM**

*(Aprovechamiento de corpus basados  
en géneros con el uso de la herramienta de  
investigación Notebook LM, impulsada por IA)*

BELÉN LABRADOR

<https://orcid.org/0000-0002-3341-1661>

[belen.labrador@unileon.es](mailto:belen.labrador@unileon.es)

*University of León* <https://ror.org/02tzt0b78>

Fecha de recepción: 17 de julio de 2025

Fecha de aceptación: 11 de diciembre de 2025

**Abstract:** Syntagmatic relations have always been at the core of corpus linguistics, where the most important element to study is the co-text, mainly collocates and patterns identified in concordance lines. Without abandoning this approach, the advent of artificial intelligence (AI) has opened new possibilities for exploiting corpora beyond traditional browser and concordancer functionalities, as AI-powered research tools offer a deeper, more holistic understanding of corpora. This paper explores the potential of Notebook LM, a generative AI (GenAI) application, to complement data retrieved by traditional corpus-analysis toolkits. Our initial hypothesis is that combining corpus and GenAI approaches leads to a more comprehensive understanding of specific genre-based texts and the language used within them. A key advantage of Notebook LM over large language models (LLMs) like ChatGPT is that it processes only uploaded texts, giving researchers precise control over the sources used. As a case study, a comparable English-Spanish corpus of online cheese descriptions was used. It contains 400 Spanish texts (121,461 words) and 600 English texts (111,871 words). This corpus was fed into Notebook LM, and its functions were used to summarize, explain, and retrieve key themes such as historical significance and production methods. The power of prompting within Notebook LM is also demonstrated, suggesting that GenAI tools are more suitable for qualitative than quantitative analysis in this context. Comparing writing conventions in both languages revealed distinct differences: English descriptions tend to be more concise and informative, while Spanish texts use more subjective, evocative language, often incorporating cultural context and a

more enthusiastic tone. These findings can assist in second-language writing and translation training by pinpointing the appropriate stylistic characteristics of promotional texts in English and Spanish. In conclusion, Notebook LM provides complementary affordances which, when combined with other corpus-analysis applications, constitute a powerful tool for leveraging genre-based corpora.

**Keywords:** Notebook LM. Generative AI applications. Genre-based corpus. Comparable corpus. Prompts.

**Resumen:** Las relaciones sintagmáticas siempre han sido el eje central de la lingüística de corpus, donde el elemento más importante es el cotexto (principalmente patrones colocacionales). Sin abandonar este enfoque, el advenimiento de la inteligencia artificial (IA) ha abierto nuevas posibilidades para explorar los corpus más allá de las funcionalidades de los navegadores y programas de concordancias, ya que las herramientas de investigación impulsadas por IA ofrecen una comprensión más profunda y holística de los corpus. Este artículo examina el potencial de Notebook LM, una aplicación de IA generativa. La hipótesis inicial es que la combinación de corpus e IA conduce a una comprensión más exhaustiva de géneros textuales y del lenguaje utilizado en ellos. Una ventaja de Notebook LM sobre grandes modelos de lenguaje como ChatGPT es que procesa solo los textos cargados, otorgando a los investigadores control preciso sobre las fuentes utilizadas. Como estudio de caso, se utilizó un corpus comparable de descripciones de quesos: 400 textos en español (121 461 palabras) y 600 en inglés (111 871 palabras). Se utilizó Notebook LM para resumir el corpus, obtener temas clave, como la importancia histórica y los métodos de producción y demostrar el poder de las instrucciones (*prompts*), más adecuadas para el análisis cualitativo. Al comparar las convenciones de escritura, las descripciones en inglés tienden a ser más concisas e informativas, mientras que los textos españoles utilizan un lenguaje más subjetivo y evocador, a menudo incorporando contexto cultural y un tono más entusiasta. Estos hallazgos pueden ser útiles en escritura en segundas lenguas y en traducción, al distinguir las características estilísticas de textos promocionales que resultan apropiadas en ambas lenguas. En conclusión, Notebook LM proporciona posibilidades complementarias que, combinadas con otras aplicaciones de análisis de corpus, constituyen una poderosa herramienta para el aprovechamiento de los corpus basados en géneros.

**Palabras clave:** Notebook LM. Aplicaciones de IA generativa. Corpus basados en géneros. Corpus comparables. *Prompts*.

## 1. Introduction

The advent of artificial intelligence (AI) and generative AI (GenAI) applications represents a milestone in human history in general, and in the progress of all fields of knowledge in particular. As far as linguistics is concerned, we can compare the arrival of AI with the arrival of corpora because both constitute breakthroughs in this area. The beginning of corpora was revolutionary at a time when the so-called “armchair linguists” (Fillmore 1992) used to work with examples that they could come up with to support and illustrate their theories. Corpora allowed linguists to get hold of thousands of empirical data to validate or reject their hypotheses. Therefore, there was a major change from deductive reasoning to an inductive methodology. In corpus linguistics, syntagmatic relationships are the center of attention (Sinclair 1991; Hoey 2005). The most important element to study is the co-text: collocations, patterns, clusters, and multiword expressions, among others, identified in concordance lines, and the way to retrieve the data is by means of corpus browsers and concordancers. Corpus linguistics constituted a huge leap in the scope, focus, and methodologies used in linguistics (Svartvik 1992; Biber *et al.* 1998; Crosthwaite 2024) and in contrastive linguistics (Aijmer *et al.* 1996; Johansson and Oksefjell 1998). Likewise, nowadays, the advent of AI has led to another significant shift.

Without abandoning the general principles and approach of corpus linguistics, the appearance of AI has opened new possibilities to exploit corpora in different ways beyond the features of the corpus browsers or concordancers. AI-powered research tools have broadened perspectives by providing a deeper understanding of corpora that includes a more holistic scope. This paper aims to explore the potential of a GenAI application, Notebook LM (Language Model) <https://notebooklm.google/> - described in more detail in Section 2.2. - in providing information that complements the data retrieved by traditional corpus-analysis toolkits. Within the realm of contrastive linguistics and comparable genre-based corpora, the research questions that initially prompted this study were: (1) How can we research differently with AI from now on in contrastive linguistics? (2) How can we benefit from the tools that we have at our disposal nowadays to explore genre-based corpora? Those “armchair linguists” had to face challenges at their time with the arrival of corpora, and they had to adapt to new methods and tools. Similarly, now we need to address the problems that AI causes, but we can also benefit from the opportunities it generates. This paper constitutes a tentative attempt to examine this new world of AI by exploring Notebook LM to find out the possibilities it can offer in linguistic research, in line with other studies that have analysed the usefulness of other AI tools for

linguists (Torrent *et al.* 2024; Curry *et al.* 2024; Oksefjell Ebeling 2025; Sanz-Valdivieso and López-Arroyo 2025).

Notebook LM has been mainly used so far in education to generate podcasts with a feature called audio overview (Alonso-Guisande and López-Fraile 2024; Mehta *et al.* 2024). This interesting feature enables users to transform their uploaded source materials into engaging, podcast-like audio summaries; consequently, students can understand content better, as they are given the chance to not only read papers, documents, or textbooks, but also listen to those materials in a more informal style. It constitutes a technique to make those materials more appealing and to facilitate comprehension. This affordance ties in with the three main guidelines of Universal Design for Learning (UDL): multiple means of engagement, representation, and action and expression. UDL claims that it is necessary to offer multiple pathways to achieve educational goals (Gadsden and Goegan 2023; Moriña *et al.* 2025). This variety in the way learning targets are presented and explained also contributes to catering to the needs of different types of students, with different preferred learning styles: visual, auditory, and kinesthetic. By listening to the podcasts that are based on the written texts uploaded to Notebook LM, students can complement their understanding of the texts, as the channel varies (from written to spoken), but also the style is less formal and academic, which makes it more accessible. These types of tools that provide multiple means of representation of knowledge foster inclusive methodologies in education.

In this study, some other features of Notebook LM have been tried to support the initial hypothesis claiming that, by combining corpus and GenAI approaches, we can gain a more comprehensive understanding of specific genre-based texts and therefore of the language used, for educational and research purposes. Until recently, corpora have been thought to offer a significant benefit over GenAI applications; it has been claimed that:

One of the main advantages of corpora is that we know exactly the domain of texts from which the corpus data is derived, something that we cannot track from current large language models underlying applications like ChatGPT (Crosthwaite and Baisa 2023).

However, Notebook LM is a special GenAI application in this respect, as it overcomes this shortcoming by allowing users to load their own trusted sources. Unlike ChatGPT, Gemini, Copilot, or any other similar AI-powered tools, Notebook LM only deals with the texts that are uploaded as the sources of data, and these sources can be our own corpora. Consequently, now, thanks to this

new GenAI research application, we can completely rely on the sources of data because we know which texts they are – only those we choose to upload. We can gain full control of the sources of data; thus, it is the type of queries and the type of data retrieved that vary compared with the traditional methods used in corpus linguistics, but the sources of data, the corpora themselves, can remain the same.

Section 2 describes the methodology followed, which makes use of a genre-based corpus as the source of data and some features of Notebook LM. Section 3 presents the case study comparing some of the results with previous findings from research on the same corpus. Finally, some conclusions are drawn in Section 4.

## 2. Methodology

In this paper, a comparable English-Spanish corpus has been used as a case study to illustrate a tentative research methodology based on Notebook LM. The two subcorpora of texts come from the same domain, so the genre is taken as the *tertium comparationis* to contrast the differences to express similar contents in the two languages involved. What was sought was how to find out information that traditional browsers or concordancers do not provide, so that an AI-powered tool such as Notebook LM can prove useful in corpus linguistics research. This methodology is not intended to replace corpus-analysis tools and procedures, but to focus on the integration of GenAI technologies within corpus-based research in order to widen the scope of the findings and provide further insight into the implications of this combination of tools and the ways to use them. The following subsections describe (1) the source of data, that is, the corpus used, (2) the tool, that is, Notebook LM, and (3) the procedure, that is, the steps followed, focusing on the comparison with previous results retrieved from the same corpus and corpus-analysis tools.

### 2.1. *The corpus: OCD*

The corpus used is called OCD, a comparable English-Spanish genre-based corpus of online cheese descriptions. It contains 267,364 words of running text, divided into two subcorpora. The English subcorpus contains 600 texts and amounts to 111,871 words. The Spanish subcorpus is similar in size, that is, the number of words is not much higher: 121,461, but it is composed of fewer texts, 400, because they are longer than the English texts. All of the texts, originally written in English and Spanish, were downloaded from freely available British and Spanish websites of cheese manufacturing companies, retailers, and general websites describing different types of cheese. These websites were selected because they offered longer cheese descriptions than other websites, where there

were only pictures and brief technical information. From each website, a variable number of texts was downloaded, generally between 7 and 10. Different types and brands of cheeses were chosen to show a wide range of different products and ensure representativeness, and despite the diverse origins of the texts, they all share similar characteristics and focus on promotional discourse in the cheese industry (for more information on OCD, see Labrador and Ramón 2024).

### 2.2. *The AI-powered research tool: Notebook LM*

Notebook LM is an AI-powered research and note-taking tool developed by Google Labs. It is designed to help users understand, synthesize, and generate insights from their documents. Unlike general AI chatbots, Notebook LM grounds its responses solely in the sources provided by the user, making it a personalized AI expert on the user's specific information – their reliable sources. As stated in Section 1, this unique characteristic makes it perfect to be used in corpus linguistics research. On the other hand, Google emphasizes that your personal data, including source uploads, queries, and model responses, remain private, that they are not used to train Notebook LM's underlying models and are not reviewed by human reviewers without permission.

Notebook LM can manage a wide range of formats: PDFs, Google Docs, Google Slides, website URLs, YouTube videos (it processes the transcript), audio files (it transcribes them), and even copied text. The user can create “notebooks” for different topics or projects, keeping their sources organized. It quickly generates summaries of individual documents or entire notebooks. You can ask specific questions about your sources, and NotebookLM will provide answers drawn directly from them. You can engage in a conversational chat with the AI about your material, and it can generate various outputs based on your sources, such as study guides (with glossaries, key concepts, and questions), FAQs, timelines, briefing documents, presentation outlines, video overview, audio overview, quiz, mind map, flashcards and there will be soon other functionalities available such as infographics and slide decks. The interface and features are rapidly changing.

### 2.3. *The procedure*

The method follows two stages that make use of some functionalities provided by Notebook LM. First, Notebook's built-in features have been used to summarize and explain the contents of the OCD corpus uploaded, and retrieve key terms and themes. These themes provided by Notebook LM have been compared with the moves and steps of the rhetorical structure that had been manually identified by the researchers who had built the OCD corpus. In

addition, some aspects that are highlighted in either the summary of the English subcorpus or the summary of the Spanish subcorpus have been compared with the frequency of occurrence of the moves and/ or steps related to those aspects in the corpus.

Secondly, some questions have been typed into Notebook LM's chat interface as prompts. The power of prompting, that is, of addressing specific queries, is crucial. The questions deal with both quantitative and qualitative data and have also been contrasted with previous findings of studies on the OCD corpus that relied on the information supplied by traditional corpus-analysis tools. This is the set of queries, addressed to the whole corpus or to each subcorpus, English and Spanish, separately, and the results are presented and discussed in the following section:

- Prompt 1: *How many words are there?* (whole corpus)
- Prompt 2: *What are the ten most frequent/ important words?* (English/ Spanish)
- Prompt 3: *What are the flavours mentioned?* (English/ Spanish)
- Prompt 4: *What are the textures mentioned?* (English/ Spanish)
- Prompt 5: *What types of evidence of quality assurance are there?* (English/ Spanish)
- Prompt 6: *What are the main differences between the two sources in terms of style and language use?* (whole corpus)

### 3. Case study

The genre of cheese descriptions has been selected as the case study to illustrate how to capitalize on genre-based corpora with the use of Notebook LM. The procedure has been divided into two main stages. The first stage, presented and discussed in Section 3.1, is based on the briefing document and list of key themes provided by Notebook LM as part of the built-in features yielded by default. This list has been compared with the rhetorical structure (moves and steps) of this genre. The second stage, presented and discussed in Section 3.2, revolves around 6 questions used as prompts (listed in Section 2.3) in Notebook LM's conversational interface and the outputs provided by Notebook LM to each of them.

#### 3.1. Built-in features

The briefing document yielded a list of key themes (on the right-hand side in Figure 1), which are very similar to the moves and steps (on the left-hand side in Figure 1) previously identified and used to tag the corpus (Labrador and Ramón 2015). They only differ slightly in the terms used, but the classification is similar: 1) Historical significance/ Historical provenance, 2) Protected Designation of Origin/ PDO, 3) Regionality and terroir/ Geographical provenance, 4)

Production methods/ Describing the process, 5) Variety of milk types/ Type of milk, 6) Importance of maturation/ Aging, 7) Use of vegetarian rennet/ Type of rennet, 8) Flavour profiles and unique characteristics/ Describing the product and describing the smell and taste, 9) Notable cheese types/ Name of the cheese and 10) Serving and Pairing suggestions/ Offering serving suggestions: temperature, presentation, food-pairing and wine-pairing.

**RESULTS: BRIEFING DOCUMENT**

**Key themes:**

- Historical significance
- Protected Designation of Origin
- Regionality and terroir
- Production methods
- Variety of milk types
- Importance of maturation
- Use of vegetarian rennet
- Flavour profiles and unique characteristics
- Notable cheese types
- Serving and pairing suggestions

**Move 1: IDENTIFYING THE CHEESE**  
Step 1: name of the cheese (and PDO (Protected Designation of Origin))  
Step 2: name of the company/ manufacturer  
Step 3: (geographical and historical) provenance

**Move 2: SHOWING A PICTURE**  
Move 3: DESCRIBING THE PRODUCT  
Step 1: characteristics: shape, size, texture, coating and weight  
Step 2: ingredients: type of milk and rennet used

**Move 4: DESCRIBING THE PROCESS**  
Step 1: preparation  
Step 2: aging

**Move 5: DESCRIBING THE SMELL AND TASTE**

**Move 6: OFFERING SERVING SUGGESTIONS**  
Step 1: temperature  
Step 2: presentation  
Step 3: food-pairing  
Step 4: wine-pairing

**Move 7: PROVIDING QUALITY ASSURANCE (quotations, awards, quality tests)**

Figure 1. List of rhetorical moves and steps in OCD (Labrador and Ramón 2015) compared with key themes retrieved by Notebook LM

Notebook LM also provided a summary of the English subcorpus and another of the Spanish subcorpus (Figures 2 and 3).

Capitalizing on genre-based corpora with the use of AI-powered research tool...

**Summary**

This document is a catalog of British cheeses, detailing their origins, production methods, flavor profiles, and suggested pairings. The entries highlight the diversity of British cheesemaking traditions, emphasizing the importance of local sourcing, time-honored techniques, and often featuring cheeses with Protected Designation of Origin (PDO) status. Many descriptions note awards received and provide serving suggestions, showcasing the cheeses' qualities and versatility. The overall purpose is to provide a comprehensive guide to a selection of notable British cheeses, celebrating their unique characteristics and regional identities.

**Key Topics**

- British cheeses
- Cheese production
- Cheese types
- Cheese flavors
- Cheese regions

Figure 2. Summary of the OCD English subcorpus and their corresponding key topics provided by Notebook LM

**Summary**

This document is a compilation of excerpts detailing the production methods and characteristics of various Spanish cheeses. The text focuses on the **processes of cheesemaking**, including milk sourcing (often from specific sheep or goat breeds), coagulation techniques (using animal rennet or plant-based alternatives), and aging procedures (in cellars, caves, or with added elements like smoke or herbs). Each cheese description highlights key **organoleptic properties** such as taste, texture, and aroma, along with suggested pairings. The overall purpose is to provide a comprehensive overview of traditional Spanish cheesemaking practices and the unique qualities of different regional cheeses.

**Key Topics**

- Cheesemaking process
- Cheese types
- Cheese flavor
- Cheese texture
- Cheese regions

Figure 3. Summary of the Spanish subcorpus and their corresponding key topics provided by Notebook LM

As can be seen in Figures 2 and 3, the summaries are quite similar; however, there are some significant differences. The summary of the English subcorpus mentions Protected Designation of Origin (PDO) status and awards received. On the other hand, the summary of the Spanish subcorpus emphasizes the type of rennet used and the cheese-making processes. These differences match previous findings based on traditional corpus analysis (Ramón and Labrador 2016) where it was found that: the rhetorical move of quality assurance (awards, prizes, PDO) occurs in almost half of the texts in English but less than 10% in Spanish (Figure 4), which accounts for the fact that PDO is mentioned in the summary of the English subcorpus but not in the summary of the Spanish subcorpus.

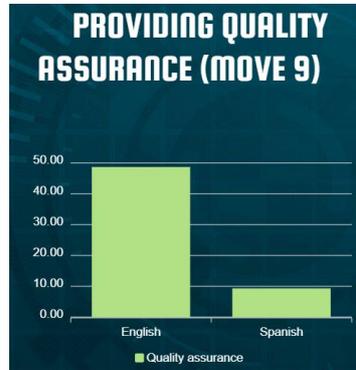


Figure 4. Number of texts including the Move: “Providing quality assurance” in the English and Spanish subcorpora in OCD (Ramón and Labrador 2016)

The previous analysis (Ramón and Labrador 2016) showed that the step of type of milk and rennet used occurs in 68% in English and 83% in Spanish (Figure 5).

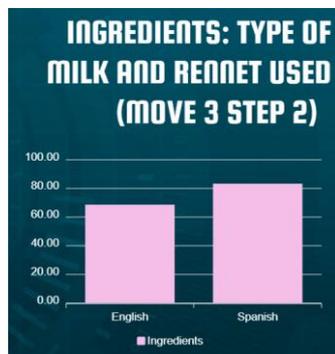


Figure 5. Number of texts including the Step: “Type of milk and rennet used” in the English and Spanish subcorpora in OCD (Ramón and Labrador 2016)

It also showed that the number of texts containing the step of preparation is 29% in English versus 44% in Spanish (Figure 6).

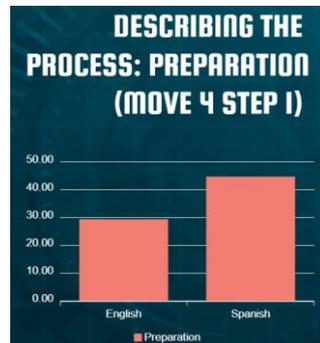


Figure 6. Number of texts including the Step: “Describing the process: Preparation” in the English and Spanish subcorpora in OCD (Ramón and Labrador 2016)

Finally, that study showed that the step of aging accounted for 18% in English versus 52% in Spanish (Figure 7). These data in Figures 5, 6 and 7 correlate with the fact that the type of rennet used and cheese-making processes are mentioned in the summary of the Spanish subcorpus but not in the summary of the English subcorpus (Figures 2 and 3), as they are not so frequent, thus not so relevant, in the English subcorpus.

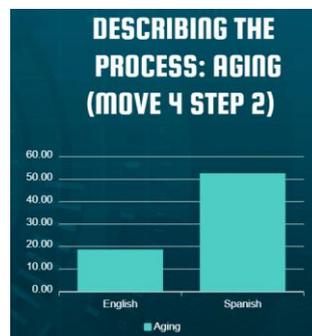


Figure 7. Number of texts including the Step: “Describing the process: Aging” in the English and Spanish subcorpora in OCD (Ramón and Labrador 2016)

These examples prove that the information supplied by Notebook LM (summaries of the subcorpora and lists of key themes and terms) aligns with the information supplied by corpus linguists (rhetorical structure of the corpus into

moves and steps) and by corpus browsers (number of texts in each subcorpus containing certain moves or steps). This supports the idea that, as well as the corpora and corpus browsers, corpus linguists can also take advantage of the data provided by Notebook LM in the rhetorical tagging process of corpus compilation.

### 3.2. Prompts

Regarding the prompts used, the first two questions referred to quantitative data: Prompt 1 was about the size of the whole corpus: *How many words are there?* And, curiously enough, Notebook LM could not answer this simple question. Its answer was:

The sources provided contain descriptions of many different types of cheeses, but do not mention the number of words in the document itself. Therefore, it is impossible to answer your query using only the provided sources.

Prompt 2: *What are the ten most frequent words?* was addressed to the English and the Spanish subcorpora separately. About the English subcorpus, again, as in Prompt 1, Notebook LM did not count the number of words but selected those that stand out because they are recurrent and because they seem to be central to understanding the texts. This was the output:

It is not possible to provide the ten most frequent words in the sources, as the provided text is a collection of excerpts from a PDF document, and not the full text. However, based on the content of the excerpts, here are some of the most frequently used and important words, along with a description of their significance. These words are central to understanding the different kinds of cheeses, their qualities, and their production methods as described in the document.

And the words provided were: *cheese, milk, texture, flavour, made, cheddar, style, British, vegetarian, and region*. When asked the same question about the Spanish corpus, as Notebook LM continues the conversations coherently, it refused to give the most frequent words:

The sources describe many different cheeses, but they do not provide any information about word frequency within the document itself. Therefore, it is **impossible to identify**

**the ten most frequent words only using the provided sources** [see conversation history]<sup>1</sup>

Consequently, the question was rephrased to ask about the most important, instead of frequent, words, and this is the list retrieved: *queso, leche, cabra, oveja, maduración, pasta, corteza, cruda, artesano, sabor*. Both *cheese, milk*, and *flavour* appear in both lists, but the other words are different. The English subcorpus seems to emphasize the importance of texture, provenance, suitability for vegetarians, and one cheese in particular, cheddar, whereas in the Spanish subcorpus, the type of milk, parts of the cheese, maturation degrees, and artisanal process are more relevant features.

The following questions addressed qualitative, rather than quantitative, aspects. Prompts 3 and 4 enquired about two important characteristics of cheeses: flavour and texture. The output to Prompt 3: *What are the flavours mentioned?* in the English subcorpus was: “The sources mention a wide variety of flavors associated with different cheeses, which can be broadly categorized as follows” and then it proceeded to provide this list, which includes not only the adjectives but also a classification into basic, specific, and combinations, descriptions and examples from the corpus, as shown in Appendix A. A similar, although slightly less elaborated, list was given as the output to the same prompt about the Spanish subcorpus. As English is the language of interaction, the list of flavours of the Spanish descriptions was provided in English, with some examples of words in Spanish as citations from the texts (Appendix B).

Similarly, Appendix C (concerning the English subcorpus) and Appendix D (concerning the Spanish subcorpus) show the lists of adjectives, explanations of the terms in this context, and some examples as answers to the question about textures, Prompt 4: *What are the textures mentioned?*

Similar lists of adjectives like those presented in Appendices A and C about the flavours and textures mentioned in the English subcorpus, have been retrieved with corpus-analysis toolkits mainly by examining the co-occurrences of English key terms like *flavour* (and related words such as *taste, tang, profile, finish, character, mouth, notes, palate, aftertaste* and *twist*) and *texture* (and related words like *body* and *consistency*) in a previous study (Ramón and Labrador 2018). Some of the adjectives found in that study are missing in the outputs provided by Notebook LM, even though the same corpus, OCD, has been used, for instance, in the case of taste: *clean, complex, strong, subtle, lasting, lingering, rich, full, mellow, zesty* and *citrus*, and in the case of texture: *close, dry, rich, velvety, light, semi-soft, hard, friable, springy*

---

<sup>1</sup> Boldface type used by Notebook LM.

and *rugged*. And vice versa, some adjectives retrieved by Notebook LM had not been found with the OCD corpus browser, for example, modifiers of taste like *sharp, earthy, herbal, lemony, floral, caramel, mushroomy, goaty, lactic, herbaceous* and *vegetal* and modifiers of texture like *gooey, pliant, mousse-like, brittle, sticky* and *melting*.

The third qualitative question, Prompt 5, was: *What types of evidence of quality assurance are there?* The answer given was interesting and complex, including more ideas than expected and reflecting a deep understanding of the texts. As can be seen in Appendix E, about the English subcorpus, and Appendix F, about the Spanish subcorpus, the types of evidence of quality assurance differ to a certain extent in both subcorpora, and more elements are mentioned that complement a previous corpus-based contrastive study on this move in OCD (Ramón and Labrador 2025). In this study, the type of quality assurance evidence found was mainly official certifications of quality, such as PDO and PGI (Protected Geographical Indication) and the awards, prizes and medals obtained. The output to Prompt 5 yielded more varied information regarding quality assurance indicators. As can be seen in Appendix E, the types of evidence in the English subcorpus included PDO, certification trademarks, handmade and traditional methods, use of raw milk, vegetarian rennet, specific ingredients such as annatto, nettle leaves, locally sourced milk, awards, maturation and grading, selection by experts, farm-specific practices, and unique terroir. In comparison, Appendix F shows the types of evidence mentioned in the Spanish subcorpus, which are: raw milk, artisanal production, controlled milk quality, natural ingredients and absence of additives, specific production processes, aging and maturation, microbiological analysis, denomination of origin, awards and recognition, controlled storage and shipping conditions. These lists include more quality indicators (and slightly different in each subcorpus) than the initial lists devised solely by traditional corpus-analysis tools and the linguists' intuition.

Finally, Prompt 6 was: *What are the main differences between the two sources in terms of style and language use?* And the answer was that they differ noticeably. As can be seen in the output provided in Figure 7, the English source tends to use a more straightforward, descriptive style, and the descriptions are concise and informative, whereas the Spanish source uses a more varied descriptive style, including more subjective descriptions and more evocative language. Spanish tends to embed more historical and cultural context and anecdotes; on the other hand, English includes historical information, but it is often briefer and more focused on the cheese production itself rather than its cultural significance. Unlike the informative and neutral tone of the English cheese descriptions, a more enthusiastic and celebratory tone is sometimes used in the Spanish descriptions. English presents a more uniform style across cheese entries, and, in

contrast, Spanish varies more in what details it includes; there is less consistency in the structure. In summary, the language in the English subcorpus tends to be more direct and factual, while the language in the Spanish subcorpus is more varied and expressive, often highlighting the historical and cultural significance of the cheeses it describes.

The two sources differ noticeably in their style and language use, reflecting their distinct origins and purposes.

**Language:** The most apparent difference is language. **x is written entirely in English**, while **y is primarily in Spanish**, with some English interspersed, especially in the "Notes" sections. This language difference alone suggests different target audiences and origins for the two documents.

**Descriptive Style:** The **English** source tends to use a **more straightforward, descriptive style**, often providing factual details about the cheeses. It focuses on describing the characteristics of each cheese in terms of flavor, texture, and production methods. The descriptions are **concise and informative**. The **Spanish** source uses a **more varied descriptive style**. While it also provides factual information, it often includes **more subjective descriptions**, such as the "sonorous name" of a cheese related to its texture and sometimes uses **more evocative language** to describe tastes and aromas. For example, some cheeses are described as "creamy and unctuous" or having a "strong and spicy flavor".

**Use of Technical Terms:** Both sources use technical terms related to cheesemaking, but in their respective languages. The English source might use terms such as "washed curd" or "affinage period", while the Spanish source uses terms such as "cuajada" (curd) or "Denominación de Origen" (appellation of origin).

**Historical and Cultural Context:** The **Spanish** source **tends to embed more historical and cultural context** into the descriptions of the cheeses. It often **includes anecdotes** about the origins of the cheese or its role in local traditions, such as Cabrales cheese having more than 1,000 years of history, or Torta del Casar being used as a form of payment in ancient times. The **English** source **includes historical information**, but it is often **briefer and more focused on the cheese production itself rather than its cultural significance**.

**Tone:** The tone of the **English** document is generally **informative and neutral**. The **Spanish** source, while also informative, sometimes uses a **more enthusiastic and celebratory tone**, reflecting a cultural appreciation for the cheeses being described. This is particularly evident in the descriptions of traditional or artisanal cheeses. For example, it refers to Cabrales as "one of the great hallmarks of Asturias".

**Structure and Organization:** Both sources organize the information by cheese type. However, the **English** source presents **a more uniform style across cheese entries**, typically including details of origin, milk type, style of cheese, vegetarian status, and other key features. The **Spanish** source often **varies in what details it includes** for each cheese type, sometimes focusing more on history, other times on production, and other times on taste, texture, or how the cheese is best served, reflecting **less consistency in structure**.

In summary, while both sources provide information about cheese, they do so using **distinct styles and language**, influenced by their respective cultural contexts and intended audiences. The **English** source is **more direct and factual**, while the **Spanish** source is **more varied and expressive, often highlighting the historical and cultural significance** of the cheeses it describes.

Figure 7. Notebook LM's answer to Prompt 6: "What are the main differences between the two sources in terms of style and language use?" (my own highlighting in yellow).

Although both subcorpora in OCD belong to the same genre, online cheese descriptions, so they deal with the same text-types, the same topic, the same domain and the same purpose, Notebook LM has revealed some cross-linguistic differences in terms of tone, style, and emphasis on some details rather than others. The use of traditional corpus-analysis tools for contrastive research on key words, their frequency rates, collocates, and concordance lines can lead linguists to infer similar interpretations about differences in the style and language use of genre-based comparable subcorpora. However, a general overview, like the output to Prompt 6, can help to contrast and complement the quantitative data found in word lists, lists of collocates and concordance lines and their corresponding interpretations.

### Conclusions

This paper has explored some of the new possibilities that can be found at the crossroads of corpus linguistics and GenAI technologies. The answers to the first two prompts clearly show that Notebook LM is not software programmed to perform functions like counting the number of words, but a research tool that emulates human cognitive processes. If someone were given a book and asked, “How many words are there?”, they would not be able to answer. Nevertheless, if they were given some time to read it, they would probably be able to remember it, understand it, explain it, and interpret it. They would be able to answer questions about the book. That is exactly what Notebook LM does, but much more quickly and accurately: in only a few seconds and without missing any details that might be otherwise overlooked. Its limitations, however, lie in providing quantitative information, which is much better provided by traditional corpus-analysis tools.

The case study presented here has answered the research questions by illustrating how Notebook LM can be used for contrastive linguistic research on genre-based corpora. The information retrieved through the summary, the key themes and terms and the outputs to the prompts is more related to meaning than form. It has to do more with keywords and meaning associations than frequency rates, percentages, or numbers. Notebook LM can help us to discover results that we would not be able to find by only using corpus analysis tools, for example, as shown in this paper, a broader list of quality assurance indicators or more adjectives that describe flavour and texture, as well as classifications and descriptions of these modifiers. It can assist at different stages of corpus-building and corpus-based research, for instance, triggering new ideas for further research paths, identifying moves and steps for rhetorical tagging, interpreting the data or drawing general conclusions about cross-linguistic contrasts. This paper has

shown that we can search in each subcorpus separately, but also in both, allowing comparisons to be made between the writing conventions in each language in a particular domain, which can enhance corpus-based contrastive studies and assist in second-language writing and translation training.

It took us some time to adapt to the types of queries that we needed as formal inputs for searches in corpora. Now we need to rethink the type of queries we want with the current tools at our disposal, like GenAI technologies. The advantages of these queries are that they can be more flexible, we can use natural language, and we can add follow-up questions, as we are dealing with dialogical applications. The main role of the linguist has tilted towards being creative in asking questions and insightful in interpreting the data, without ceasing to rely on traditional corpus-analysis tools for the retrieval of quantitative data.

All in all, the main conclusion drawn is that Notebook LM constitutes a powerful tool to capitalize on genre-based corpora without interfering with other corpus-analysis applications; together, they can provide a more holistic approach.

## References

- AIJMER, Karin; ALTENBERG, Bengt; JOHANSSON, Mats (eds.), *Languages in contrast. Papers from a symposium on text-based cross-linguistic studies. Lund 4-5 March, 1994*. Lund: Lund University Press, 1996.
- ALONSO-GUISANDE, Miguel Ángel; LÓPEZ FRAILE, Luis Antonio, “Herramientas de inteligencia artificial generativa aplicadas en la producción de podcasts”. In: *Edu Review*, 12, 2, 2024, pp. 19-32. Accessed at: <https://doi.org/10.62701/revedu.v12.5409>
- BIBER, Douglas; CONRAD, Susan; REPPEN, Randi, *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press, 1998.
- CROSTHWAITE, Peter; BAISA, Vit, “Generative AI and the end of corpus-assisted data-driven learning? Not so fast!”. In: *Applied Corpus Linguistics*, 3, 3, 2023. Accessed at: <https://doi.org/10.1016/j.acorp.2023.100066>
- CROSTHWAITE, Peter (ed.), *Corpora for language learning: Bridging the Research-Practice Divide*. London/ New York: Routledge, 2024.
- CURRY, Niall; BAKER, Paul; BROOKES, Gavin, “Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT”. In: *Applied Corpus Linguistics*, 4, 12024, 2024. Accessed at: <https://doi.org/10.1016/j.acorp.2023.100082>
- FILLMORE, Charles J., “‘Corpus linguistics’ or ‘Computer-aided armchair linguistics’”. In: Svartvik, Jan (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Berlin/ New York: De Gruyter Mouton, 1992, pp. 35-60. Accessed at:

- <https://doi.org/10.1515/9783110867275.35>
- GADSDEN, Amy Dominique; GOEGAN, Lauren. Denise, “Informing Inclusive Practice in Post-Secondary Environments: Perspectives of Post-Secondary Instructors with Learning Disabilities”. In: *The Canadian Journal for the Scholarship of Teaching and Learning*, 14, 2, 2023. Accessed at: <https://doi.org/10.5206/cjsotlrcacea.2023.2.8020>
- HOEY, Michael, *Lexical Priming: A New Theory of Words and Language*. London/New York: Routledge, 2005.
- JOHANSSON, STIG; OKSEFJELL, SIGNE (eds.), *Corpora and cross-linguistic research. Theory, method and case studies*. Amsterdam/ New York: Rodopi, 1998.
- LABRADOR, Belén; RAMÓN, Noelia, “Positive Evaluation in the Translation of Online Promotional Discourse in the Cheese Industry”. In: *IEEE Transactions on Professional Communication*, 67, 3, 2024, pp. 316-332. Accessed at: <https://doi.org/10.1109/TPC.2024.3417056>
- , “‘Perfectly Smooth, Creamy and Full Flavoured’: Online Cheese Descriptions”. In *Procedia - Social and Behavioral Sciences*, 198, 2015, pp. 226-232. Accessed at: <https://doi.org/10.1016/j.sbspro.2015.07.440>
- MORINA, Anabel; CARBALLO, Rafael; DOMÉNECH, Ana, “Transforming higher education: a systematic review of faculty training in UDL and its benefits”. In: *Teaching in Higher Education*, 2025, pp. 1-18. Accessed at: <https://doi.org/10.1080/13562517.2025.2465994>
- MEHTA, Neil; AGRAWAL, Anoop; BENJAMIN, Jennifer; MEHTA, Seysha, MACNEILL, Heather; MASTERS, Ken, “Pedagogy and generative artificial intelligence: Applying the PICRAT model to Google NotebookLM”. In: *Medical Teacher*, 47, 5, 2024, pp. 788-790. Accessed at: <https://doi.org/10.1080/0142159X.2024.2418937>
- OKSEFJELL EBELING, Signe, “Corpus-based contrastive studies and AI-generated translations”. In: *Languages in Contrast*, 25, 2, 2025, pp. 289-315. Accessed at: <https://doi.org/10.1075/lic.00051.ebe>
- RAMÓN, Noelia; LABRADOR, Belén, “Describing cheese in English and Spanish: A corpus-based contrastive study”. In: *8<sup>th</sup> International Conference on Corpus Linguistics (CILC 2016)*. Málaga, 2-4 March 2016.
- , “Selling cheese online: key nouns in cheese descriptions”. In: *Terminology*, 2018, pp. 210-235. Accessed at: <https://doi.org/10.1075/term.00019.ram>
- , “Comparing the expression of quality assurance in English and Spanish online cheese descriptions” Presented in: *16<sup>th</sup> International Conference on Corpus*

Capitalizing on genre-based corpora with the use of AI-powered research tool...

*Linguistics Meaning in Corpus: Tools, Methods and Approaches to 'Aboutness'*.  
Salamanca, 15-17 May 2025.

SANZ-VALDIVIESO, Lucía; LÓPEZ-ARROYO, Belén, “Human vs. ChatGPT corpus annotation: Data augmentation using LLM fine-tuning”. In: Rabadán, Rosa; Ramón, Noelia (eds.), *Cross-linguistic mediated communication. Hybrid text production English-Spanish*. Berlin: Peter Lang, 2025.

SINCLAIR, John, *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

SVARVTIK, Jan (ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Berlin/ New York: Mouton de Gruyter, 1992.

TORRENT, Tiago; HOFFMANN, Thomas; LORENZI, Arthur; TURNER Mark, *Copilots for Linguists: AI, Constructions, and Frames*. Cambridge: Cambridge University Press, 2024.

**Appendix A.** Notebook LM's output to Prompt 3: "What are the flavours mentioned?" about the English subcorpus.

**Basic Taste Descriptors:**

**Creamy:** Many cheeses are described as having a creamy flavor, which refers to a smooth, rich, and often buttery taste. This is a common characteristic of many British cheeses.

**Mild:** This term is used to describe cheeses with a gentle and subtle taste.

**Buttery:** This term is used to describe cheeses that have a rich flavor reminiscent of butter.

**Sharp:** This term is frequently used for Cheddar, indicating a strong, tangy, and sometimes acidic flavor.

**Sweet:** Some cheeses have a sweet taste, which can come from the milk itself or from added ingredients such as fruit.

**Tangy:** Many cheeses are described as having a tangy flavor, indicating a slight sharpness and acidity.

**Salty:** Some cheeses, especially those stored in brine like Halloumi, have a salty taste.

**Savory:** This term is used to describe cheeses that have a rich, non-sweet flavor, sometimes with a meaty or umami quality.

**Nutty:** This flavor is common in many aged cheeses and is reminiscent of nuts.

**Specific Flavor Notes:**

**Fruity:** Several cheeses are described with fruity notes, often from the milk or added fruits, and including notes of **apricot, pineapple, cranberry, or citrus.**

**Earthy:** Some cheeses have an earthy flavor, often associated with the rind or the aging process.

**Herbal:** This term describes cheeses with flavors of herbs such as **chives, garlic, nettle, and sage.**

**Smoked:** Some cheeses are smoked, resulting in a smoky flavor.

**Spicy:** Cheeses can have a spicy kick from added ingredients such as chili.

**Lemony:** Some cheeses have a lemony flavor with a citrusy tang.

**Floral:** Some cheeses have a delicate floral aroma and taste.

**Caramel:** A caramel-like flavor is noted in some mature cheeses.

**Mushroomy:** Some cheeses are said to have a mushroomy flavor, particularly brie styles.

**Goaty:** Some goat's milk cheeses have the characteristic "goaty" flavor.

**Peppery:** Some cheeses have a peppery, spicy flavor that is sometimes associated with blue cheeses.

**Lactic:** This term indicates a fresh, milky taste, often slightly acidic, which is a common characteristic of fresh cheeses or cheeses that have not been aged long.

**Grassy:** Some cheeses have a grassy flavor that is fresh, like fresh cut grass or hay.

**Herbaceous:** Some cheeses have an herbaceous taste, reminiscent of green herbs.

**Vegetal:** Some cheeses have vegetal notes that can be similar to that of porcini mushrooms.

**Minerally:** Some cheeses have a flinty, mineral taste.

**Flavor Combinations:**

The sources also mention specific flavor combinations, such as **Cheddar with caramelized onions, Wensleydale with cranberries, Double Gloucester with chives and onions, Cheddar with balsamic onions,** and others that indicate specific pairings of tastes.

**Appendix B.** Notebook LM's output to Prompt 3: "What are the flavours mentioned?" about the Spanish subcorpus.

**Acidic/Slightly Acidic:** This is a common characteristic, often described as "ligeramente ácido". Some cheeses have a more pronounced acidity, while others are described as having a "delicate" or "elegant" acidity.

**Lactic:** Many cheeses are noted for their milky or lactic flavors, sometimes with a fresh, "dulzón" or sweetish note.

**Salty:** The level of saltiness varies, with some cheeses having a "correct point of salt", while others are described as "slightly salty".

**Buttery:** A creamy, buttery flavor is often mentioned, with some cheeses having a "mantecoso" (buttery) texture.

**Smooth/Mild:** Several cheeses are characterized by their "suave" or mild flavor profiles, often with a creamy or "mantecosa" texture.

**Strong/Intense:** In contrast, some cheeses have a "fuerte" or strong flavor, sometimes described as "intenso".

**Piquant/Spicy:** A "picante" or spicy flavor is also noted in some cheeses.

**Aromatic:** Many cheeses have notable aromas, ranging from "lácticos" (milky) and "afrutado" (fruity), to "herbáceo" (herbaceous) and "ahumado" (smoky).

**Nutty:** Some cheeses have a flavor reminiscent of nuts, such as almonds or hazelnuts.

**Sweet:** A few cheeses are described as having a "dulzón" or sweetish flavor.

**Earthy/Fungal:** Some cheeses have a flavor profile that includes notes of "setas" (mushrooms), "tierra húmeda" (damp earth) or "bosque" (forest).

**Caramelized:** The Idiazabal cheese has a flavor reminiscent of "caramelo quemado" (burnt caramel).

**Winey:** The Murcia al Vino cheese has a flavor with notes of wine due to the way the rind is washed during production.

**Smoked:** Several cheeses have a smoky flavor from the way they are cured.

**Herbal:** Some cheeses have flavors of rosemary or other herbs.

**Appendix C.** Notebook LM's output to Prompt 4: "What are the textures mentioned?" about the English subcorpus.

**Creamy:** Many cheeses are described as having a **creamy** texture. This indicates a smooth, rich, and often spreadable consistency. Some examples include Colston Bassett Stilton, Sharpham Elmhirst, and Blacksticks Blue.

**Crumbly:** This texture is frequently mentioned, referring to a cheese that breaks apart easily into small pieces. Examples include Cheshire, Wensleydale, and Cotherstone.

**Firm:** This indicates a cheese that is solid and holds its shape well. Mature Cheddar, Double Gloucester, and Old Harry are examples of cheeses with a firm texture.

**Smooth:** This term is used for cheeses that have a uniform, even consistency without lumps or grains. Examples include Ashmore Farmhouse, Sharpham Elmhirst, and Coquetdale.

**Soft:** This texture refers to cheeses that are pliable and easily yield to pressure. Examples include Slipcote, Baron Bigod, and Dorstone.

**Dense:** This term is used for cheeses that are compact and tightly packed. Examples include Abbaye de Belloc, Spenwood, and Old Ford.

**Open:** Some cheeses are described as having an open texture, meaning they have small holes or spaces throughout. This characteristic is found in cheeses like Creamy Lancashire and Isle of Mull.

**Moist:** This term indicates that a cheese has a high moisture content. Some cheeses that are described as moist are Bedale and Lincolnshire Poacher.

**Silky:** This texture refers to a smooth and delicate feel in the mouth. Examples include Baron Bigod and Sharpham Elmhirst.

**Flaky:** This texture refers to cheese that breaks into thin layers or small pieces. Examples include Montgomery's Cheddar and Sparkenhoe.

**Gooey:** This term refers to cheeses that have a soft, liquid-like texture. Examples include Winslade and Sharpham.

**Runny:** This refers to cheeses with a very soft, liquid texture, particularly when ripe. Examples include Basing, Innes Brick and Renegade Monk.

**Grainy:** This texture describes cheeses with small, distinct particles that can be felt in the mouth. Examples include Ricotta and Isle of Mull.

**Supple:** Some cheeses are described as supple, meaning they are flexible and yielding. Examples include Miss Muffet and Sharpham Savour.

**Chewy:** This term describes cheeses that are firm and elastic. An example is British Gruyere.

**Pliant:** This term describes a texture that is easily bent and flexible. An example is Sharpham Savour.

**Mousse-like:** This term describes a very light and airy texture. Examples include Dorstone and Tysboro.

**Brittle:** This term describes a texture that is easily broken. Example: Montgomery's Cheddar.

**Sticky:** Some cheeses, particularly those with washed rinds, are described as sticky. Examples include Basing and Colston Bassett Stilton.

**Melting:** Some cheeses are described as having good melting characteristics. Examples include Creamy Lancashire, Teifi, and Mayfield.

**Appendix D.** Notebook LM's output to Prompt 4: "What are the textures mentioned?" about the Spanish subcorpus.

**Creamy/Untuosa:** This is a frequently used term ("cremoso" or "untuosa") to describe cheeses that are smooth, rich, and often spreadable, such as the Torta del Casar, and some fresh cheeses. These cheeses are often described as having a high level of "mantecosidad" or butteriness.

**Soft/Blanda:** Many cheeses are described as having a "blanda" or soft texture, which can range from very soft and almost liquid to a pliable, yielding texture. The Monte Enebro and other goat cheeses are often in this category.

**Firm/Firme:** Some cheeses are noted for their "firme" or firm texture, which can be either dense and compact, or slightly more yielding. The Manchego is described as having a firm paste.

**Hard/Dura:** Cheeses that are described as "dura" are dense, solid, and often aged for a long time, like the Castellano or the Roncal cheese.

**Semi-hard/Semi-dura:** Many cheeses fall into the "semi-dura" category, indicating a texture that is neither too soft nor too hard, such as the Cantabria, or the Serrat.

**Compact/Compacta:** This term indicates a dense texture with few or no air pockets, such as the interior of the Castellano and the Pedroches.

**Granular/Granulosa:** Some cheeses are described as having a "granulosa" or granular texture, which can range from a slightly grainy feel to a more pronounced graininess, such as the Blanco cheese from Asturias.

**Laminar:** Some cheeses, like the Castellano, have a laminar texture, meaning they have thin, layered structures within the paste.

**Friable:** This term describes a texture that is easily crumbled or broken, as seen in aged cheeses.

**Elastic/Elástica:** Cheeses described as "elástica" are those that have a springy texture and can stretch or flex without breaking, like some of the mixed milk cheeses.

**Soluble:** Some cheeses are described as "soluble", meaning that they readily melt in the mouth, such as the tierno cheese.

**Adhesive/Adhesiva:** This describes cheeses that tend to stick to the mouth when eaten, as seen in the Flor de Tauste or the Franxón.

**Moist/Húmeda:** This indicates that the cheese has a certain level of moisture, especially if it has been recently made or is fresh, like the Drap, and some of the fresh cheeses mentioned. Some cheeses are described as having a "moderately" moist surface.

**Dry/Seca:** Some aged cheeses are described as dry or having a dry texture, like the Mahón, indicating a lower moisture content.

**Oily/Aceitosa:** Some cheeses are described as having an oily or unctuous texture, especially when they are aged.

**Pasty:** The Torta de la Serena can become so soft that it is described as having a pasty or even liquid consistency.

**Appendix E.** Notebook LM's answer to Prompt 5: "What types of evidence of quality assurance are there?" about the English subcorpus.

**Protected Designation of Origin (PDO):** Several cheeses are identified as having a PDO status. This designation means that the cheese must be produced in a specific region using traditional methods, with milk from that area. This serves as a guarantee of the cheese's authenticity and quality.

**Certification Trademarks:** Some cheeses have certification trademarks, like Stilton, which indicate that the cheese meets specific quality and production standards.

**Handmade and Traditional Methods:** Many cheeses are described as **handmade** or made using **traditional methods**. This suggests a commitment to quality and craftsmanship, as opposed to mass-produced cheeses. For example, West Country Farmhouse Cheddar must be made on a farm by hand using authenticated methods. Some specific traditional methods include:

**Hand-cheddaring.**

**Cloth-binding.**

**Hand-ladling of curd.**

Using **pint starters** instead of powdered versions.

**Use of Raw (Unpasteurized) Milk:** Some cheeses are made with unpasteurized or raw milk. This is often associated with more complex and stronger flavors and is often an indicator of traditional farmhouse production.

**Vegetarian Rennet:** The use of vegetarian rennet is noted for many cheeses, which indicates that no animal products are used in the cheese-making process.

**Specific Ingredients:** The sources sometimes note the use of particular ingredients, such as:

**Annatto,** a natural dye used to give cheeses their orange hue.

**Nettle leaves** wrapped around Cornish Yarg.

**Locally sourced milk.**

**Awards:** Many cheeses are described as award-winning, which indicates that they have been recognized for their high quality and taste by expert judges.

**Maturation and Grading:** Some cheeses are subject to specific maturation periods and regular grading throughout their life. For example, West Country Farmhouse Cheddar must be matured for a minimum of 9 months.]

**Selection by Experts:** Some cheeses are hand-selected by experts at a certain point in their aging process, ensuring that only the best cheeses are sold.

**Farm Specific Practices:**

Some cheesemakers have their own herds and use milk from their own cows.

Others use milk from local farms.

Some farms have been making cheese for multiple generations, which indicates a commitment to traditional methods and quality.

**Unique Terroir:** Some cheeses benefit from the unique characteristics of their local environment, such as the salt deposits in the Cheshire area which contribute to the milk's unique flavor.

**Appendix F.** Notebook LM's answer to Prompt 5: "What types of evidence of quality assurance are there?" about the Spanish subcorpus.

**Raw Milk (Leche Cruda) and its Source:** Several cheeses are made with raw milk, which is unpasteurized. The use of raw milk is presented as a sign of quality and tradition, as it retains the natural flavors and characteristics of the milk. The specific breeds of animals providing the milk, such as the Payoya goat, or the Lacha sheep, are often mentioned as contributing to quality. Some farms, like the one that produces the San Pelegrín, control the feed of their animals, which is also presented as an important factor for quality.

**Artisanal Production (Elaboración Artesanal):** Many descriptions emphasize that the cheese is made using traditional, artisanal methods. This often involves manual processes, such as hand-molding, and a focus on time-honored techniques passed down through generations.

**Controlled Milk Quality:** Some producers ensure the quality of their milk through their own controlled herds, or buy it from local farms that are carefully checked for health and hygiene. This control is noted as a key factor in the quality of the final product.

**Natural Ingredients and Absence of Additives:** Many cheeses are described as being made with natural ingredients and without artificial preservatives or additives. The use of natural rennet (cuajo natural) is also highlighted. Some producers only use salt, natural rennet and other natural ingredients like herbs, spices, and olive oil.

**Specific Production Processes:** The sources detail specific steps in the cheesemaking process, such as the type of coagulation, the temperature at which the milk is heated, how the curds are cut, how the cheeses are pressed or molded, and the brining method. These details suggest a careful and controlled approach to cheesemaking. For instance, the slow acidification of the milk in the production of some cheeses, or the washing of the rind, such as in the Murcia al Vino, which uses red wine, are described as unique and quality-enhancing procedures.

**Aging and Maturation:** The length of maturation is a significant factor for quality, as this influences the flavor and texture of the cheese. Some cheeses are aged in natural caves, which is described as contributing to their unique characteristics. Some are also aged in contact with different materials, such as olive oil, or covered in manteca (lard).

**Microbiological Analysis:** Several sources state that their cheeses are tested to be free of harmful microorganisms, and are subject to physical, chemical and microbiological analysis, indicating a focus on safety.

**Denomination of Origin (D.O.P.):** Some cheeses have a protected Denomination of Origin status which guarantees they are produced in a specific region using specific methods, ensuring quality and authenticity.

**Awards and Recognition:** Many cheeses have won national and international awards, like the World Cheese Awards, or the Cincho awards, mentioned as evidence of their high quality. Some cheeses have also won awards in the UK and US, further confirming their quality.

**Controlled Storage and Shipping Conditions:** Some producers emphasize the importance of maintaining the cold chain and specific temperatures to preserve the quality of their fresh cheeses.