Journal of English for Academic Purposes 65 (2023) 101285

Contents lists available at ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Using legitimation code theory to investigate English medium lecturers' knowledge-building practices

Irina Argüelles-Álvarez [a,*], Tom Morton [b]

[a] Department of Linguistics Applied to Science and Technology, Universidad Politécnica de Madrid, Spain
[b] Department of English Philology, Universidad Autónoma de Madrid, Spain

ARTICLE INFO

Handling editor: Guangwei Hu

Keywords:
English medium instruction (EMI)
Legitimation code theory (LCT)
Knowledge building
Semantics
Semantic density
Semantic gravity

ABSTRACT

This study uses Legitimation Code Theory (LCT) to examine the knowledge-building practices of two lecturers teaching computing courses through English as a medium of instruction (EMI) at a Spanish university. LCT is a sociological framework for exploring and improving knowledge practices across academic and other fields of activity. The study uses the LCT dimension of Semantics, which sees knowledge building in terms of condensation of meaning (semantic density) and context-dependence (semantic gravity). Five video-recorded sessions were transcribed and coded using the text annotation software CorpusTool, and the analysis traced variation in semantic density and gravity over the teaching sessions (both within and across the two lecturers' practices). The findings show that the lecturers dealt with complexity of meaning and context-relatedness in ways which reflected the nature of the content topic and the teaching activity. There was evidence that they used "semantic waves" (movements between higher and lower semantic density and stronger and weaker semantic gravity) to build knowledge cumulatively over the sessions. We identify implications for the professional development of EMI lecturers, arguing that LCT Semantics has the potential to help lecturers see connections between their disciplinary knowledge building practices and the communicative resources used to enact them.

1. Introduction

In teaching disciplinary content, university lecturers in all subjects need to deal with complex ideas and to place them in some kind of context in order to facilitate comprehension on the part of their students. This is something which is true irrespective of medium of instruction, whether the first or any additional language(s). However, relatively little is known about how lecturers across disciplinary areas and different medium of instruction scenarios deal with this aspect of knowledge building. The study reported in this article examines how complexity and context-embeddedness are dealt with in university English Medium Instruction (EMI), where English is used as the language of teaching and learning in a non-Anglophone context. Specifically, it draws on the sociological framework of Legitimation Code Theory (LCT), particularly its Semantics dimension (Maton, 2014) to explore how two EMI lecturers in a Spanish university dealt with complexity of meaning and its embeddedness in context in two computer education modules.

The use of a sociological, knowledge-building framework to explore the lecturers' knowledge building practices is deliberate. On the one hand, it addresses the issue of "ownership" of EMI, which as Macaro and Aizawa (2022) argue, has been rather "one-sided" in that research has been dominated by specialists in applied linguistics and language education, to the exclusion of disciplinary content

* Corresponding author.
E-mail address: irina.arguelles@upm.es (I. Argüelles-Álvarez).

https://doi.org/10.1016/j.jeap.2023.101285
Received 9 January 2023; Received in revised form 20 July 2023; Accepted 25 July 2023
Available online 28 July 2023
1475-1585/

specialists. As shown in several studies (e.g., Moncada-Comas & Block, 2021), content lecturers in university EMI contexts are often unwilling to be "CLIL-ised" by taking on language teaching duties. Even calls for a focus on disciplinary literacies and more cross-fertilization with work in EAP such as the use of genres (Dafouz, 2021; Wingate & Hakim, 2022) may meet resistance. Starting from disciplinary knowledge building practices may speak more directly to the concerns of subject lecturers as for them, the primordial concern is with their students' learning of disciplinary knowledge and skills. Specifically, it can be productive to focus on how they deal with complexity of meaning and how they relate concepts to different contexts. This can enable them to reflect in a more focused manner on the linguistic and communicative resources they use to build knowledge in their teaching practices.

The exploratory study reported in this article describes how two EMI lecturers deal with complexity of meaning and context embeddedness in their classroom discourse. The two modules are on computing topics, and the principal aim is to show how they play out in different, but related, contexts. Focusing on these two main notions of complexity and context embeddedness can provide conceptual tools for an initial reflection on existing pedagogical practices as seen in classroom discourse, and a point of departure for the planning of instructional sequences drawing on constructs from applied linguistics, language and literacy studies. In the next section, we introduce Legitimation Code Theory, with a specific focus on its Semantics dimension. Then, we go on to show how its conceptual framework fed into the methodological approach of the study.

2. Legitimation code theory: Semantics

Legitimation Code Theory (LCT) is a sociological framework for the investigation and improvement of knowledge practices which emerged in the late 1990s and early 2000s (Maton, 2014; Maton et al., 2016). It draws on and develops Bernstein's "code theory" (1990; 2000) and Bourdieu's "field theory" (e.g., Johnson & Bourdieu, 1993), seeing participants' knowledge practices within specific fields in terms of competing claims to legitimacy, with "legitimation codes" representing what is valued by actors as a basis for success in the field. It seeks to overcome what Maton (2014) describes as "knowledge blindness" in the social sciences, in which knowledge itself, its forms, effects and consequences are rarely if ever taken into account. Thus, for LCT, knowledge is real and has real effects in the world, and the organizing principles of knowledge practices, the "rules of the game" can be uncovered through empirical research using LCT tools. In revealing the organizing principles of knowledge practices, LCT provides tools to overcome fragmentation in academic fields, and to contribute to more cumulative knowledge building. The different dimensions of LCT (see below) provide a set of empirical tools which can be applied to problems and issues concerning knowledge building across all fields, and which allow researchers to cumulatively build knowledge in addressing similar problems.

Current research in LCT draws mainly on three dimensions of the framework: Specialization, Semantics and Autonomy. Specialization sees educational activity as knowledge-knower structures with varying strengths of emphasis on what is to be known and who the knowers are. Semantics explores how content is dealt with in more or less complex ways, with stronger or weaker embeddedness in specific contexts. Autonomy explores how practices can be seen in terms of boundaries between different types of knowledge objects and the uses to which they are put. The current study uses Semantics, and this dimension is further elaborated on here.

In this dimension the key concepts are *semantic gravity* (SG) and *semantic density* (SD). When semantic gravity is strong (SG+), knowledge is firmly grounded in a specific context, perhaps with examples from everyday life, or an artifact or visual present in the physical surroundings. With weak semantic gravity (SG-), knowledge is more "free-floating", removed from any specific context. High semantic density (SD+) indicates that ideas, concepts, texts, or visuals are highly condensed or "packed" with meaning. They may be very technical and will need to be "unpacked" for non-experts. When semantic density is lower (SD-), meanings are less complex, more transparent, and do not normally need unpacking.

In LCT studies it is not appropriate to code data directly using the theory's conceptual categories or codes. It is first necessary to develop a "translation device" (Maton & Chen, 2016) which relates the LCT concepts to categories emerging in the data. Translation devices can be generic in that they can be applied to a very wide range of data, or they can be developed for specific datasets. A basic translation device for Semantics analysis can be in the format of a table with three columns. The first column includes the LCT concepts applied in the analysis (in this case semantic gravity or density), while the second column can include indicators to enable the location of examples in the data. The third column can include extracts from the data itself. The translation devices for the current study are described and shown in detail in the methodology section below.

Semantic gravity and density can be analyzed either separately or in combination. When analyzed together they can also be visualized along two axes to reveal four *semantic codes.* In this exploratory study, we coded the data for both semantic gravity and density independently since we were more interested in seeing how semantic density and gravity varied as the sessions unfolded rather than how they interact to reveal distinct codes. One of the strengths of measuring semantic density and gravity separately is that it can show how the values change over time as teaching/learning events unfold. Thus, variation in the values of semantic density and gravity are not random affairs, as some combinations of their values as instruction moves through a specific topic or a larger chunk of curriculum content can be more effective pedagogically than others.

The changes produced between semantic gravity and semantic density over time lead to semantic profiles, one of which can be represented by means of semantic waves (Maton, 2014, 2020). Semantic waves are movements in variation of semantic density and gravity through time in a teaching/learning event. Stronger or weaker SG (more concrete or more abstract forms of knowledge) and/or higher or lower SD (more or less condensed meanings) can be visually shown by means of wave strokes that move up and down and over time, thus providing a useful representation of SG and/or SD flow. When semantic gravity is weak and semantic density high, this is the high point of a wave – meanings will be abstract and highly condensed and removed from any specific context of use or application. When semantic gravity is strong and semantic density is low, this is a low point of a wave, and meanings will be less condensed, more "everyday" and grounded in a specific context or concrete reality.

Fig. 1 shows the three basic semantic profiles within LCT. The vertical axis on the left shows the varying strengths of Semantic Gravity (SG) and Semantic Density (SD) in inverse relationship, and the horizontal axis represents time unfolding. From the point of view of making knowledge building more cumulative, "waving" semantically (Fig. 1: C) is preferable to maintaining either a "high semantic flatline" (Fig. 1: A) where meaning is abstract and removed from context, or a "low semantic flatline" (Fig. 1: B) where learners may be stuck at the level of concrete experience without having the opportunity to make links to conceptual frameworks (Maton, 2020).

According to Maton (2016: 18) the movements from abstract to practical or vice versa, or from complex to simpler meanings in semantic waves are crucial for cumulative knowledge building. Clarence (2021: 82–87) expands on this by describing how teaching and learning benefit from starting with the basics and gradually increasing complexity in order to accumulate knowledge and practice before moving to the next level where knowledge and its application are more complex. She uses a "surfing metaphor" to visualize different wave shapes and sizes that learners can navigate, for example by moving from general definitions (up) to specific cases (down).

LCT Semantics has been applied in a number of disciplines in higher education, including law (Clarence, 2016a), biology (Mouton, 2020), and chemical engineering (Dorfling et al., 2019). In English for Academic Purposes, semantic waves have been used in teaching critical thinking skills in academic writing (Brooke, 2017), and in an English for Specific Academic Purposes (ESAP) context, semantic gravity has been used to help students compose written arguments in line with the expected standards of the discipline of International Relations and Development (Munn, 2021). There are as yet, to our knowledge, no studies using LCT semantics in higher education computing education, though semantic waves are gaining popularity as an approach to teaching computer science at school level. Waite and Sentance (2021) point to the potential of semantic waves as a suitable pedagogical tool for the "concept-rich, yet practically applied subject of computer science" (p. 24), and semantic waves have been used in the teaching of algorithmic thinking to secondary school students (Ritter & Standl, 2023). Given the increasing use of LCT Semantics across disciplines in higher education, its specific application in EAP, and its clear potential for computing education, we sought to explore its potential as a conceptual and empirical tool for investigating knowledge building practices in higher education EMI computing education. The study reported here uses the Semantics dimension of LCT to examine lecturers' knowledge-building practices in real-time classroom discourse. Our aim was to identify and analyze the knowledge-building and meaning-making practices in the discourse as we found them.

Bringing together the different strands in the Semantics dimension of LCT, we formulated the following three research questions to guide our study:

RQ1: How does complexity of meaning (semantic density) vary between and within the two lecturers' practices?
RQ2: How does context-embeddedness (semantic gravity) vary between and within the two lecturers' practices?
RQ3: What semantic profiles emerge in the lecturers' practices? Is there evidence of semantic waves?

3. Methodology

3.1. Context and participants

The study focuses on the practices of two lecturers from the School of Engineering and Telecommunications Systems at a large public technical university in Madrid, Spain. One of the lecturers, who is given the pseudonym Pablo, was delivering a module on the cloud computing service Microsoft Azure for doctoral students, focusing on how they could use it for their research projects. He had two years' experience teaching this module in English. There were 22 students in the group, the majority of whom were Spanish, but with a third of the group being international students from Europe, Latin America, and the Middle East. The students were following doctoral studies in a range of disciplines, including biology, agronomy, and health sciences. This module was taught intensively online in June 2021. The other lecturer, pseudonym Manuel, was teaching a third-year undergraduate module on developing embedded systems with Raspberry Pi, and the recorded sessions focused on the use of the open-source software Git for tracking students' work. At the time of the study, Manuel also had two years of experience teaching various modules in English. The group consisted of 23 students, the majority of whom (15) were pursuing electronics degrees. The others were following computer science programs. The majority of the group were Spanish, but there were five Erasmus students from Eastern Europe (Hungary, Poland and Romania). The module was taught face-to-face in the second semester of the academic year 2020–21.

The modules were selected based on two main features: first, both were taught in English because they were addressed to both

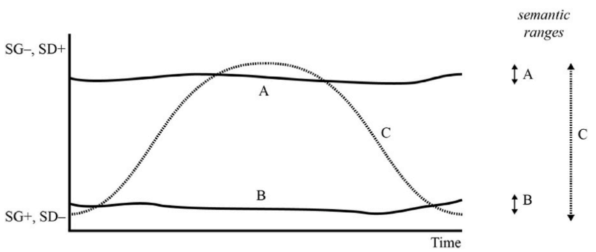

Fig. 1. Semantic profiles (Maton et al., 2016, p. 17).

national and international students; second, they claimed similar aims and followed an analogous teaching structure. In both cases a tool was presented to the students for them to use as part of their research or group work. The teaching sessions varied in length according to the demands of the topic or the working formats, which were either teacher-led explanations or walkthroughs, or periods of individual or group practice on the part of the students. Although it is not the focus of the study, the fact that one course was taught face-to-face and the other online may have had an influence on the lecturers' knowledge-building practices. This is clearly a topic for further research, but our aim here is exploratory, that is, to begin to map out the potential for LCT Semantics to shed light on knowledge-building practices in an EMI context.

3.2. Data

Two sessions from Pablo and three sessions from Manuel were video-recorded and transcribed, and informed consent from participants was obtained. The sampling criteria for the sessions were that a specific topic or chunk of the module content would be introduced and developed. This would provide us with sufficient data to see how complexity and context-embeddedness were dealt with in dealing with a single topic. Table 1 shows the corpus of video data collected.

The variation in length of the sessions can be explained by the fact that the lecturers shared the videos already edited according to the topic presented. These topics did not always coincide with timetabled sessions. Pablo's sessions were delivered synchronously and recorded using Microsoft Teams. On the other hand, Manuel's sessions 1 and 2 took place on the same day, while the second session on Git took place on a different day. The video recordings used for our analysis, as seen in Table 1, were limited to parts of the sessions in which direct teaching occurred. Parts of the sessions where students worked independently on tasks were not included.

3.3. Data analysis

The transcripts were converted to text files and uploaded to UAM CorpusTool, a text annotation software program (O'Donnell, 2021). This program allows segments of data to be tagged using coding schemes developed for the purposes of specific studies. For the current study, we developed and used translation devices (see below) to code segments of the data according to the varying strengths of semantic gravity and density. The length of the segments can vary from a few lines of transcript to much longer stretches of discourse, as this depends on semantic criteria and the shifts in the coding.

Two translation devices were developed to code the data, one for semantic gravity (SG) and one for semantic density (SD), with four levels of strength for each (Tables 2 and 3). For semantic gravity, the translation device distinguishes between four levels of context-embeddedness, from abstract ideas which are not linked to any context (SG-), down to highly grounded situations which relate strongly to the here and now. It should be noted that for semantic gravity the weaker values (SG– and SG- -) are placed at the top, to represent how increasing gravity pulls objects, ideas, and actions closer to the ground. The relation SG/SD is represented inversely in LCT since, put simply, weak semantic gravity usually implies low density and strong semantic gravity implies high density.

To identify different strengths of semantic density, we adapted Maton and Doran's (2017) "generic translation device" which allows for the analysis of relative complexity at the level of words and word combinations. Words are divided into two broad categories: Wording (±technical) from -technical and everyday words to +technical (as in Maton & Doran, 2017). For example, we can distinguish between everyday usage of a term like "application" when it is used in the context of a computer program or a smartphone with limited relationships in terms of epistemological meanings, and technical usage involving a relatively complex constellation of epistemological meanings including design principles or technology architecture. At the level of word-grouping (±condensed) from –condensation to +condensation of meanings, the concern is with the relation of the words with other meanings of terms or concepts.

The transcribed sessions were coded independently by the first author, and then checked by the second author, with small discrepancies resolved by referring to the translation devices. Having completed and agreed on the coding, we used CorpusTool's statistics feature to produce descriptive statistics which allowed us to address the first two research questions. For the third research question, we sampled semantic density and gravity at 5-minute intervals as the sessions unfolded, choosing the mode (the most frequently used coding) to first delimit and then characterize each episode. As regards the 5-minute intervals, the mean number of words is 540 for Pablo (min. 315; max. 672; median 548) and 535 in the case of Manuel (min. 200; max. 659; median 585). Episode boundaries were established by analyzing wave changes in the semantic profile once it had been graphically represented (see results in Section 4) and correlating wave phases with changes in content. In order to further contextualize our coding and data, we selected representative examples of the different knowledge building practices from the corpus for qualitative analysis.

Table 1
Video-recorded data for the study.

Lecturer	Session/topic/length	Session/topic/length	Session/topic/length	Total length	Words
Pablo	(1) Microsoft Azure Sentinel/basics of cloud security (72 min)	(2) Identity, governance, privacy and compliance (33 min)		105 min	11,720
Manuel	(1) Introduction to Raspberry Pi (57 min)	(2) Introduction to version control with Git (20 min)	(3) Version control with Git (26 min)	103 min	13,415

Table 2
Translation device for semantic gravity.

Semantic Gravity (SG)	Indicator	Examples from data
SG –	General features, actions, principles or results not contextualized in the here and now	*A security information management system first collects all the information, then analyzes this information against known threats database in order to detect threats* (Pablo) *All the hardware is more or less quite similar in terms of performance in terms of peripherals, in terms of possibilities* (Manuel)
SG -	Actions are carried out by systems or programs on their own or are not dependent on a programmer, person or specific context	*So here what Microsoft Azure applies is the politic of "denied by default", so everything is closed, everything is denied, no connection can be performed from outside from the Internet* (Pablo) *Okay, because the embedded Linux is the element that is controlling all the application* (Manuel)
SG +	Real world and local context actions or elements slightly detached from the most concrete or physical here and now	*So, if in your company, you have an active directory already deployed, it's very easy to migrate from the local active directory to the Azure, okay* (Pablo) *So, I need to generate an answer or a response to a specific event in a specific time* (Manuel)
SG ++	Real world easily identifiable or even personal contexts; here and now	*And then if you if you take a look here at the network interface, you can see the name of the network interface, okay?* (Pablo) *Okay, we're talking about 1000 euros per component. So, in your computer in the windows the station that you have here you have in this lab, you have the Git client installed on Windows okay?* (Manuel)

Table 3
Translation device for semantic density.

Semantic Density (SD)	Indicator	Examples from data
SD ++	Includes technical vocabulary and complex nominal groups or clauses which need high level processing in non-experts	*This DevOps Starter option here is a very fast way if you want to deploy your own code, and run the code on Azure without taking care about virtual machine operating system installing SDKs or whatever* (Pablo) *The use of a specific real time operating system uses for instance for the case of ARM, the use of CNCS head-to-toes solutions using FDX or using free or RDOS okay. Okay, it's quite useful and in GitHub if you have Mel installed you can invoke Mel using this merge tool when you do that typically you see something like this a graphical tool* (Manuel)
SD +	Includes technical vocabulary and/or nominal groups and acronyms difficult to recognize by non-technical professionals	*In Raspberry Pi version three, we have a Wi-Fi element in the reading all the hardwired that you need to deploy and use the Wi-Fi* (Manuel) *This is the IP of your virtual machine, the public IP, this is the port number - you can leave it like that* (Pablo)
SD -	Technical vocabulary not widely used but easy to recognize on the part of non-technical professionals	*So there was some time ago that the normal attacks were the kinds of denial of service, okay, so the idea was to stop the proper function of your servers* (Pablo) *And additionally, we have different interfaces to connect a display, high resolution displays, or to connect your typical monitor screens to display your high-resolution graphics, okay* (Manuel)
SD –	General words and technical terminology used as everyday language that is widely understood by non-technical professionals	*You just define the rules and then you decide what happens when one of the rules happens okay and one of the of these rules are met* (Pablo) *So for us, the important is that we have a very cheap solution something that is very affordable* (Manuel)

4. Results

The overall results for semantic density (Table 4) show a significantly greater proportion of high semantic density in Manuel's teaching, with a medium effect size for SD++. When SD++ and SD + are combined, Pablo has 20.2% and Manuel has 43.3%. That is, strong semantic density (SD++ and SD+) accounted for 20.2% of all the segments coded for semantic density in Pablo's data, while

Table 4
Semantic density in frequency per 1000 words and percentage distribution for each lecturer.

Semantic density	Pablo		Manuel		Comparison (%)			
	Freq. per 1000 wds	%	Freq. per 1000 wds	%	ChiSqu	P	Signif	Effect Size
SD++	1.1	2.0	6.7	12.4	55.30	0.0000	+++	0.440
SD+	10.2	18.2	16.6	30.9	29.99	0.0000	+++	0.297
SD-	22.5	40.0	11.7	21.8	54.24	0.0000	+++	0.398
SD - -	22.4	39.8	18.7	34.9	3.64	0.0564	+	0.102
TOTAL	**56.2**	**100.0%**	**53.7**	**100.0%**				

this figure was 43.3% in Manuel's case. It can thus be argued that semantic density is twice as high in Manuel's practices. Conversely, at the lower levels of semantic density, Pablo has significantly more examples, especially at SD-, where there is a medium effect size. These results show that Manuel made significantly greater use of technical vocabulary and complex nominal groups or clauses which would require unpacking for non-experts. For example, when discussing constraints and their solutions in developing embedded systems, Manuel used a series of abbreviations and a compound adjective "head-to-toe" which is removed from everyday meaning:

> As soon as you increase the complexity of the system, you move to a use of the operating system and the market now has two different approaches: the use of embedded system solutions is the approach that we are going to use here or the use of a specific real time operating system uses for instance for the case of ARM, the use of CNCs head-to-toe solutions using FDX or using free DoS okay.

This result may be explained by the fact that the students in Manuel's module were expected to develop the embedded systems themselves by writing code. They needed a higher level of technical expertise to do this than the level required by doctoral students to use a software program to protect their data in the cloud. Pablo's practices also show relatively high levels of semantic density as PhD students learning about Microsoft Azure also needed to be able to be familiar with technical terminology, in this case when describing the DevOps Starter option in Azure:

> So it's as easy as going to the DevOps Starter, you select whatever you want to deploy, you have.NET, C, C#, node.JS, PHP, Java, HTML, Ruby, Go, or Python.

Turning to semantic gravity (Table 5), for both lecturers gravity was on the strong side. Combining SG+ and SG ++, Pablo had 72.1% and Manuel had 71%, suggesting a very similar weighting overall for both. However, when broken down into SG + and SG ++, differences between the two lecturers' practices emerge. Manuel has significantly more instances of SG + than Pablo, but the situation is reversed in the case of SG ++, where Pablo has double the amount (29.6% SG++) than Manuel (14.8% SG++). This result is significant, with a low to moderate effect size which suggests that Manuel embedded a considerable proportion of what he was talking about in specific contexts slightly detached from the here and now, whereas Pablo tended to rely more on real-world, easily relatable contexts. As expected from their contexts, while both endeavored to embed the teaching of their topics in context, Pablo had a stronger orientation to keeping what he was doing closer to the ground.

For example, in discussing the concept of "Defense in Depth", Pablo gives the example of mobile phones as a personal possession:

> If you have an Android or an iOS cell phone, they have both this kind of hardware security modules and all your keys, your fingerprints etc.

He also refers to the possible real-life scenario of having one's mobile phone stolen:

> If someone stole your phone and using high level engineering tools get access to your mobile, in the worst case it will get access to your, let's say, to all your documents, pictures, chats etc.

In Manuel's sessions, context embeddedness was often in the form of the use of diagrams on PowerPoint slides, and this was made apparent in the discourse through the use of deictic terms such as 'this' and 'here', as can be seen in these examples:

> So, have a look at the content of this slide later, because again, this is a summary of the different technologies that we have.

> We are going to connect here the accelerometer and the color sensor and we are going to connect here interface and USB to the serial line interface to connect these to a computer.

The Semantics analysis shows that there can be considerable variation in knowledge building practices even within related disciplinary areas with broadly similar theory/practice or pure/applied orientations. Higher semantic density is related to technicality in the sense of meaning condensation in terms, images and abbreviations that are removed from everyday knowledge (Martin, 2020, p. 123). Regarding semantic gravity, it is strengthened in different ways such as providing examples from concrete situations in everyday experience or grounding the discourse in the visible material reality available in the physical surroundings.

We traced the strengths of semantic density and semantic gravity over time separately in order to identify the "semantic profile" of each lecturer's knowledge building practice. All the recorded sessions from each lecturer were treated as a single block of content, in which concepts were introduced and applied over time to show knowledge-building practices in the context of a specific topic from a module. The waves obtained in the semantic profile are evidence that the inverse relation of SD and SG largely used in the LCT

Table 5
Semantic gravity in frequency per 1000 words and percentage distribution for each lecturer.

Semantic gravity	Pablo		Manuel		Comparison (%)			
	Freq. per 1000 wds	%	Freq. per 1000 wds	%	ChiSqu	P	Signif	Effect Size
SG- -	6.3	10.8	1.2	2.6	33.30	0.0000	+++	0.344
SG-	9.9	17.1	11.9	26.4	16.73	0.0000	+++	0.228
SG+	24.6	42.5	25.3	56.2	24.13	0.0000	+++	0.274
SG++	17.2	29.6	6.7	14.8	40.62	0.0000	+++	0.362
TOTAL	**57.9**	**100.0%**	**45.1**	**100.0%**				

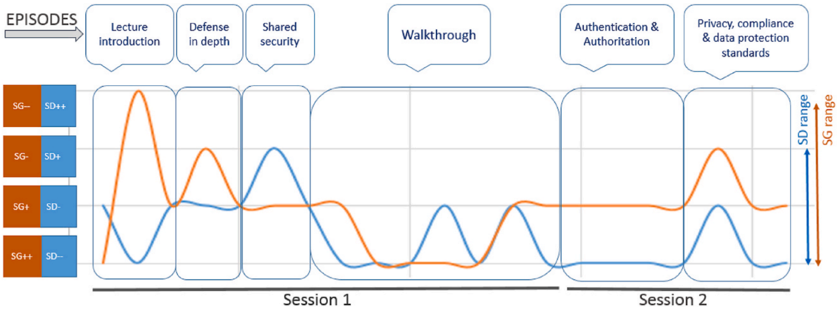

Fig. 2. Semantic profile for Pablo's sessions.

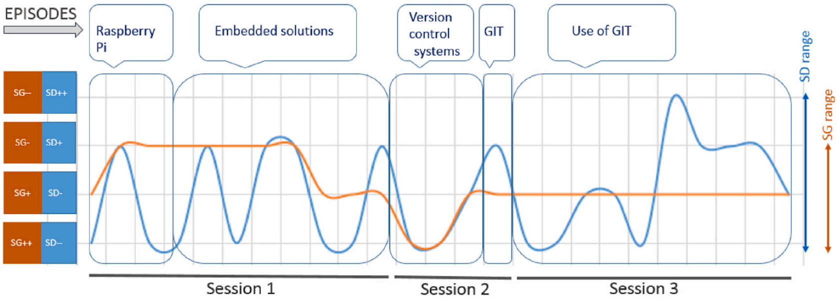

Fig. 3. Semantic profile for Manuel's sessions.

literature to describe pedagogical practices (Maton, 2020, pp. 68–77) does not always apply, especially over a series of sessions. They will more likely vary independently in an undetermined number of occasions (Maton, 2014, p. 131). This can be seen in Figs. 2 and 3 which represent the semantic profile for each lecturer's group of sessions.

Pablo's session starts with very weak semantic gravity. Here, he talked about the importance in general of security in the cloud and how it was important for different situations. As the sessions progressed, semantic gravity gradually strengthened, reaching its strongest point in the walkthrough episode where the activity was firmly grounded in the immediate context. Towards the end of this episode, there was a weakening of semantic gravity and a small strengthening of semantic density as Pablo repacked the practical work in terms of more general standards for data protection. As was seen in the statistics reported above, semantic density in Pablo's sessions was on the low side, and after a small peak in the episode of shared security, it never moved above SD-. Noticeably, during the long period in middle of the two sessions, the episodes of shared security and the walkthrough, semantic density shifted from quite high (SD+) in shared security to overall low (SD-, SD - -) in the walkthrough, which indicates that this was a practical activity which did not require the use of terms with high condensation of meaning.

For Manuel, sematic gravity was fairly weak until the middle of the episode on embedded solutions, and from there it remained relatively strong until the end of the sessions. This reflects the fact that Manuel talked a lot about Raspberry Pi, its origins and uses, and the nature of embedded systems at a quite general level for most of the first session. As the sessions progressed, however, the discourse became more grounded in objects and images visible in the shared physical space and on specific actions such as the use of Git to track work. Semantic density throughout the sessions had a "ripple" effect with shifts from low to fairly high, as quite condensed meanings in technical terms related to developing embedded systems were interspersed with more transparent meanings which would not require unpacking by non-experts, as can be seen in this example:

> And of course, we have here different other peripherals that are very useful to develop applications (SD-). Of course, we have this connection to connect your graphical output, we have the option to connect a camera, we have an audio output (SD+), we have a micro-SD connector to insert the SD card that we will use to boot our operating system solution (SD++). And finally, we have here a connector, a 40 pins connector (…) to connect different GPIO elements I squared C chips, SPI element serial line devices (SD++). We will use that in our lab okay to connect peripherals and connect sensors to implement application, okay (SD-).

The analyses here point to the practical nature of both courses, as semantic gravity was relatively strong most of the time, with some more "waving" in semantic density. There was no evidence of a "high semantic flatline" in which highly abstract ideas and concepts were dealt with in isolation from specific contexts of practice. Neither were participants "stuck" at the level of specific examples or procedures, as semantic density oscillated between more and less condensed meanings, without reaching the highest levels of condensation for any sustained period of time.

In order to take a closer look at the context in which the lecturers enact their knowledge building practices, we analyze one extract from each lecturer in which shifts between different strengths of semantic density and gravity can be seen at a more granular level. The extract from Pablo's module (Extract 1) comes from the second session, where the purpose was to review the different identity services that are available in the Microsoft Azure platform, including governance features, privacy solutions and compliance agreements. In the extract, Pablo is explaining the difference between the concepts of authentication and authorization.

Extract 1: Authentication and authorization.

So, authentication and authorization (SD+SG-), these are two words
that maybe you know (SD+/SG++), but apply it to IT security, it has,
they have different meanings (SD+/SG-) with authentication we control
who is accessing any asset okay (SD+/SG-). So, with the
authentication service, we try to identify the person, the client, it could
be also a machine, okay, that is asking access to a specific resource.
(SD+/SG-). We can do that using access credentials, for example, a
username and a password and this is the basic feature to start
controlling the access to our resources okay. (SD-/SG+). Once we have
identified
who is trying to access, if that person that server or whatever is Access
Granted, then we can also control what this this entity is allowed to do
in our - to perform in our servers in our services, okay. (SD+/SG+). So,
with authentication, we control who accesses it and with authorization,
we control what this person is able to, to do in the in the servers. (SD-
/SG-). These two services authentication and authorization are the
basics to control or to manage the roles that the all the persons in our company or external like clients or users etc. are allowed to do in our services. (SD+/SG-)

In Extract 1, line 1, Pablo links the two terms to students' existing knowledge of the general meaning of the words outside the field of IT security. He highlights that, although the students may be familiar with their more general meanings, they are different when applied in this specific field. Semantic density is kept relatively strong (SD+) as these words will have more condensed meanings when applied to the field of IT security than with more general applications, although they do not reach the highest levels of technicality. Semantic gravity is relatively weak, apart from the brief dip when he relates the ideas to what they already know, as the concepts are not yet tied down to any specific context or example (SG-). From line 6, semantic density is lowered (SD-), as Pablo introduces "access credential" and unpacks it by giving examples of more transparent and familiar terms with less condensed meanings such as "username" or "password". Semantic gravity is strengthened (SG+) as he focuses on actual things you can use if you want to restrict who can get access to your resources. From line 8, he moves from authentication to authorization, without using the latter term. He clarifies what it is, and it seems he expects the students to make the link to the second term, when this arrives as a reformulation in which both terms are glossed. Semantic density is relatively high (SD+), as he uses the term "access granted" and links this to actions performed on servers. Semantic gravity remains relatively strong (SG+) as he is still focusing on what can be done to control access – there is a context, and that is that some entity may be trying to access your resources, and these are things you can do to prevent it. From line 12, both semantic density and gravity are weakened (SD-/SG-) as he gives a more general, not context-embedded summary of the meanings of the two terms. Meanings are not very condensed, and no specific context of use is appealed to. From line 14, semantic density is strengthened (SD+) as more condensed terms relating to internal and external organizational roles are used, but still without reference to any specific context of use (SG-).

Extract 2 comes from Manuel's third session, in which he is taking students through the steps in using Git to track each other's work in developing code for their embedded systems projects using Raspberry Pi. In this sequence, he is explaining how to use the branching and merging functions in Git to avoid conflicts when two developers are working on code together.

Extract 2: Git branching and merging.

So this means that if you create a branch you switch to that branch
and you perform the modifications in that branch.
Once you check that that part of the code is working your colleague
can bring that changes to their branch OK?
And this is something that we typically call merge. (SD+/SG+)
Merge is like we are mixing the things okay?
In principle if you are developing one thing and your colleague is
developing another thing when you perform the merge typically
there are no problems or issues.
And the point here is that two developers are doing the same thing
in that case we have the typical thing that we call conflicts.
OK this is the thing that you should avoid.
So the idea is that at the end of the course when you are
developing the main project of the course (SD-/SG+), you use this
idea of branches to more or less organise the development of the code (SD+/SG+).

In Extract 2, semantic density starts off relatively high as the terms with condensed meanings ("branch" and "merge") are used (lines 1–5). "Branch", of course, has everyday meanings that the students could be expected to be familiar with, although Manuel does not appeal to these differences as shown in Pablo's example above. Here it has condensed meaning (SD+) which can only be interpreted in relation to knowledge of Git, and he assumes familiarity with it. However, in line 5, he introduces the term "merge", which can also have meanings outside the current context of computing. In this case, he does draw attention to word meaning, by explicitly referring to what a process he has just described is "called". In this sense, he is producing an expansion reformulation in the form of an explanation (Hyland, 2007, p. 274), by using "merge" as a technical term for the idea he has already conveyed in a more accessible way, albeit in somewhat non-standard grammar: "bring that changes to their branch OK?" From line 6, semantic density is lowered (SD-), as Manuel "unpacks" the term "merge" with less condensed meaning: "Merge is like we are mixing the things okay?" Semantic density rises towards the end of the extract as he returns to the idea of organizing the development of the code.

Throughout Extract 2, semantic gravity is held relatively strong (SG+) as terms are embedded in the specific context of using Git when working with partners to develop code. The relatively strong semantic gravity is reflected in language use as there is quite a high degree of "iconicity" (Martin & Matruglio, 2020, p. 107). There is congruency between the linguistic forms and the aspects of reality they construe. In the following example, participants are realized as nouns, and processes as verbs: "… if you are developing one thing and your colleague is developing another thing when you perform the merge typically there are no problems or issues".

5. Discussion

Turning again to the research questions which guided this study, in this section we present a brief discussion of the findings. The first research question inquired as to the variation in the complexity of meaning (semantic density) between and within the two lecturers' practices. The results show variation in semantic density in the lecturers' discourse, both from a synoptic perspective when comparing them at an overall level, and in terms of phases of the sessions as they unfold over time. As Maton and Doran (2017) point out, by analyzing the complexity of knowledge practices, we can understand the bases of achievement in different educational contexts, what they describe as the "rules of the game" (p. 17). The results of this study show that there was more complexity in the knowledge practices in the undergraduate module than in the postgraduate one. Although this result may seem counter-intuitive if we take the educational levels (undergraduate and postgraduate) at face value, it makes sense when we consider the content and aims of the two courses. In this sense, it is logical that an introductory course for non-experts, as is Pablo's at the postgraduate level, aims to reduce complexity. On the other hand, Manuel's module is for telecommunications engineers in the third year of their degree, and, as such, is expected to be more technical and, therefore, more complex in terms of content.

The second research question asked about variation in context-embeddedness (semantic gravity) between and within the two lecturers' practices. The results suggest that, while both lecturers strengthened semantic gravity by embedding knowledge in specific contexts, they did so in different ways - either by invoking real-life scenarios (Pablo) or emphasizing the physical materiality of the surroundings (Manuel). Strengthening or weakening semantic gravity can be related in social semiotic terms to "presence" (Hood, 2020; Martin & Matruglio, 2020). When presence is stronger, language more iconically represents phenomena in reality, that is, verbs represent actions and nouns represent things. Furthermore, knowledge may be seen as more negotiable, not simply "right" or "wrong" and its objects may be referred to more implicitly, for example by using deictic expressions such as "this" and "here" to refer to things present in the shared physical surroundings. The results as seen in the short extracts show evidence of more iconic and implicit language, particularly by Manuel. As Hood (2020) shows, lecturers use a wide range of semantic resources for modulating the strength of semantic gravity by increasing or decreasing "presence", and the results of this study provide evidence of these lecturers' ability to do so through English.

The third research question addressed the issue of the semantic profiles emerging in the lecturers' practices, with a particular interest in semantic waves. The findings show evidence of "waving" in both lecturers' knowledge-building practices. For both lecturers, semantic density had a rather choppy wave profile. This is more evident in Manuel's case, which shows how they "unpacked" and "repacked" concepts as they moved through the sessions. The evidence of "waving" suggests that both lecturers were working effectively to build knowledge cumulatively over time rather than leaving their students stuck on a "high flatline" of very dense material, or a "low flatline" of mostly everyday concepts unlinked to principles or theory. In this sense, it is important not to characterize semantic gravity and density as static constructs which can typify a body of knowledge as "complex" or "context-embedded", but as dynamic concepts in constant movement, with educators displaying pedagogic skills in how they manage upwards and downwards shifts as they cumulatively build knowledge through their discourse. These shifts can be seen over longer periods of time, as in the analysis of whole sessions or sequences of sessions (Figs. 2 and 3), or in a more fractal way in shorter episodes, as in Extracts 1 and 2. Shifts in semantic density and gravity can thus be seen as important both in planning for instruction, and in instruction as it is enacted in real time.

6. Conclusion

The study reported here uses the Semantics dimension of LCT to examine two lecturers' knowledge-building practices in real-time classroom discourse. Our three main research questions concerned: (RQ1) How complexity of meaning (semantic density) varies between and within the two lecturers' practices; (RQ2) How context-embeddedness (semantic gravity) varies between and within the two lecturers' practices; and (RQ3) The semantic profiles emerging in the lecturers' practices, especially if there is evidence of semantic waves. The results show that there is variation in semantic density in the lecturers' discourse, and that both lecturers strengthened semantic gravity by embedding knowledge in specific contexts, although they do so in different ways. The results also confirm that there is "waving" in both lecturers' knowledge-building practices.

A possible limitation of the study is that the difference in delivery format (online and face-to-face) is not taken into account in the analysis. It would be a very interesting avenue for further research to explore the constraints and affordances of these delivery formats when it comes to dealing with complexity and context-embeddedness in EMI contexts. Another limitation of the study is that it draws on a relatively small amount of data, and thus the findings have to be interpreted with caution.

Our aim was to identify and analyze the knowledge-building and meaning-making practices in the discourse as we found them. Therefore, the Semantics dimension of LCT was not used in the planning of the sessions as part of any kind of intervention. However, as an exploratory study, it can open up theoretical and methodological options for linking EMI lecturers' knowledge-building practices with the communicative resources used to enact them. As such, the study has some implications and possible applications for lecturer professional development in EMI contexts and further studies could take a more interventionist approach by introducing lecturers to LCT Semantics.

The dimension of Semantics can be presented as a practical tool for reflecting on their own classroom discourse practices, and guiding their teacher talk in preparation for teaching (Brooke, 2020). Specific software tools for video-based teacher professional development such as the VEO app (Video Enhanced Observation) (Miller & Haines, 2021) could be used, drawing on the possibilities of "tagging" classroom practices recorded on video. The tags for these activities can be drawn from the Semantics dimension of LCT, and tagged videos can be used as a basis for lecturers' reflection on their practices, most profitably in small professional development learning communities.

A possible benefit of such an approach based on the Semantics dimension of LCT is that it engages with practitioners at the level of their disciplinary knowledge practices. As Clarence (2016b) points out, LCT Semantics can help to open up more "pointed", focused and specific conversations with practitioners about their disciplinary knowledge practices. In the context of EMI, such conversations can be a starting point to enable lecturers to reflect on their language and communicative practices in the context of how they deal with complex meanings and embed them in contexts. Lecturers can reflect on specific subject-related concepts, objects, principles, and from there it is possible to look at how these are introduced, exemplified and developed in ongoing instruction. LCT Semantics can focus on very specific resources of English required for meaning making in disciplinary topics. For example, lecturers can be alerted to the ideas of "unpacking" and "repacking", in which, for example, complex nominalizations with dense meanings are "opened up", with their participants and processes in more "congruent" ways, possibly with real-world examples. Using the Semantics dimension of LCT can complement already existing approaches to professional development in EMI by highlighting the relationships between disciplinary knowledge-building practices and the communicative resources needed to enact them. In this sense, the Semantics dimension of LCT can be an "emerging zone" (Dafouz, 2021) which can help to bridge disciplinary boundaries between practitioners and researchers in EAP and EMI, and between both fields and other subject specialists.

Authorship statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the *Journal of English for Academic Purposes*.

Acknowledgements

This research has been funded by the Dirección General de Investigación e Innovación Tecnológica (Comunidad de Madrid) and the Universidad Autónoma de Madrid (SI2/PBG/2020-00009: Integrating language and knowledge building through video-enhanced observation in EMI).

References

Bernstein, B. (1990). *The structuring of pedagogic discourse* (Vol. 4). London: Routledge & Keegan Paul.
Bernstein, B. (2000). *Pedagogy, symbolic control and identity*. Oxford, England: Rowman & Littlefield Publishers, Inc.
Brooke, M. (2017). Using 'semantic waves' to guide students through the research process: From adopting a stance to sound cohesive academic writing. *Asian Journal of the Scholarship of Teaching and Learning, 7*(1), 37–66.
Brooke, M. (2020). Guiding teacher talk in the content and language integrated learning classroom using semantics from legitimation code theory. *Teaching in Higher Education, 25*(7), 812–824.
Clarence, S. (2016a). Surfing the waves of learning: Enacting a semantics analysis of teaching in a first-year law course. *Higher Education Research and Development, 36*(5), 920–933.
Clarence, S. (2016b). Exploring the nature of disciplinary teaching and learning using Legitimation Code Theory Semantics. *Teaching in Higher Education, 21*(2), 123–137.
Clarence, S. (2021). *Turning access into success: Improving university education with legitimation code theory*. London: Routledge.
Dafouz, E. (2021). Crossing disciplinary boundaries: English-Medium education (EME) meets English for specific purposes (ESP). *Iberica, 41*, 13–38.
Dorfling, C., Wolff, K., & Akdogan, G. (2019). Expanding the semantic range to enable meaningful real-world application in chemical engineering. *South African Journal of Higher Education, 33*(1), 42–58.
Hood, S. (2020). Live lectures: The significance of presence in building disciplinary knowledge. In J. R. Martin, K. Maton, & Y. J. Doran (Eds.), *Accessing academic discourse: Systemic functional linguistics and legitimation code theory* (pp. 211–235). London: Routledge.
Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics, 28*(2), 266–285.
Johnson, R., & Bourdieu, P. (1993). *The field of cultural production: Essays on art and literature*. Cambridge: Polity.

Macaro, E., & Aizawa, I. (2022). Who owns English medium instruction? *Journal of Multilingual and Multicultural Development*. https://doi.org/10.1080/01434632.2022.2136187
Martin, J. R. (2020). Revisiting field: Specialized knowledge in secondary school science and humanities discourse. In J. R. Martin, K. Maton, & Y. J. Doran (Eds.), *Accessing academic discourse: Systemic functional linguistics and legitimation code theory* (pp. 114–147). London: Routledge.
Martin, J. R., & Matruglio, E. (2020). Revisiting mode: Context in/dependency in Ancient History classroom discourse. In J. R. Martin, K. Maton, & Y. J. Doran (Eds.), *Accessing academic discourse: Systemic functional linguistics and legitimation code theory* (pp. 89–113). London: Routledge.
Maton, K. (2014). *Knowledge and knowers: Towards a realist sociology of education*. London: Routledge.
Maton, K. (2016). Legitimation code theory: Building knowledge about knowledge building. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in legitimation code theory* (pp. 1–23). London: Routledge.
Maton, K. (2020). Semantic waves: Context, complexity and academic discourse. In J. R. Martin, K. Maton, & Y. J. Doran (Eds.), *Accessing academic discourse: Systemic functional linguistics and legitimation code theory* (pp. 59–85). London: Routledge.
Maton, K., & Chen, R. T. H. (2016). LCT in qualitative research: Creating a translation device for studying constructivist pedagogy. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in legitimation code theory* (pp. 27–48). London: Routledge.
Maton, K., & Doran, Y. J. (2017). Semantic density: A translation device for revealing complexity of knowledge practices in discourse, part 1—wording. In *Onomázein, Número especial LSF y TCL sobre educación y conocimiento* (pp. 46–76).
Maton, K., Hood, S., & Shay, S. (2016). *Knowledge-building: Educational studies in legitimation code theory*. London: Routledge.
Miller, P., & Haines, J. (2021). From teaching to learning: The development of the VEO app. In P. Seedhouse (Ed.), *Video enhanced observation for language teaching: Reflection and professional* (pp. 45–60). London: Bloomsbury Academic.
Moncada-Comas, B., & Block, D. (2021). CLIL-ised EMI in practice: Issues arising. *Language Learning Journal, 49*(6), 686–698.
Mouton, M. (2020). A case for project-based learning to enact semantic waves: Towards cumulative knowledge building. *Journal of Biological Education, 54*(4), 363–380.
Munn, D. (2021). Learning transfer from an in-sessional ESAP module to International Relations and Development students' core disciplinary writing assignments: Teaching semantic gravity to support cumulative knowledge building. *Journal of Academic Language and Learning, 15*(1), 187–213.
O'Donnell, M. (2021). UAM CorpusTool (Version 3.3) [Computer software]. http://www.corpustool.com/index.html.
Ritter, R., & Standl, B. (2023). Promoting student competencies in informatics education by combining semantic waves and algorithmic thinking. *Informatics in Education, 22*(1), 141–160. https://doi.org/10.15388/infedu.2023.07
Waite, J., & Sentance, S. (2021). *Teaching programming in schools: A review of approaches and strategies*. Raspberry Pi Foundation Research Report.
Wingate, U., & Hakim, A. (2022). Moving beyond 'infancy': Towards a cross-fertilization between EMI and EAP scholarship. *ELT Journal, 76*(4), 529–537.

Irina Argüelles-Álvarez is a Professor in the Department of Linguistics Applied to Science and Technology at the Universidad Politécnica de Madrid. Her research focuses on language teaching, e-material development and English as a Medium of Instruction in the context of higher education.

Tom Morton is Beatriz Galindo Distinguished Research Fellow at the Universidad Autónoma de Madrid, Spain, where he is a member of the UAM-CLIL Research Group. His research focuses mainly on content and language integration and teacher knowledge and identity in CLIL and EMI contexts.