

1. Rúbrica Longitudinal — Protocolo IA-Socrático USACH

Versión: 1.7 Fecha: Junio 2026 Instrumento de análisis para el investigador

Cambios v1.6 → v1.7 (refactor de notación — IP, 2026-06-16): - Convención de notación: las decisiones de diseño pasan de D1–D39 a DD_N (DD_1–DD_39) en todo el corpus, para eliminar la colisión con las dimensiones D1–D4 de esta rúbrica. Las dimensiones D1–D4 NO cambian. Nota de nomenclatura (§1) actualizada. Ver doc_inv_Glosario_Conceptual §“Convención de notación”.

Cambios v1.5 → v1.6 (revisión SFL — Ingrid Westhoff Podestá, 2026-05-31): - Agregado Anexo A — “Fundamentos SFL para codificación”. Incluye: (A.1) mapa de metafunciones → dimensiones, (A.2) glosario operacional de 10 recursos lingüísticos con definiciones, ejemplos positivos del dominio del proyecto, ejemplos que no califican, y dimensión asociada, y (A.3) instrucciones de codificación para IA con procedimiento de 7 pasos y principio de codificación conservadora. Diseñado como referencia autónoma para LLMs y codificadores humanos sin conocimiento previo de SFL.

Cambios v1.4 → v1.5 (revisión SFL — Ingrid Westhoff Podestá, 2026-05-31): - §2: Agregada columna “Criterio de codificación lingüística” en las 4 tablas D1–D4. Cada nivel ahora incluye un criterio observable basado en recursos lingüísticos (nexos causales, participantes con modificación numérica, marcadores evidenciales, cláusulas condicionales). Fundamentado en doc_inv_SFL_Análisis_v1.1.md §2.2 y §2.4. - §3: Agregada tabla “Criterios de codificación lingüística por clase” que traduce las preguntas conceptuales de cada dimensión a qué buscar en el texto del estudiante (umbrales numéricos, tipos de marcadores, estructuras esperadas). - §4: Agregada nota metodológica sobre asimetría de géneros entre M1 (género cerrado, 9 secciones) y M4 (género abierto, respuesta libre). Declara que la diferencia de exigencia estructural es una limitación que debe reportarse en el paper. Fundamentado en doc_inv_SFL_Análisis_v1.1.md §4. - §6: Agregado Paso 0 — Calibración de codificadores (pre-piloto). Define procedimiento de 5 pasos con textos simulados, clasificación de desacuerdos (observación vs. criterio), umbral $\kappa \geq 0.85$ para aprobar calibración, y documentación de casos de borde como anexo metodológico.

Cambios v1.3 → v1.4 (inspección IP 2026-05-30): - Umbral de Cohen’s κ unificado: $\kappa \geq 0.80$ para reporte publicable (ponderado, lineal/cuadrático). $\kappa \geq 0.70$ solo exploratorio. Alineado con MarcoMetodológico v1.5.

Cambios v1.2 → v1.3 (auditoría IP 2026-05-27, decisión metodológica): - Agregado M4 — Cierre (reflexión DD_30) como momento independiente de medición. - Δ_{intra} redefinido: M4 - M1 (cierre vs. rastro), alineado con DD_30 de Premisas v1.0. M3 (evaluación BUILD) se mantiene como indicador independiente de evaluación crítica (D3/DD_27). - Plantilla de registro: agregada columna M4 y filas Δ_{intra} para C1 y C5 (antes solo C2–C4). - C5: agregado cierre breve (min 75–80) para habilitar Δ_{intra} comparable con C1–C4.

Cambios v1.1 → v1.2 (auditoría contra Premisas de Diseño v1.0, 2026-05-19): - Tabla de aplicación por clase: agregado entregable BUILD específico que evalúa el alumno

(DD_13). - Escalación de errores DD_8 nombrada explícitamente (C2 obvios, C3 sutiles, C4 profesionales). - Nuevo indicador observable: calidad de defensa del alumno ante contraargumentación BUILD (DD_27).

Cambios v1.0 → v1.1 (alineación con Premisas de Diseño v1.0, 2026-05-16): - D4 aplica en TODAS las clases (C2 tiene decisión no-binaria, C3 tiene decisión de intervención). - Momentos de codificación actualizados: Papel → Post-PLAN → Post-BUILD (DD_5, DD_24). - Agregado rol de AGENT_ANALISTA_SFL como codificador primario en tiempo real (DD_9, DD_16). - Plantilla de registro actualizada a estructura v2.0.

1.1 1. Dimensiones de análisis

Nota de nomenclatura: D1–D4 en este documento son **dimensiones de análisis del razonamiento** (qué se mide en el texto del estudiante). No confundir con DD_1–DD_39 de doc_inv_Premisas_Diseño, que son **decisiones de arquitectura del protocolo** (qué se diseña y por qué; DD = Decisión de Diseño). Desde 2026-06-16 las dos notaciones son distintas (D1–D4 para dimensiones, DD_N para decisiones) para eliminar la colisión que existía cuando ambas usaban D + número.

Cód.	Dimensión	Descripción operacional	Clases donde aplica
D1	Complejidad causal	¿El estudiante identifica síntomas o mecanismos? ¿Menciona cadenas causales de más de un eslabón?	1, 2, 3, 4, 5
D2	Especificidad técnica	¿Las variables mencionadas son medibles? ¿El lenguaje es operacional o vago?	1, 2, 3, 4, 5
D3	Consciencia epistémica	¿Distingue lo que sabe de lo que supone? ¿Reconoce vacíos en su información?	1, 2, 3, 4, 5
D4	Decisión bajo incertidumbre	¿Toma una posición explícita o evade? ¿Nombra el riesgo que asume al decidir?	1, 2, 3, 4, 5

1.2 2. Niveles de desempeño por dimensión

1.2.1 D1 — Complejidad causal

Nivel	Descriptor	Indicadores observables	Criterio de codificación lingüística
1 — Des-criptivo	Menciona observaciones sin relación causal	“El agua está turbia”, “el pH subió”	Solo procesos relacionales (“es”, “está”, “tiene”) o materiales simples (“subió”, “bajó”). Cero nexos causales explícitos entre cláusulas.
2 — Causal simple	Establece una relación causa-efecto de un eslabón	“El pH alto causa turbidez”	Exactamente 1 nexo causal explícito (“porque”, “entonces”, “debido a”). Relación lineal $A \rightarrow B$.
3 — Causal múltiple	Identifica cadenas de 2+ eslabones o causas paralelas	“La temperatura sube \rightarrow pH sube \rightarrow eficiencia del cloro baja \rightarrow ORP cae \rightarrow proliferación microbiológica”	≥ 2 nexos causales explícitos encadenados o en paralelo. Al menos una nominalización como participante de proceso (“la caída de presión”).
4 — Sistémico	Integra subsistemas, retroalimentaciones o condiciones de borde	“El filtro obstruido reduce caudal, lo que aumenta tiempo de residencia del agua caliente, acelerando la degradación química”	Combinación de ≥ 2 tipos de relaciones lógico-semánticas: realce causal + elaboración (“lo que aumenta...”) o extensión (“y además...”). Nominalizaciones como participantes en múltiples procesos.

1.2.2 D2 — Especificidad técnica

Nivel	Descriptor	Indicadores observables	Criterio de codificación lingüística
1 — Vago	Lenguaje genérico, sin variables medibles	“Algo anda mal con el sistema”	Participantes indeterminados (“algo”, “el sistema”, “la cosa”). Procesos con verbos generales (“funciona”, “anda”). Sin variables técnicas nombradas.
2 — Nominal	Nombra componentes o variables sin valores	“El filtro puede ser el problema”	Participantes técnicos por nombre (“pH”, “filtro”, “bomba”) sin modificación numérica ni unidades.

Nivel	Descriptor	Indicadores observables	Criterio de codificación lingüística
3 — Opera- cional	Usa variables con valores y unidades	“El ΔP pasó de 12 a 38 kPa en 14 horas”	≥ 3 participantes técnicos con valor numérico + unidad (“28 kPa”, “7.8 pH”, “72 m ³ /h”).
4 — Diag- nóstico	Usa umbrales técnicos para clasificar estados	“ORP bajo 650 mV indica desinfección comprometida; pH sobre 7.8 reduce eficiencia del cloro libre en ~80%”	≥ 1 participante técnico con umbral de clasificación explícito (“ORP bajo 650 mV indica...”). Lenguaje metalingüístico que clasifica el estado del sistema.

1.2.3 D3 — Consciencia epistémica

Nivel	Descriptor	Indicadores observables	Criterio de codificación lingüística
1 — Ausente	No distingue supuestos de certezas	Afirma diagnóstico sin señalar incertidumbre	Solo declaraciones asertivas sin modalización. Cero operadores modales (“podría”, “quizás”), marcadores evidenciales (“según”, “los datos muestran”) o cláusulas condicionales.
2 — Implí- cita	Usa lenguaje hedged (“probable- mente”, “podría ser”)	“Podría ser el filtro” sin explicar por qué no está seguro	≥ 1 operador modal epistémico (“probablemente”, “podría ser”) o expresión de duda, pero sin marcador de fuente que explique la incertidumbre.
3 — Explí- cita	Nombra datos que faltan y supuestos que asume	“Asumo que no hay fuga; necesitaría el dato de cloro libre residual para confirmar”	≥ 1 marcador evidencial (“según los datos”, “asumo que”) + ≥ 1 declaración de vacío epistémico (“no tengo información suficiente”, “necesitaría”).
4 — Meta- cogni- tiva	Evalúa la calidad de su propia evidencia y los límites de su análisis	“Mi hipótesis depende de que el sistema no tenga fuga. Si la tuviera, el caudal decreciente tendría otra explicación”	≥ 1 cláusula condicional epistémica (“si tuviera X, podría Y”) + evaluación de los límites de su propio análisis (“mi hipótesis depende de que...”).

1.2.4 D4 — Decisión bajo incertidumbre

Nivel	Descriptor	Indicadores observables	Criterio de codificación lingüística
1 — Eva- sión	No toma posición; pide más datos sin decidir	“Necesito más información antes de recomendar”	Sin verbos de decisión en primera persona (“recomiendo”, “decido”, “propongo”). Predominan actos de habla de opinión o peticiones de datos sin decisión.
2 — Deci- sión sin riesgo	Decide pero no nombra consecuencias de error	“Recomiendo retrolavado” (sin análisis de qué pasa si no funciona)	Verbo de decisión en primera persona, pero sin cláusula condicional de consecuencia.
3 — Deci- sión con riesgo	Decide y nombra explícitamente el riesgo asumido	“Recomiendo retrolavado. Si no resuelve en 2h, el riesgo es exposición a agua con ORP < 500 mV”	Verbo de decisión + ≥1 cláusula condicional que nombra una consecuencia negativa (“si no resuelve, el riesgo es...”).
4 — Deci- sión contin- gente	Decide, nombra riesgo, define criterio de escalamiento y plan B	“Retrolavado inmediato. Si ORP no sube a 650+ en 2h: cerrar sistema y dosificar shock de hipoclorito. Riesgo de demora: exposición sanitaria”	Verbo de decisión + cláusula condicional + criterio de escalamiento (“si X no sube a Y en Z horas: entonces W”) + referencia a destinatario o autoridad.

1.3 3. Aplicación por clase (trayectoria longitudinal)

1.3.1 Entregable BUILD que evalúa el alumno en cada clase (DD_13)

Clase	Entregable BUILD	Escalación de errores (DD_8)
C1	— (solo PLAN socrático básico)	—
C2	Informe técnico de diagnóstico	Obvios — fácilmente detectables con conocimiento básico

Clase	Entregable BUILD	Escalación de errores (DD_8)
C3	Reporte de análisis de tendencias SCADA	Sutiles — requieren conocimiento técnico específico
C4	Protocolo de emergencia + análisis económico	Profesionales — parecen correctos, requieren criterio experto
C5	— (chatbot NEUTRO, no genera documento)	—

Nota DD_27: Cuando el alumno señala un error, el chatbot BUILD **defiende su posición con contraargumentos técnicos**. La calidad de la defensa del alumno ante esta contraargumentación es un indicador observable adicional (ver sección 5).

1.3.2 Tabla de dimensiones por clase

Dimensión	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
D1	Línea base: ¿identifica causas o solo síntomas?	¿Discrimina entre hipótesis competidoras (mecanismo, no solo correlación)?	¿Construye cadenas multicausales distinguiendo señal de ruido en datos ambiguos?	¿La causalidad sostiene la decisión bajo presión adversarial?	¿Aplica razonamiento causal a caso nuevo sin guía?
D2	Línea base: ¿usa variables medibles?	¿Usa las 6 variables con valores y unidades para evaluar hipótesis?	¿Usa umbrales cuantitativos para clasificar alarmas (real vs. falsa)?	¿Parámetros de control con valores específicos y justificados?	¿Nombra variables del caso nuevo con precisión?

Dimensión	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
D3	Línea base: ¿sabe qué no sabe?	¿Reconoce que la intervención del operador introduce incertidum- bre? ¿Detecta errores en el informe BUILD?	¿Distingue sensor sin calibrar, sesgo del operador y señal real? ¿Cuestiona el reporte BUILD?	¿Nombra modos de falla no cubiertos y limitaciones de su diseño? ¿Detecta omisiones en protocolo BUILD?	¿Identifica vacíos sin que nadie los señale?
D4	Decisión preliminar (sin presión)	Decisión no-binaria con trade-offs (abrir/cerrar/restringir) + evaluación crítica del informe BUILD	Decisión de intervención basada en datos (¿cuándo actuar con incertidum- bre?)	Decisión defendible bajo presión adversarial + restricciones (presupues- to, personal, política)	Recomendación autónoma completa

1.3.3 Criterios de codificación lingüística por clase

La tabla anterior describe qué mide conceptualmente cada dimensión en cada clase. Esta segunda tabla traduce esas preguntas conceptuales a **qué buscar en el texto del estudiante** durante la codificación. Los criterios son acumulativos: el nivel esperado en C3 presupone el de C2.

	C1 — qué buscar en el texto	C2 — qué buscar en el texto	C3 — qué buscar en el texto	C4 — qué buscar en el texto	C5 — qué buscar en el texto
D1	Presencia de ≥ 1 nexo causal explícito vs. solo procesos descriptivos.	≥ 2 nexos causales + marcadores de contraste entre hipótesis (“por un lado... por otro”, “a diferencia de”).	Cadena causal con ≥ 3 eslabones + distinción léxica entre señal y ruido (“correlaciona pero no causa”, “es un artefacto del sensor”).	Cadena causal que articula múltiples restricciones (técnica, presupuestaria, humana, política) en una decisión unificada.	Mismos criterios que C4, producidos sin presión del chatbot.
D2	≥ 1 variable nombrada con valor numérico + unidad.	≥ 4 de las 6 variables del caso con valor + unidad, usadas para evaluar hipótesis.	Variables con umbrales de clasificación (normal/anormal) + al menos una variable cuestionada por calidad del sensor.	Parámetros de control con valores específicos y justificación numérica (no solo nombrados).	Mismos criterios que C4, aplicados a variables del caso nuevo.
D3	Presencia de ≥ 1 marcador de duda o vacío epistémico.	≥ 1 marcador evidencial que referencia la intervención del operador Muñoz como fuente de incertidumbre.	Distinción léxica explícita entre dato confiable (“el sensor de caudal está calibrado”) y dato dudoso (“el pH podría tener offset”).	Declaración de ≥ 2 modos de falla no cubiertos por el protocolo + limitaciones del diseño propio.	Identificación de vacíos epistémicos sin que el chatbot los señale.

	C1 — qué buscar en el texto	C2 — qué buscar en el texto	C3 — qué buscar en el texto	C4 — qué buscar en el texto	C5 — qué buscar en el texto
D4	≥1 verbo de decisión en primera persona (aunque sea preliminar).	Decisión no-binaria con ≥1 trade-off explícito ("abrir parcialmente", "restringir hasta confirmar").	Decisión con estructura condicional ("si el pH supera X, entonces intervenir; si no, monitorear").	Decisión + riesgo + criterio de escalamiento + acto de habla de comunicación a autoridad ("le informo al alcalde que...").	Mismos criterios que C4, para un sistema nuevo sin guía.

1.4 4. Medición del desplazamiento cognitivo

El **desplazamiento cognitivo** es la diferencia entre el nivel alcanzado en una dimensión y el nivel de la sesión anterior. Se mide en dos escalas:

1.4.1 Desplazamiento intra-sesión

Compara el rastro inicial (pre-chatbot) con la reflexión de cierre de la misma sesión.

$$\Delta_{intra} = M4 \text{ (cierre DD_30)} - M1 \text{ (rastros papel)}$$

Nota metodológica — Asimetría de géneros entre M1 y M4 (SFL): M1 (ficha Pre-AI) es un género cerrado de 9 secciones obligatorias que fuerza la activación de las metafunciones ideacional, interpersonal y textual. M4 (cierre DD_30) es un género abierto de respuesta libre donde el estudiante elige qué metafunción activar. Las escalas D1-D4 se aplican a ambos momentos, pero la exigencia estructural no es equivalente: un nivel bajo en M4 puede reflejar que la pregunta no forzó ese recurso lingüístico, no que el estudiante no lo posee. Esta asimetría constituye una limitación metodológica que debe declararse en el paper (ver `doc_inv_SFL_Analisis_v1.1.md` §4). No invalida Δ_{intra} , pero obliga a interpretar sus valores con cautela.

1.4.2 Desplazamiento inter-sesión

Compara el nivel alcanzado en una clase con el de la clase anterior.

$$\Delta_{inter} = Nivel_{clase_N} - Nivel_{clase_(N-1)}$$

1.4.3 Interpretación

Δ	Interpretación
0	Estancamiento (posible techo cognitivo)
+1	Progresión normal esperada
+2 o más	Salto significativo (posible insight)
-1 o menos	Regresión (posible sobrecarga o confusión)

1.5 5. Criterios de calidad para el paper

Nota de formato: Las tablas de esta sección no incluyen la columna “Criterio de codificación lingüística” presente en §2. La tabla de cambio superficial vs. estructural es un marco conceptual (no un criterio de codificación por nivel). La tabla DD_27 es un indicador conductual de interacción con el chatbot (no un análisis de texto del estudiante). Para la codificación de D1-D4 sobre los textos, remitirse a los criterios lingüísticos de §2 y al Anexo A.

1.5.1 Cambio superficial vs. cambio estructural

Tipo de cambio	Indicador	Ejemplo
Superficial	Cambia palabras pero no la estructura del argumento	“Es el pH” → “Es el ORP” (sin mecanismo nuevo)
Estructural	Cambia la lógica del razonamiento	“El agua está turbia” → “La temperatura acelera consumo de cloro, lo que baja ORP y permite proliferación”

1.5.2 Indicador de calidad de defensa ante contraargumentación BUILD (DD_27)

Nivel	Descriptor	Indicadores observables
1 — Capitulación	Acepta la contraargumentación del chatbot sin resistir	“Ah, tienes razón” — cambia de posición inmediatamente
2 — Resistencia superficial	Insiste en su posición pero sin evidencia nueva	“No, yo creo que sí es un error” — sin argumento técnico
3 — Defensa técnica	Sostiene su posición con evidencia técnica específica	Cita valores, umbrales o relaciones causales que contradicen al chatbot
4 — Refutación fundamentada	Desmonta la contraargumentación del chatbot identificando la falacia o el dato incorrecto	“Tu argumento asume X, pero los datos muestran Y, por tanto Z”

Uso: Este indicador se registra en M3 (Post-BUILD) para C2-C4. Es dato complementario a D3 y D4 — mide la capacidad argumentativa bajo presión adversarial del chatbot. M3 es un indicador independiente de evaluación crítica y NO se usa para calcular Δ_{intra} (ver Paso 3).

1.5.3 Indicador de techo cognitivo

Si $\Delta_{\text{inter}} = 0$ durante dos sesiones consecutivas, considerar: - ¿El andamiaje es insuficiente?
- ¿El estudiante alcanzó su máximo con la información disponible? - ¿Hay un factor externo (tiempo, motivación)?

1.6 6. Protocolo de codificación

1.6.1 Paso 0: Calibración de codificadores (pre-piloto)

Antes de codificar datos reales de estudiantes, los dos codificadores humanos deben completar una fase de calibración con textos simulados. El objetivo es alinear su interpretación de los criterios de codificación lingüística (§2) antes de tocar el corpus.

Procedimiento:

1. **Textos de calibración.** El equipo de investigación construye 3 textos de estudiante simulados que cubran la variedad anticipada: uno de perfil predominantemente descriptivo ($D1 \cdot N1$, $D2 \cdot N1$, $D3 \cdot N1$), uno de perfil intermedio con progresión parcial ($D1 \cdot N2$,

D2 · N2, D3 · N2), y uno de perfil avanzado (D1 · N3, D2 · N3, D3 · N3, D4 · N3). Los textos no provienen de alumnos reales. Se construyen deliberadamente para incluir casos de borde: textos con D1 alto y D4 bajo, textos con D2 alto y D1 bajo.

2. **Codificación independiente.** Cada codificador asigna nivel 1-4 en D1-D4 para cada texto de calibración usando exclusivamente los criterios de codificación lingüística de §2. No se consultan entre sí.
3. **Confrontación.** Se comparan las asignaciones. Para cada desacuerdo, se clasifica la divergencia:
 - **De observación:** un codificador no vio un recurso que el otro sí identificó en el texto. Se resuelve señalando el recurso en el texto.
 - **De criterio:** ambos ven el mismo recurso pero discrepan sobre qué nivel le corresponde. Se resuelve refinando la redacción del criterio en §2.
4. **Iteración.** Se repite el ciclo con nuevos textos de calibración hasta alcanzar $\kappa \geq 0.85$ (ponderado, lineal/cuadrático) en al menos 3 textos consecutivos. Solo entonces se procede al corpus real.
5. **Documentación.** Los desacuerdos residuales, los casos de borde y las decisiones de criterio adoptadas se documentan como anexo metodológico del paper. Esto fortalece la transparencia y la replicabilidad.

1.6.2 Paso 1: Recopilar evidencias

Para cada estudiante, por cada clase, recopilar hasta CUATRO momentos: - **M1 — Rastro papel:** Foto del escrito a mano (procesada por AI Vision, almacenada en BD) - **M2 — Post-PLAN:** Conversación completa con chatbot socrático (logs en PostgreSQL) - **M3 — Post-BUILD:** Evaluación del documento generado por el chatbot — texto libre en el chat donde el alumno acepta, cuestiona o argumenta contra los errores (logs en PostgreSQL). Solo aplica en C2-C4. - **M4 — Cierre (reflexión DD_30):** Respuesta del alumno a “¿qué cambiarías de tu escrito inicial y por qué?” — disparada por el profesor desde el Dashboard (logs en PostgreSQL). Aplica en C1-C5.

Para C1: M1 (papel), M2 (post-socrático básico) y M4 (cierre). No hay BUILD ni M3. Para C5: M1 (papel), M2 (interacción con chatbot neutro) y M4 (cierre). No hay BUILD ni M3.

1.6.3 Paso 2: Codificación dual (IA + humano)

Codificación primaria (tiempo real): AGENT_SESSION invoca a AGENT_ANALISTA_SFL para aplicar D1-D4 sobre M1, M2, M3 y M4 automáticamente después de cada sesión. Resultado: niveles provisorios + justificación textual. AGENT_SESSION usa el resultado para generar feedback al alumno (proceso, no calificación) y al profesor (informe analítico).

Codificación para el paper: Dos codificadores humanos independientes asignan nivel (1-4) en cada dimensión, para cada momento. Trabajan sobre datos codificados (sin nombres). La codificación de la IA sirve como referencia pero NO reemplaza al codificador humano.

1.6.4 Paso 3: Calcular desplazamientos

Registrar Δ_{intra} y Δ_{inter} por estudiante y por dimensión. - $\Delta_{intra} = M4 - M1$ (diferencia entre reflexión de cierre DD_30 y rastro papel dentro de la misma sesión). M4 es la evidencia oficial de Δ_{intra} porque es el único momento que vuelve explícitamente sobre el rastro inicial (DD_30). - $\Delta_{inter} = M1_{clase_N} - M1_{clase_(N-1)}$ (diferencia entre rastros papel de sesiones consecutivas)

Nota: M3 (evaluación BUILD) se analiza como indicador independiente de capacidad evaluativa (D3, DD_27). No se usa para calcular Δ_{intra} .

1.6.5 Paso 4: Analizar patrones

- ¿Qué dimensiones muestran mayor desplazamiento?
- ¿En qué clase se produce el mayor salto?
- ¿Hay correlación entre dimensiones?
- ¿Clase 5 (transferencia) mantiene los niveles de Clase 4?
- ¿Los alumnos que aceptaron BUILD sin cuestionar (DD_29 activado) muestran D3/D4 más bajos?
- ¿La calidad argumentativa contra BUILD predice Δ_{inter} ?

1.6.6 Fiabilidad inter-codificador

Calcular Cohen’s Kappa (κ) ponderado entre codificadores humanos (lineal o cuadrático, por la naturaleza ordinal de la rúbrica: un desacuerdo 1↔4 pesa más que 1↔2). Umbral mínimo para reporte publicable: $\kappa \geq 0.80$ (substantial-strong, Landis & Koch). $\kappa \geq 0.70$ acepta solo análisis exploratorio. Adicionalmente: calcular κ entre AGENT_ANALISTA_SFL y codificadores humanos (dato metodológico para el paper — ¿puede la IA codificar D1-D4 de forma confiable?). Si $\kappa_{IA-humano} \geq 0.80$, es hallazgo publicable sobre viabilidad de codificación automatizada de razonamiento; si queda entre 0.70 y 0.79, se reporta solo como análisis exploratorio.

1.7 7. Plantilla de registro por estudiante

Estudiante (código): _____

Fecha Clase 1: _____

Clase	M1 (Papel)	M2 (Post-PLAN)	M3 (Post-BUILD)	M4 (Cierre DD_30)	Notas
	D1 D2 D3 D4	D1 D2 D3 D4	D1 D2 D3 D4	D1 D2 D3 D4	
C1	__ __ __ __	__ __ __ __	- (no BUILD)	__ __ __ __	
C2	__ __ __ __	__ __ __ __	__ __ __ __	__ __ __ __	
C3	__ __ __ __	__ __ __ __	__ __ __ __	__ __ __ __	
C4	__ __ __ __	__ __ __ __	__ __ __ __	__ __ __ __	
C5	__ __ __ __	__ __ __ __	- (no BUILD)	__ __ __ __	

Δ_{intra} por clase (M4 - M1):

C1: D1=___ D2=___ D3=___ D4=___

C2: D1=___ D2=___ D3=___ D4=___

C3: D1=___ D2=___ D3=___ D4=___

C4: D1=___ D2=___ D3=___ D4=___

C5: D1=___ D2=___ D3=___ D4=___

Δ_{inter} por clase (M1_N - M1_(N-1)):

C1→C2: D1=___ D2=___ D3=___ D4=___

C2→C3: D1=___ D2=___ D3=___ D4=___

C3→C4: D1=___ D2=___ D3=___ D4=___

C4→C5: D1=___ D2=___ D3=___ D4=___

Indicadores BUILD (M3 - evaluación crítica, independiente de Δ_{intra}):

C2: ¿Aceptó sin cuestionar? ___ ¿Push DD_29 activado? ___ ¿Recapacitó? ___ Calidad
defensa (1-4): ___

C3: ¿Aceptó sin cuestionar? ___ ¿Push DD_29 activado? ___ ¿Recapacitó? ___ Calidad
defensa (1-4): ___

C4: ¿Aceptó sin cuestionar? ___ ¿Push DD_29 activado? ___ ¿Recapacitó? ___ Calidad
defensa (1-4): ___

Observación general de trayectoria:

1.8 Anexo A — Fundamentos SFL para codificación (referencia para IA y codificadores humanos)

Este anexo proporciona las definiciones operacionales de los recursos lingüísticos mencionados en los criterios de codificación (§2). Está diseñado para que cualquier LLM o codificador humano pueda identificar estos recursos en textos de estudiantes sin conocimiento previo de Lingüística Sistémico-Funcional. Cada entrada incluye: definición operacional, ejemplos positivos del dominio de la piscina/torre de enfriamiento, ejemplos que NO califican, y la dimensión de la rúbrica asociada.

Referencia teórica: Halliday, M.A.K. & Matthiessen, C.M.I.M. (2014). Halliday's Introduction to Functional Grammar (4th ed.). Routledge. La terminología SFL se traduce aquí a criterios operacionales para codificación de textos en español técnico de ingeniería.

1.8.1 A.1 Mapa de metafunciones → dimensiones de la rúbrica

Metafunción SFL	Pregunta que responde	Dimensión	Qué buscar en el texto
Ideacional	¿Cómo construye el lenguaje la experiencia?	D1, D2	Nexos causales, tipos de procesos, participantes técnicos, nominalizaciones, valores numéricos con unidades
Interpersonal	¿Cómo se posiciona el hablante frente a lo que dice?	D3, D4	Operadores modales, marcadores evidenciales, hedging, verbos de decisión en primera persona, cláusulas condicionales
Textual	¿Cómo se organiza la información?	D4 (parcial), Transferencia C5	Estructura de etapas del género, cohesión, progresión temática

1.8.2 A.2 Glosario operacional de recursos lingüísticos

Cada recurso se define con ejemplos del dominio del proyecto. Los ejemplos positivos muestran el recurso presente en un texto de estudiante simulado. Los ejemplos negativos muestran textos donde el recurso está AUSENTE (y por tanto corresponden a un nivel inferior en la rúbrica).

Nexo causal explícito Definición operacional: Palabra o locución que establece una relación de causa-efecto entre dos cláusulas. La relación debe ser explícita en el texto, no inferida por el lector.

Marcadores típicos en español: porque, ya que, puesto que, debido a, por lo tanto, e ntonces, así que, lo que, lo cual, por eso, a causa de, como consecuencia de.

Ejemplos positivos (D1 · N2 o superior): - “El ORP bajó **porque** el pH está alto.” (1 nexos) - “El filtro está obstruido, **lo que** reduce el caudal.” (1 nexos, subordinación) - “La temperatura sube → el pH sube → el cloro se degrada **y como consecuencia** el ORP cae.” (3 eslabones, D1 · N3) - “El retrolavado fue insuficiente, **por eso** el ΔP sigue subiendo, **lo cual** indica que el filtro no se limpió.” (2 nexos encadenados, D1 · N3)

Ejemplos que NO califican (D1 · N1): - “El agua está turbia y el pH subió.” (yuxtaposición sin nexos causal) - “Hay un problema con el filtro. La presión está alta.” (dos oraciones sin

conector causal) - “Se observa turbidez, ORP bajo, pH alto.” (lista de síntomas sin relación causal)

Dimensión asociada: D1 — Complejidad causal.

Proceso relacional Definición operacional: Verbo que establece una relación de identidad, atribución o posesión, sin expresar acción ni cambio. Los más comunes son **ser**, **estar**, **tener**, **parecer**, **consistir en**.

Ejemplos (típicos de D1 · N1 cuando dominan el texto sin nexos causales): - “El pH es 7.8.” - “El agua **está** turbia.” - “El sistema **tiene** un problema.”

Diferencia con proceso material: Un proceso material expresa una acción o cambio observable: **subir**, **bajar**, **aumentar**, **disminuir**, **filtrar**, **bombear**, **degradar**. “El pH **subió** de 7.2 a 7.8” es proceso material. “El pH **es** alto” es proceso relacional. Un texto dominado por procesos relacionales sin nexos causales es indicador de D1 · N1.

Dimensión asociada: D1 — Complejidad causal.

Participante con valor numérico y unidad Definición operacional: Una variable técnica nombrada que aparece acompañada de un valor numérico y su unidad de medida correspondiente.

Ejemplos positivos (D2 · N3): - “ ΔP de **28 kPa**” - “pH de **7.8**” (pH es adimensional — la mención explícita del valor cuenta) - “caudal de **72 m³/h**” - “ORP en **580 mV**” - “temperatura de **23 °C**”

Ejemplos que NO califican (D2 · N1 o D2 · N2): - “El pH está alto.” (sin valor numérico — D2 · N2) - “El caudal bajó.” (sin valor numérico — D2 · N2) - “Algo anda mal con el sistema.” (sin variable nombrada — D2 · N1)

Umbral para D2 · N3: ≥ 3 participantes técnicos con valor + unidad en el texto completo.

Dimensión asociada: D2 — Especificidad técnica.

Nominalización técnica Definición operacional: Un proceso o cualidad expresado como sustantivo. En ingeniería, las nominalizaciones empaquetan fenómenos complejos en participantes que pueden entrar en relaciones causales.

Ejemplos positivos: - “La **caída de presión** en el filtro” (en lugar de “la presión cayó”) - “La **degradación del cloro** por temperatura” (en lugar de “el cloro se degrada”) - “La **obstrucción del filtro**” (en lugar de “el filtro se obstruyó”) - “El **consumo de cloro**” (en lugar de “se consume cloro”)

Relevancia: La presencia de nominalizaciones técnicas como participantes de procesos causales es indicador de D1 · N3 o superior. Permiten construir cadenas causales compactas: “La **obstrucción** del filtro causa la **reducción** del caudal” — dos nominalizaciones en relación causal.

Dimensión asociada: D1 — Complejidad causal.

Operador modal epistémico Definición operacional: Palabra o locución que gradúa el grado de certeza del hablante sobre lo que dice. Indica que el hablante reconoce que su afirmación no es una verdad absoluta.

Marcadores típicos en español: probablemente, posiblemente, quizás, tal vez, podría ser, puede que, es probable que, no estoy seguro, me parece que.

Ejemplos positivos (D3 · N2 o superior): - “**Probablemente** el filtro esté obstruido.” - “**Podría ser** que el pH esté alto por la dosificación manual.” - “**No estoy seguro** de que el ORP refleje la desinfección real.” - “**Quizás** el sensor de pH no está calibrado.”

Ejemplos que NO califican (D3 · N1): - “El problema es el filtro.” (asertivo, sin modalización) - “Hay que cerrar el sistema.” (asertivo, sin modalización) - “La causa es el pH alto.” (asertivo, sin modalización)

Dimensión asociada: D3 — Consciencia epistémica.

Marcador evidencial Definición operacional: Palabra o locución que indica la fuente de la información que el hablante está usando. Responde a “¿cómo sabes eso?” o “¿de dónde sale ese dato?”.

Marcadores típicos en español: según, de acuerdo a, los datos muestran, la tabla indica, se observa que, como se ve en, basado en.

Ejemplos positivos (D3 · N3 o superior): - “**Según los datos** del SCADA, el ORP empezó a bajar a las 10:00.” - “**La tabla muestra** que el ΔP sube consistentemente.” - “**Asumo que** no hay fuga porque no tengo ese dato.” - “**Los datos indican** una tendencia descendente en el caudal.”

Ejemplos que NO califican (D3 · N1 o D3 · N2): - “El ORP bajó.” (sin fuente — D3 · N1) - “Probablemente el ORP bajó.” (modalizado pero sin fuente — D3 · N2)

Dimensión asociada: D3 — Consciencia epistémica.

Cláusula condicional epistémica Definición operacional: Estructura que expresa qué sabría o podría afirmar el hablante si tuviera un dato que actualmente le falta. Típicamente usa la construcción “si [dato faltante], [conclusión modalizada]”.

Ejemplos positivos (D3 · N4): - “**Si** tuviera el dato de cloro libre residual, **podría** confirmar si el ORP bajo es real.” - “**Si** el sensor de pH estuviera calibrado, **sabría** si el offset es sistemático.” - “**Si** conociera la curva de calibración, **determinaría** si el sesgo es constante.”

Diferencia con cláusula condicional deóntica (D4): La condicional epistémica expresa qué se SABRÍA (“podría confirmar”, “sabría si”). La condicional deóntica expresa qué se HARÍA (“si el ORP baja de 650, cerrar el sistema”). La primera es D3; la segunda es D4.

Dimensión asociada: D3 — Consciencia epistémica.

Verbo de decisión en primera persona Definición operacional: Verbo en primera persona (explícita o implícita) que expresa que el hablante toma una posición sobre qué acción ejecutar. Es un acto de habla directivo o compromisivo.

Verbos típicos: recomiendo, propongo, decido, sugiero, optaría por, mi decisión es.

Ejemplos positivos (D4 · N2 o superior): - “**Recomiendo** hacer un retrolavado completo de 5 minutos.” - “**Mi decisión** es cerrar parcialmente hasta confirmar el ORP.” - “**Propongo** monitorear el ΔP cada hora antes de intervenir.” - “**Decido** no abrir el sistema hasta que el ORP supere 650 mV.”

Ejemplos que NO califican (D4 · N1): - “Habría que hacer un retrolavado.” (impersonal, sin sujeto que decide) - “Se recomienda cerrar.” (pasiva, sin sujeto que decide) - “Creo que el problema es el filtro.” (opinión, no decisión) - “Necesito más información.” (petición, no decisión)

Dimensión asociada: D4 — Decisión bajo incertidumbre.

Cláusula condicional de consecuencia (deóntica) Definición operacional: Estructura que nombra qué acción tomar si ocurre una condición, especificando la consecuencia negativa de no actuar o de que la acción falle. Combina condición + acción + riesgo.

Ejemplos positivos: - “Recomiendo retrolavado. **Si no resuelve en 2 horas**, el riesgo es exposición a agua con $ORP < 500$ mV.” (D4 · N3) - “**Si el ORP no sube a 650+ en 2 horas**, cerrar el sistema y dosificar shock de hipoclorito. **Si el pH supera 7.8**, ajustar dosificación de ácido.” (D4 · N4 — múltiples condiciones con escalamiento) - “**Si el ΔP sigue subiendo** después del retrolavado, hay que inspeccionar el filtro físicamente.” (D4 · N3)

Ejemplos que NO califican: - “Recomiendo retrolavado.” (decisión sin condicional — D4 · N2) - “Si el ORP está bajo es porque el cloro no funciona.” (condicional explicativa, no deóntica — D1)

Dimensión asociada: D4 — Decisión bajo incertidumbre.

Acto de habla de comunicación a autoridad **Definición operacional:** Fragmento del texto donde el estudiante formula lo que le diría a un superior, cliente o autoridad. Indica que el estudiante no solo decide, sino que sabe traducir su decisión técnica a un acto comunicativo profesional.

Ejemplos positivos (D4 · N4): - “**Le informo al alcalde** que el sistema no puede operar a plena capacidad hasta que se verifique la calibración del sensor de pH.” - “**Le comunico al jefe de operaciones** que el riesgo de exposición sanitaria existe si se abre antes de las 09:00.” - “**Mi recomendación al supervisor** es retrolavado inmediato. Si no resuelve, escalar a cierre total.”

Dimensión asociada: D4 — Decisión bajo incertidumbre (nivel 4).

1.8.3 A.3 Instrucciones de codificación para IA

Cuando un LLM actúe como codificador automático (AGENT_ANALISTA_SFL, DD_9), debe aplicar los criterios de §2 siguiendo este procedimiento:

1. **Segmentar** el texto del estudiante en cláusulas (cada cláusula tiene un verbo principal).
2. **Para D1:** Contar nexos causales explícitos entre cláusulas (ver A.2 — Nexo causal explícito). Clasificar el tipo predominante de procesos (relacionales vs. materiales). Detectar presencia de nominalizaciones como participantes de procesos causales.
3. **Para D2:** Identificar participantes técnicos. Verificar cuántos aparecen con valor numérico + unidad. Detectar presencia de umbrales de clasificación.
4. **Para D3:** Detectar operadores modales epistémicos, marcadores evidenciales y cláusulas condicionales epistémicas (ver A.2).
5. **Para D4:** Detectar verbos de decisión en primera persona, cláusulas condicionales de consecuencia, criterios de escalamiento y actos de habla de comunicación a autoridad (ver A.2).
6. **Asignar nivel** según el umbral cuantitativo de la columna “Criterio de codificación lingüística” en §2. Si el texto cumple parcialmente (ej: tiene 2 variables con valor pero no 3), asignar el nivel inferior.
7. **Justificar** la asignación citando el recurso lingüístico específico encontrado en el texto, no con juicios holísticos (“el razonamiento es sólido”). Ejemplo de justificación correcta: “D1 · N3: 3 nexos causales explícitos (‘porque’, ‘lo que’, ‘debido a’) + nominalización ‘caída de presión’ como participante.”

Principio de codificación conservadora: Ante la duda entre dos niveles, asignar el inferior. La codificación debe ser replicable: otro codificador con los mismos criterios debe llegar al mismo nivel leyendo el mismo texto.

Versión: 1.6 **Fecha:** 2026-05-31 **Referencia SFL:** doc_inv_SFL_Analisis_v1.1.md — fuente de la traducción de descriptores D1-D4 a marcadores lingüísticos observables.
