

# Ritmo y organización temporal del habla: Propiedades estructurales y correlatos acústicos

Antonio Romano, UniTO

## Resumen

En este artículo recopiló una serie de reflexiones sobre el tema de la organización temporal y la estructura rítmica del habla, destacando las propiedades específicas del idioma y los fenómenos que dependen de las características individuales de los hablantes y del uso accidental que hacen de ellas. Sin olvidar el papel del oyente (destacado desde Allen 1972), intentaré identificar los principales correlatos acústicos de este nivel de organización lingüística.

Como es bien sabido gracias a numerosas publicaciones que se han sucedido a lo largo de décadas, el ritmo lingüístico se clasifica habitualmente en referencia a la controvertida idea, conocida como la hipótesis de las clases rítmicas, de que todas las lenguas pueden agruparse según unidades rítmicas organizativas (casi-)isócronas que corresponden a sílabas o grupos acentuales (Abercrombie 1967). Sin embargo, en opinión de muchos (desde Roach 1982 hasta Arvaniti 2021, entre otros), las tipologías rítmicas han recibido poco apoyo de los estudios empíricos y las métricas alternativas basadas en otros constructos lingüísticos, como las relaciones entre las duraciones de los segmentos consonánticos y vocálicos en el habla (por ejemplo, Ramus *et al.* 1999, Grabe & Low 2002), aplicadas a muestras de muchas lenguas diferentes, han dado lugar a dispersiones de valores muy irregulares.

Como intenté demostrar en una conferencia invitada en la Universidad de Sala-

manca, dirigida a estudiantes y colegas no todos especialistas en este campo, las ramificaciones interdisciplinarias en el examen de este fascinante problema implican diversas consideraciones técnicas y suposiciones de carácter general sobre el papel de este nivel de estructuración del habla<sup>1</sup>.

Partiendo de la referencia a Keller (2007), es posible imaginar las razones por las que, en las producciones de un mismo grupo de hablantes, pueden prevalecer diferentes estrategias de organización rítmica en diferentes condiciones.

Parece evidente que un mismo hablante puede optar por estilos de elocución que dejan aflorar diferentes propiedades rítmicas. Una atención excesiva a la naturaleza corpuscular (granular < segmental) del habla puede, de hecho, conducir a un flujo sonoro «rígido», marcado, es decir, a ese ritmo lingüístico típico de las lenguas isosilábicas (¿e isosegmentales?). Por el contrario, el predominio de la atención a la naturalidad del flujo gobernado por íctus, aceleraciones y ralentizaciones, deja emerger una naturaleza ondulatoria (¿eurítmica?) del habla. A este nivel, parece evidente que el predominio de los fenómenos de compresión y rarefacción produce inevitablemente un habla más «accidentada» en la

---

<sup>1</sup> El texto aquí propuesto es una versión adaptada de la conferencia que impartí por invitación de la profesora Carmen Quijada Van Den Berghe en la inauguración de los *V Fonética Open Days* (Salamanca, 31 de marzo - 2 de abril 2025).

que los sonidos (¿y las sílabas?) son menos importantes, no dominantes —precisamente— en la fonología de una lengua.

Precisamente en consideración de los diferentes niveles de comprensión y de compensación identificados y descritos a lo largo de la vasta producción científica de P.M. Bertinetto, y en analogía con una representación ondulatoria-corpúscula de la luz, también la fonología de una lengua debe describirse otorgando la misma importancia a lo segmental y a lo suprasegmental, introduciendo instrumentos que permitan detectar cuándo, en el conjunto de las características sistémicas de una lengua o en sus manifestaciones individuales, una tiende a ser primitiva y la otra derivada.

Lo demostramos aquí, pasando en revista a modelos y procedimientos analíticos, pero concluyendo con la propuesta de una de las técnicas más prometedoras en el estudio de fenómenos recurrentes. Se trata del método de los llamados *Recurrence plots*, cuyas cualidades descriptivas (y predictivas) se demuestran aquí de forma experimental, a la luz de evaluaciones que, además de distinguir entre métricas consonánticas y vocálicas y devolver importancia a otros parámetros como  $f_0$  e intensidad, vuelven a centrar la atención en los dos niveles de organización temporal, intra- e inter-silábico, que se reflejan en la caracterización rítmica de las lenguas.

### Riassunto

In quest'articolo raccolgo un insieme di riflessioni attorno al tema dell'organizzazione temporale e della struttura ritmica del parlato, mettendo in evidenza proprietà che dipendono dalla lingua e fenomeni che dipendono dalle caratteristiche indi-

viduali dei parlanti e dagli usi accidentali che ne fanno. Senza dimenticare il ruolo dell'ascoltatore (messo in evidenza sin da Allen 1972), cercherò di identificare i principali correlati acustici di questo livello di organizzazione linguistica.

Come è ben noto grazie a numerose pubblicazioni apparse nel corso dei decenni, il ritmo linguistico viene solitamente classificato in riferimento alla controversa idea, nota come ipotesi delle classi ritmiche, secondo cui tutte le lingue possono essere raggruppate in base a unità ritmiche organizzative (quasi-)isocrone che corrispondono a sillabe o gruppi accentuali (Abercrombie 1967). Tuttavia, secondo molti (da Roach 1982 ad Arvaniti 2021 e altri), le tipologie ritmiche hanno ricevuto scarso sostegno dagli studi empirici e dalle metriche alternative basate su altri costrutti linguistici, come le relazioni tra le durate dei segmenti consonantici e vocalici nel parlato (ad esempio, Ramus *et al.* 1999, Grabe & Low 2002), applicate a campioni di molte lingue diverse, hanno dato luogo a dispersioni di valori molto irregolari.

Come ho cercato di dimostrare in una conferenza tenuta all'Università di Salamanca, rivolta a studenti e colleghi non tutti specialisti in questo campo, le ramificazioni interdisciplinari nell'esame di questo affascinante problema implicano diverse considerazioni tecniche e ipotesi di carattere generale sul ruolo di questo livello di strutturazione del parlato.

Partendo dal riferimento a Keller (2007), è possibile immaginare i motivi per cui, nelle produzioni dello stesso gruppo di parlanti, possono prevalere diverse strategie di organizzazione ritmica in condizioni diverse.

Sembra inoltre evidente che uno stesso parlante possa optare per stili di elocuzione che lasciano emergere diverse proprietà ritmiche. Un'attenzione eccessiva alla natura corpuscolare (granulare < segmentale) del parlato può, infatti, portare a un flusso sonoro «rigido», marcato, cioè a quel ritmo linguistico tipico delle lingue isosillabiche (e isosegmentali?). Al contrario, il predominio dell'attenzione alla naturalezza del flusso governato da ictus, accelerazioni e rallentamenti, fa emergere una natura ondulatoria (euritmica?) del parlato. A questo livello, sembra evidente che il predominio dei fenomeni di compressione e rarefazione produca inevitabilmente un linguaggio più «irregolare» in cui i suoni (e le sillabe?) sono meno importanti, non dominanti — proprio — nella fonologia di una lingua.

Proprio in considerazione dei diversi livelli di compressione e compensazione identificati e descritti nella vasta produzione scientifica di P.M. Bertinetto, e in analogia con una rappresentazione ondulatoria-corpuscolare della luce, anche la fonologia di una lingua deve essere descritta attribuendo la stessa importanza al segmentale e al soprasegmentale, introducendo strumenti che consentano di rilevare quando, nell'insieme delle caratteristiche sistemiche di una lingua o nelle sue manifestazioni individuali, una tende a essere primitiva e l'altra derivata.

Lo dimostriamo qui, passando in rassegna modelli e procedure analitiche, ma concludendo con la proposta di una delle tecniche più promettenti nello studio dei fenomeni ricorrenti. Si tratta del metodo dei cosiddetti *Recurrence plots*, le cui qualità descrittive (e predittive) vengono qui di-

mostrate in modo sperimentale, alla luce di valutazioni che, oltre a distinguere tra metriche consonantiche e vocaliche e a restituire importanza ad altri parametri come  $f_0$  e intensità, riportano l'attenzione sui due livelli di organizzazione temporale, intra- e inter-sillabico, che si riflettono nella caratterizzazione ritmica delle lingue.

## Introducción

El problema del estudio de la organización temporal y la estructura rítmica del habla ha recibido numerosas contribuciones relevantes en las últimas décadas, lo que ha requerido notables recursos experimentales, computacionales y estadísticos.

No oculto, por tanto, que me siento bastante avergonzado al escribir este texto, porque mi contribución en este ámbito solo puede ser, digamos, de naturaleza cultural y especulativa, con algunas consideraciones más técnicas que, en ciertos casos, pueden resultar limitadas en comparación con los trabajos de estudiosos mucho más competentes que asistieron a mi conferencia o que tal vez lean este artículo.

Empiezo precisamente por una cita que tomo de Victoria Marrero, de una charla que dio en el congreso de la *AISV* en Turín en el año 2014:

«Rhythm is one of the most pervasive aspects of the human condition; it is in the world around us and in the world within us, in our bodies and our minds, our living and our thinking. [...] human language, quite

predictable, is deeply rhythmic as well» (Auer, Couper-Kuhlen & Müller 1999, p. 3).

Parto de ella, que me parece una buena premisa, porque habla del ritmo como uno de los aspectos omnipresentes de la condición humana, porque —tanto en nuestros cuerpos como en nuestras mentes— tienen lugar actividades que presentan regularidades temporales.

No es casualidad que la pulsación y los ciclos respiratorios correspondan precisamente a una periodicidad de acontecimientos que suelen asumir una regularidad rítmica.

Por lo tanto, ¿estas sucesiones regulares de eventos se definen a partir de propiedades lingüísticas intrínsecas? ¿O son simplemente un fenómeno que observamos en el habla? ¿Corresponden a una estructuración interna profunda? ¿O son propiedades emergentes?

Esto equivale a preguntarse si el ritmo lingüístico es una propiedad fonológica, es decir, si pertenece a la lengua compartida en la comunidad, o si es una mera evidencia fonética que podemos detectar en los datos del habla individual<sup>2</sup>.

---

<sup>2</sup> De las publicaciones de mayor alcance que han intentado explorar estos temas a lo largo de un siglo de avances, surgen diferentes visiones que tienden a reconsiderar el ritmo del habla, no

Son preguntas que varios estudiosos ya se han hecho —cíclicamente, podemos decir— incluso durante estas décadas en las que la investigación lingüística ha buscado pruebas experimentales<sup>3</sup>.

## 1. La importancia del ritmo lingüístico

### 1.1. *El ritmo en la adquisición del lenguaje y en las tradiciones culturales*

Creo que es importante comenzar con una breve información sobre cómo surgen las actividades rítmicas durante la adquisición del lenguaje y algunas referencias al fenómeno por el cual el ritmo del habla de un hablante cambia como consecuencia del envejecimiento. Desde el punto de vista subjetivo, individual y ontogenético, intentaré pasar al filogenético, extrayendo algu-

---

como un hecho objetivo (sucesión de eventos), sino como una potencialidad subyacente del sistema, que se vuelve inaprensible debido a la variedad de usos individuales (Bertinetto 1981, Roach 1982, Wenk & Wiolland 1982, Dauer 1983). Por el contrario, se han afirmado estudios que lo han reclasificado como una característica secundaria, resultante en el habla de combinaciones de elementos determinados en otros planos de construcción (contribuciones relevantes en este sentido se recogen ahora en Arvaniti 2021).

<sup>3</sup> Sin embargo, como veremos, no han faltado contribuciones que, sin excluir estos efectos, han tratado de evaluar las cualidades fonéticas emergentes indicativas de componentes fonológicos combinados en varios planos (Bertinetto & Vékás 1991, Barbosa 2006, Bertinetto & Bertini 2011, Romano 2020).

nas consideraciones interesantes de la observación de las tradiciones orales que también nos permiten comprender qué papel puede haber desempeñado esta variable en la evolución humana. De hecho, en las diversas comunidades sociales, las cualidades rítmicas del lenguaje se han proyectado en el canto, por lo tanto, en relación con las regularidades que se pueden observar en las tradiciones musicales.

No me detendré en estos temas, porque no entran exactamente dentro de mi ámbito de aplicación cotidiano, pero es importante no perderlos de vista, con vistas a un análisis global de los fenómenos y de las propiedades lingüísticas subyacentes.

### *1.2. Ritmo, acento y entonación*

En el §2.10 me detendré un poco más en las relaciones entre el ritmo y otros fenómenos suprasegmentales; sin embargo, conviene destacar desde el principio su codependencia de los patrones acentuales, de la velocidad del habla y, posteriormente, de la estructuración entonacional de los enunciados.

Se han definido varias técnicas para analizar distintamente estos fenómenos y todas ellas han permitido arrojar luz sobre aspectos relevantes que asumen los índices acústicos de la prominencia en la codificación fonológica distintiva que les atribuyen las distintas lenguas.

Ahora nos encontramos en una fase en la que intentamos volver a aplicar métodos que analizan conjuntamente estos correlatos. En el §3.8, presentaré los resultados de un sencillo experimento que, aunque realizado de manera informal, anticipa las conclusiones estructuradas de otros estudios recientes, llevados a cabo con técnicas de investigación más sofisticadas.

### *1.3. Ritmo y organización temporal*

El ritmo, por definición una secuencia de *beats*, no puede vincularse únicamente a la duración, porque la duración de los acontecimientos corresponde a la organización temporal, a lo que llamamos *timing*. Así pues, el problema que señalaré a modo de conclusión es el de la necesidad de distinguir los patrones rítmicos de los patrones de organización temporal.

Evidentemente, se trata de fenómenos suprasegmentales que coexisten, pero que delimitan un espacio que conviene distinguir.

## **2. Primeras intuiciones sobre el ritmo del habla**

Empiezo hablando del ritmo del habla, desde los fundamentos, desde conceptos que quizá ahora todo el mundo comparte, porque venimos tratando este tema en particular desde que un estudioso galés, Arthur Lloyd James, consultor de la *BBC*, elaboró unas pautas en las que distinguía di-

stintos modelos de pronunciación relativos a lenguas que parecen estructurarse según un modelo de código Morse, es decir, línea-punto, segmentos largos frente a segmentos (muy) cortos, frente a lenguas que, por el contrario, presentan una mayor regularidad con la que se producen estos segmentos e inducen una elocuencia que recuerda a la regularidad del fuego de ametralladora<sup>4</sup>.

Esta figura fue retomada por Kenneth Pike, quien, a finales de los años cuarenta, propuso la definición de lenguas de tiempo acentual, es decir, lenguas con isocronía acentual, por oposición a las lenguas de tiempo silábico, es decir, lenguas con isocronía silábica, lenguas en las que dominan las duraciones de los constituyentes silábicos<sup>5</sup>.

Estos tipos de lengua, estos polos extremos, digamos, de lo que más adelante veremos que es un continuo, reconocen también propiedades segmentales particularmente recurrentes. De ahí que se plantee el problema de si considerar el ritmo como el resulta-

do de estas cualidades segmentales o, por el contrario, como podemos ver en una reelaboración de esta distinción, isoacentual/isosilábica propuesta por Abercrombie en 1967, partir en cambio de una visión en la que el ritmo es precisamente la causa por la que la organización segmental, y por tanto la articulación en sonidos, puede diferir de una lengua a otra<sup>6</sup>.

### 2.1. *Ritmo e isocronía*

Una valoración de este tipo estaba presente en la lingüística histórica, en la lingüística románica, porque ya se reflexionaba sobre el hecho de que en la evolución del latín vulgar hacia las lenguas románicas debía de haber cierta regularidad con la que se afirmaban ciertos fenómenos evolutivos debidos a propiedades de la estructura de la sílaba (la presencia de una consonante en la coda bloquea la vocal —sea corta o larga—, mientras que la ausencia de esta última conduce al alargamiento: las vocales del latín sólo se habían alargado en sílabas abiertas; cfr. Grammont 1913).

En su momento, este condicionamiento silábico de la evolución de la lengua se denominó isocronía porque actuaba como una igualación de longitud entre sílabas abiertas y cerradas, pero no hay que suponer que se trate

---

<sup>4</sup> V. «morse-code» vs. «machine-gun (languages)» en James (1940).

<sup>5</sup> Según Pike (1945), la expresión «Stress-timed» se reserva para las lenguas cuyo ritmo está dominado por los patrones de acentuación, mientras que la expresión «Syllable-timed» se refiere a las lenguas cuyo ritmo está dominado por los patrones silábicos. Su postura, importante de señalar, se basa en la afirmación de que las evidencias segmentales predicen las cualidades y los cambios rítmicos.

---

<sup>6</sup> Al contrario de lo sostenido por Pike (véase la nota anterior), Abercrombie (1967) plantea la hipótesis que sean las cualidades rítmicas las que determinen las realizaciones fonéticas.

de una propiedad permanente de los diferentes contextos que se originaron de este modo<sup>7</sup>.

Volveré sobre este importante tema en el §2.3.

## 2.2. Ritmo y percepción

Otro aspecto importante que hay que desarrollar, como se ha anticipado, es lo relacionado con las observaciones de otro estudioso que en los años 70 demostró cómo el ritmo puede ser de hecho una emergencia psicoacústica: en una secuencia de impulsos equiespaciados, todos idénticos, el oyente puede percibir diferencias y comenzar a agrupar los impulsos a partir de uno que no es objetivamente saliente, sin que exista un índice acústico que autorice a asignar a ese impulso el comienzo de una regularidad, de una recurrencia que sugeriría una agrupación: tenemos así la emergencia de un ritmo basado en la percepción de tiempos más fuertes que en realidad son exactamente idénticos a los demás<sup>8</sup>.

---

<sup>7</sup> Es el mismo Abercrombie (1967) quien retoma el principio según el cual, en las lenguas románicas, la «isocronía» se refiere a la duración de las sílabas abiertas frente a las sílabas cerradas (comparando (C)CV: con (C)CVC).

<sup>8</sup> Con un experimento sobre una serie de impulsos idénticos, Allen (1975) demuestra que en la percepción de un grupo de oyentes surgen *beats*, es decir, impulsos que se perciben como más fuertes a distancias regulares. En sus conclusiones el ritmo del habla no es necesariamente una propiedad física de las señales, sino principalmente una percepción subjetiva.

Ante tal evidencia, deberíamos ver el ritmo como una necesidad por parte del oyente, una necesidad de orden, y relacionado con esto está uno de los primeros estudios de Plinio Barbosa, que en los años 90, durante su doctorado, trabajó sobre estos «centros de prominencia» que surgen en correlación con la disposición humana a percibir un ritmo incluso donde no lo habría.

El oyente, en la práctica, asignaría una saliencia particular a ciertos impulsos en torno a los cuales —o a partir de los cuales— todo se organiza en intervalos regulares.

En su momento, esto también fue observado por otros estudiosos que incluso habían conseguido alinear este centro de percepción, el *p-center*, con los eventos acústicos, señalando un resultado interesante, a saber, que los oyentes tienden a asignar este centro de prominencia al ataque vocálico, es decir, justo al principio de una vocal. La vocal, núcleo de sonoridad de la sílaba, parece ocupar un lugar privilegiado en la percepción del ritmo<sup>9</sup>.

---

<sup>9</sup> Los *P-centers* se definen como *centros de prominencia* o *centros perceptivos* (Barbosa 1994) y corresponden al momento (psicoacústico) de la señal al que el oyente atribuye su percepción auditiva de los eventos que ocurren a intervalos regulares (véase Scott 1993 y Janker 1995). En el habla, se ha demostrado que los *P-centers* se encuentran cerca del inicio vocálico (véase ahora Deighton MacIntyre et al. 2022).

Volvamos, pues, a Auer *et al.* (1999) y reformulemos el problema en términos de correspondencias entre el plano acústico y el perceptivo:

«Rhythm [...] is not a property of physical signals such as acoustic waveforms; acoustic data may contain phonetic events [...] and these may occur at objectively equal or similar time intervals. Yet the perception of rhythm is not directly or automatically related to (or derived of) these physical events. It is the human mind which perceives certain physical cues as forming a rhythmic pattern or Gestalt. More precisely, the human receptor of the acoustic signal must perform a number of interpretative tasks to hear its rhythm» (Auer, Couper-Kuhlen & Müller 1999, p. 23).

### 2.3. *Isocronías y isocronismos*

Reconociendo el papel determinante de la percepción, se continuó, como veremos, con la investigación experimental sobre los datos para verificar las razones acústicas por las que algunas producciones lingüísticas se clasificaban de una manera y otras de otra. Y es en este ámbito donde las evaluaciones experimentales han llevado a superar la necesidad o la posibilidad de confirmar la permanencia de una isocronía lingüística en el habla.

De hecho, aunque exista una *isocronía* silábica, la variación de la velocidad del habla y la variación de los estilos de

un mismo hablante no permiten poner de relieve una regularidad particular. En cambio, quizá sea la referencia al concepto de *isocronismo*, basado en otros tipos de regularidad, lo que supera la necesidad de encontrar sílabas de duraciones idénticas y empuja, en cambio, a buscar las razones por las que estas duraciones varían y en qué posiciones, y evaluar así cómo puede demostrarse experimentalmente la clasificación perceptiva de una muestra de habla hacia uno u otro polo.

Un mismo hablante puede cambiar a veces de estilo de habla y ello conlleva variaciones en la organización temporal que no entran en conflicto con la idea de que en sus actividades de producción surge el mismo tipo rítmico (Loukina *et al.* 2011).

Por otra parte, algunos estudiosos ya han demostrado que la variación intraindividual no incide de manera incisiva en la posibilidad de apreciar regularidades rítmicas en un mismo hablante y entre hablantes de una misma comunidad, induciendo a atenuar los conceptos «stress-timed / syllable-timed» en términos de «stress-based / syllable-based» y animando a cuantificar las dimensiones de la variación (Schmid & Dellwo 2013).

Lo que dificulta estas evaluaciones es el hecho de que el tipo rítmico, la regularidad rítmica que se da en el habla de una misma lengua, de un mismo hablante, depende de diversas variables,

entre ellas también la organización de los enunciados en términos de patrones acentuales y la organización más fina de los gestos articulatorios que se realizan localmente.

Esto aparece también haciendo referencia a un modelo sobre el que volveremos más adelante (§3.4), que es el de los osciladores: pensemos en cómo oscilan en el habla dos niveles de organización, uno de tipo silábico y otro de tipo enunciativo o rítmico-entonativo.

Pero lo que importa es que el ritmo varía con la velocidad del habla y que los diferentes tipos de habla que observamos en el ritmo del lenguaje podrían depender de fenómenos de coordinación (Dellwo & Wagner 2003), exactamente igual que en la observación de las diferentes posibilidades de sincronización y acoplamiento de diferentes osciladores (como, por ejemplo, en el caso de las patas y los pasos de marcha de los caballos).

#### 2.4. Tipos diferentes de sincronización

Las analogías con las oscilaciones observadas en la marcha de los caballos son interesantes porque en la práctica es como si hubiera dos osciladores sincronizados según el tipo de marcha.

Podemos observar los movimientos en fase o contrafase de las patas anteriores/posteriores o los de las patas derechas e izquierdas que se sincronizan de forma diferente dependiendo precisamente de la velocidad. En los

distintos idiomas existen formas de referirse a estos aires del caballo que reflejan la observación de que, a menor velocidad, las patas delanteras y traseras se sincronizan de una manera determinada.

A medida que aumenta la velocidad, este modo de sincronización cambia, por lo que es como si en una lengua al aumentar la velocidad se pasara realmente de un tipo rítmico a otro.

Esto lo observó Eric Keller en varios estudios a principios de la década de 2000, pero desde entonces otros estudiosos lo han comprobado de otras maneras: es importante tenerlo en cuenta porque todo lo que trataremos a continuación tiene que ver con estas variables de la producción del habla<sup>10</sup>.

---

<sup>10</sup> Keller (2007) precisa «Canter is [a horse gait] situated between walking and galloping, and usually consists of the right front foot leading the set of four hooves hitting the ground in sequence» y luego sostiene «This pattern probably emerges because the front and the back portions of the horse coordinate a rapidly occurring action in an articulated structure of bone, muscle and fatty tissue [...] . However when the movement is not as rapid as in a walk or in a trot, a balanced step usually emerges between the four legs». Concluye entonces que [As for horse gaits,] «the origin of beats in speech might be found in inferential patterns involving linguistic material that is sufficiently similar and occurs in sufficient temporal proximity» (Keller 2007: 357-358).

### 2.5. *Diversos osciladores: actividades motoras portantes y portadas*

Si bien el ritmo sigue siendo detectable mediante la medición de propiedades físicas, no olvidemos que su percepción se ve afectada por las proyecciones, conocimientos y expectativas del oyente. No todo el mundo ha perfeccionado las mismas habilidades técnicas para reconocer los índices de una organización rítmica diferente: no es una propiedad que podamos detectar de forma predecible a partir de los datos, del análisis de una única variable acústica, porque la sensibilidad a ciertas regularidades depende de la exposición que haya tenido el hablante y de la atención que prestó a ciertos fenómenos.

En términos acústicos, debemos esperar que un conjunto de correlatos dé cuenta de la oscilación mandibular, por ejemplo, porque podemos ver claramente que dejamos oscilar la mandíbula mientras hablamos. En esta oscilación, en este movimiento de apoyo, anclamos todas las actividades más sutiles de la articulación del sonido. Esto se ha observado en la construcción de un modelo teórico que denominamos «Frame then content». MacNeilage y Davis observaron en niños la aparición del balbuceo, del *babbling*, precisamente en asociación con el inicio de la oscilación mandibular. A partir de un balbuceo canónico, basado siempre en los mismos sonidos, se desarrolla un balbuceo abigarrado, que corresponde al momento en que el con-

tenido de este marco silábico comienza a diferenciarse. Esto ocurre precisamente en la fase de desarrollo en la que el niño comienza a realizar actividades rítmicas con sonajeros o tambores y similares<sup>11</sup>. Más adelante, precisamente, se desarrolla la atención a una construcción interna de los constituyentes silábicos. Una diferenciación lingüística de los sonidos que realizan esta estructura interna está asociada a una especialización de las regularidades del marco de los gestos llevados, con consecuencias para el tiempo de los gestos llevados, como muestra la fonología articulatoria gestual de Browman & Goldstein (1989) o como se puede observar en segmentos organizativos de mayor extensión en los que se aplican otros modelos (véase, entre otros, el modelo  *$\pi$ -gesture*; cf. ahora Byrd & Krivokapić 2021)<sup>12</sup>.

---

<sup>11</sup> Véase el sugerente cuadro ofrecido por MacNeilage & Davis (1990) y MacNeilage (1998).

<sup>12</sup> El modelo de análisis fonológico propuesto por Browman & Goldstein (1989) se basa precisamente en el concepto de gesto articulatorio, y los resultados objetivos (fonéticos) de la actividad de producción lingüística pueden depender de la forma en que se coordinan los gestos (§3.4). A un nivel más alto de organización, la escansión interna de los enunciados en segmentos entonativos depende de la interacción de dinámicas léxicamente especificadas de los gestos segmentales con los gestos prosódicos que dependen de una modulación que varía en el tiempo (Byrd & Krivokapić 2021).

## 2.6. Sincronización: avances y retrocesos a lo largo de la vida

En el plano de la modelización fonética, esta estructura dual vuelve a aparecer en el modelo de O'Dell & Nieminen (1999) basado en dos osciladores sincronizados y acoplados (tema al que volveremos): es muy interesante referirse al marco en el que Roman Jakobson habla del hacer y deshacer del lenguaje, es decir, de cómo en la adquisición del lenguaje se producen fenómenos de alineamiento progresivo hacia la lengua adulta y de cómo uno se aleja de esta lengua precisamente en la fase de envejecimiento debido a las dificultades progresivas para coordinar las actividades articulatorias<sup>13</sup>.

Esto se observa en los pacientes disártricos. Así como en los niños la actividad del lenguaje se desarrolla asociada a los movimientos de las extremidades, en los ancianos la disartria se asocia a la apraxia, la incapacidad de coordinar bien los movimientos de las extremidades superiores o inferiores etc.<sup>14</sup>.

Yo no he trabajado en este dominio, salvo en la medida en que, junto con Paolo Mairano (véase §3.2), proporcionamos la herramienta a algunos

estudiosos que utilizaron nuestras técnicas de medición para trabajar sobre este tipo de datos (relacionados con fenómenos de apraxia)<sup>15</sup>.

Por otra parte, es interesante ver cómo, en el estudio de la adquisición del lenguaje, otro modelo basado en la evaluación de variables lingüísticas de primer nivel —morfemas, lexemas, sintaxis etc.— sigue siendo compatible con estas consideraciones porque aquí también vemos una superposición de actividades lingüísticas en varios niveles que se combinan, que se entrecruzan.

El modelo desarrollado por Kate Demuth —a quien tuve la oportunidad de conocer precisamente durante los años en que trabajaba en este modelo— sigue siendo compatible con el propuesto por la teoría de MacNeilage & Davis aunque —como suele ocurrir— estos estudiosos no tengan en cuenta los avances de los demás en sus distintos campos de aplicación<sup>16</sup>. Así, mientras algunos inve-

---

<sup>15</sup> Véase Fougeron & Pillot-Loiseau (2015) and, more recently, Hernandez *et al.* (2020) y Yeo *et al.* (2021).

<sup>16</sup> A principios de la década de 1990, K. Demuth trabajaba en la adquisición del lenguaje y obtuvo resultados interesantes sin hacer referencia a investigaciones sobre el ritmo realizadas en otros ámbitos y en idiomas distintos del inglés. Sin embargo, en los años siguientes, dado que varios investigadores habían despertado un interés general por el ritmo en la percepción del habla y la adquisición del lenguaje, Demuth (1996) incluyó eficazmente el ritmo como base para el *bootstrapping*.

---

<sup>13</sup> Véase Jakobson (1941-42 > 1947).

<sup>14</sup> Véanse los numerosos estudios sobre la caracterización rítmica en el habla de pacientes con disartria (Liss *et al.* 2009, Marrero 2014, Maffia *et al.* 2023).

stigadores siguen líneas de investigación y desarrollan modelos teóricos potencialmente convergentes, los estudiosos de otros campos hacen otras observaciones útiles que no siempre se incluyen en las investigaciones de otros.

### *2.7. El ritmo en las culturas y en la historia de la humanidad*

En apoyo del estudio de las formas en que el ritmo se establece o se pierde en las producciones individuales de los niños o los ancianos, se extiende también la reflexión sobre la evolución filogenética de las lenguas sobre una base rítmica: en efecto, es importante correlacionar el ritmo del habla también con los modos de transmisión de las tradiciones orales, con el ritmo de las tradiciones musicales. Hay una serie de estudiosos que se ocupan de esto, y el trabajo de Leroi-Gourhan ha sido muy impactante, ayudando a reconstruir el nacimiento de las lenguas en relación con el desarrollo del habla articulada que puede remontarse al momento de la historia humana en que se establecieron las actividades rítmicas estructuradas. Y esto se encuentra aún hoy en la analogía que puede establecerse entre un mensaje en una lengua hablada y otro en un código rítmico «artificial» como el que se observa, por ejemplo, en las lenguas llamadas tamborinadas, que transmiten un mensaje recodificado según el modelo de la prosodia de la lengua hablada.

Hay reflexiones importantes por las que remito a Leroi-Gourhan (1964-65)

y a algunos ejemplos preciosos sobre las lenguas tamborinadas recopilados y analizados por Cloarec-Heiss (1999) que pueden ilustrarse como en la figura 1.

Creo, en efecto, que es igualmente importante estudiar las tradiciones orales, en las que se puede observar que la versificación popular se sirve de estructuras métricas que ayudan a la memorización de largos tramos de versos, en los cuentos, en las canciones infantiles y en todas las actividades humanas relacionadas, digamos, con este ámbito, en el desarrollo y la conservación de un patrimonio de tradiciones culturales. Así lo confirman una serie de trabajos que he realizado en este ámbito, que muestran una recurrencia particular de patrones, la prominencia de la estructura rítmica en la composición y transmisión de textos para los que desempeña la función esencial de soporte mnemotécnico<sup>17</sup>.

---

<sup>17</sup> Sin poder extenderme aquí sobre los resultados de mi investigación sobre el tema, destaco sin embargo que es posible detectar una extrema regularidad en las producciones de algunos portadores de la tradición. En el caso de un hablante cuyas producciones fueron grabadas por Vittorio Zacchino (Galatone, Lecce, 1969), pude estudiar 81 cuartetos de endecasílabos (336 versos) producidos espontáneamente por el hablante con la ayuda de una mímica rítmica. La observación permitió identificar patrones recurrentes en términos de escanción métrica y modulación melódica, con una periodicidad aún mayor que la estrofa (Romano 2020).

mā fóró tǎ-mā fófórò nē

I/real+fool/myself/neg.+fool/not  
I am not fooling myself

yē ngbǎrò àndì ngbǎngbǎrò nē

you all/real+seem/fool/neg.+seem/not  
don't play the fools

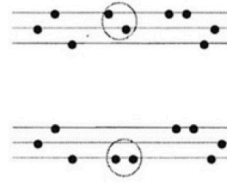


Fig. 1. Esquema de correspondencia entre esquemas tonales y transposiciones tamborinadas [extraído de Cloarec-Heiss (1999, p. 151)].

### 2.8. Módulos rítmico-melódicos en la versificación y en el text-setting musical

Por otra parte, la correlación con la música y el hecho de que las actividades rítmicas que caracterizan el habla también estén presentes en el canto y en la poesía nos da que pensar (véase Kruckenberg & Fant 1993, Fant & Kruckenberg 1996, Patel 2008, Colonna 2022).

En esencia, la organización del habla varía debido a la diferente atención que se presta a fenómenos locales o a fenómenos más globales, en una organización de patrones acentuales que coexisten con patrones entonativos y con patrones de la estructura informativa del enunciado. También en el canto, la organización rítmico-melódica debe tener en cuenta las pausas, es decir, el modo en que se segmenta el aire, con el que se crea la modularidad: de ello se han ocupado etnomusicólogos y musicólogos, así como lingüistas que han reflexionado sobre ello a distintos niveles, estudiando preci-

samente las regularidades con que aparecen los acentos en la versificación y la puesta en texto musical, el *text-setting*, la configuración textual, es decir, el modo en que el texto se corresponde con un módulo musical o viceversa, cómo un módulo musical puede adaptarse a un texto predefinido (v. §2.9)<sup>18</sup>.

Hay muchas investigaciones sobre estos temas. Aquí sólo remito a las consideraciones de Patel (2008), quien dedica varios capítulos a aclarar las diferentes relaciones entre una producción métrica y melódica adaptada a los esquemas musicales y

<sup>18</sup> Adaptar las palabras a la música a menudo implica la creación de versos antimetros en la composición contemporánea (aunque el fenómeno no es ajeno a la tradición, cf. Proto & Dell 2013). Además, la diferencia en el tipo de relación entre la prosodia y la música también se establece en función de las cualidades intrínsecas de las lenguas: Rodríguez Vázquez (2010) ha demostrado, por ejemplo, la imposibilidad de utilizar una configuración textual típica española en inglés y viceversa (véase también §2.9).

ca - pi - to ca - pi - to ca - pi - tò ca - pi - to ca - pi - to ca - pi - tò

5 ca - pi - to ca - pi - to ca - pi - tò ca - pi - to ca - pi - to ca - pi - tò

9 ca - pi - to ca - pi - to ca - pi - tò ca - pi - to ca - pi - to ca - pi - tò

Fig. 2. Reorganización rítmico-melódica de secuencias de palabras con prominencias léxicas en posiciones distintas. Tres repeticiones del italiano *capito*, *capito*, *capitò* (esdrújula, llana y aguda) las dos primeras recitadas (primero como lista / secuencia ditrocaica: *capito ca/pito capi/tò*; y después como secuencia de tres troqueos: *capí/to ca/pítò capi/tò*); la tercera, cantada [partitura definida por Arturo Romano con el software *Musescore* basándose en las versiones producidas por el autor en su tonalidad más adecuada].

las regularidades de otra naturaleza presentes en el habla<sup>19</sup>.

### 2.9. Restricciones acentuales y patrones musicales

Además de hacer referencia a diferentes tradiciones musicales (siguiendo las indicaciones de Pamies 2010), es importante tener en cuenta las características específicas de las lenguas.

<sup>19</sup> El habla es, de hecho, muy irregular desde el punto de vista musical, no es como el canto. Depende de la velocidad del habla y de las pausas, que no son tan regulares como en la música (por ejemplo, tempo largo, adagio... pausas de duración fija) e incluso los *beats* no están espaciados de manera uniforme, al igual que la entonación no es un movimiento entre notas musicales de altura fija (Patel 2008).

En italiano, por ejemplo, al igual que ocurre con las lenguas ibero-romances, a pesar de las posibilidades de variación en el acto de hablar, el acento léxico de las palabras polisilábicas contribuye a definir posiciones de mayor prominencia a diferentes distancias que pueden corresponder a los *beats* de patrones musicales apropiados<sup>20</sup>.

Sin embargo, las recurrencias rítmicas de un esquema musical a priori pueden no corresponder con las del

<sup>20</sup> Por el contrario, en lenguas cuyo léxico consiste principalmente en palabras de una o dos sílabas, se observa un mayor grado de regularidad en la distancia entre los acentos, lo que favorece patrones más equidistantes y, por lo tanto, compases más regulares (cf. Bertinetto & Bertini 2010).

texto hablado, al que se le exige cierta elasticidad. Una misma secuencia de palabras con posiciones acentuales precisas puede, por lo tanto, escanearse de distintas maneras, y esta escansión puede corresponder *tant bien que mal* a una versión cantada, reasignando las posiciones fuertes según un esquema rítmico que corresponde a las restricciones de la estructura musical (figura 2).

He tratado este tema incidentalmente, a raíz de algunos trabajos fin de grado que investigaban las elecciones que se hacen en la creación de letras en italiano a partir de canciones en inglés en la música pop, pero también en el jazz y en la música clásica (*euro-culta*): el texto en las traducciones a las distintas lenguas requiere reajustes que conllevan fenómenos de alargamiento o acortamiento<sup>21</sup>.

Queda mucho por decir, pero, en resumen, la duda de que las medidas de duración por sí solas no son la solución al problema de la clasificación

rítmica de las lenguas (v. §3) ya surge aquí: la necesidad de conciliar el ritmo de la composición original con las posibilidades de acentuación del texto en la lengua de llegada parece depender no tanto de las duraciones de los sonidos individuales como del momento mismo en que se ataca el sonido.

Volveremos sobre este tema, pero no puedo detenerme en él porque es importante, en cambio, reflexionar sobre la organización acentual, sobre cómo se pueden evaluar estas regularidades en los intervalos entre acentos, que existen y pueden ser más o menos rígidas según las lenguas.

#### 2.10. Patrones de prominencia y fonología métrica

Para mostrar la regularidad con la que se presentan los patrones de prominencia en función de la organización sintáctica y rítmico-entonativa de los enunciados, me detendré aquí solo en un ejemplo tomado de otra actividad experimental realizada con P.

---

<sup>21</sup> En los últimos números del *Bollettino LFSAG* se publicaron dos trabajos sobre este tema en el caso de cantantes de ópera profesionales (Ginepro 2024, Romano 2024). Además, se obtuvieron demostraciones experimentales sobre algunas canciones comerciales a partir de las investigaciones realizadas por Alice Felto (2015), y se analizaron las *performances* de varios intérpretes de jazz en la tesis de Amaryllis Russo (2019), que tuve el placer de supervisar junto con Valentina Colonna. Sin embargo, lo que parece más importante señalar es que en el último congreso internacional *ICPhS* se

dedicó una sesión especial a este mismo tema («Interplay or intermezzo? Structures and processes in prosody and music»): ponencias de estudiosos de un horizonte lingüístico muy amplio han mostrado precisamente cómo las tradiciones musicales que se difunden desde distintos países requieren ajustes lingüísticos en función de la lengua en la que tienen lugar. Un ejemplo de las contribuciones presentadas en esta sesión, aunque dedicado en su mayor parte al análisis de las características generales de las lenguas consideradas, puede ser el ofrecido por Kubozono & Mizoguchi (2023).

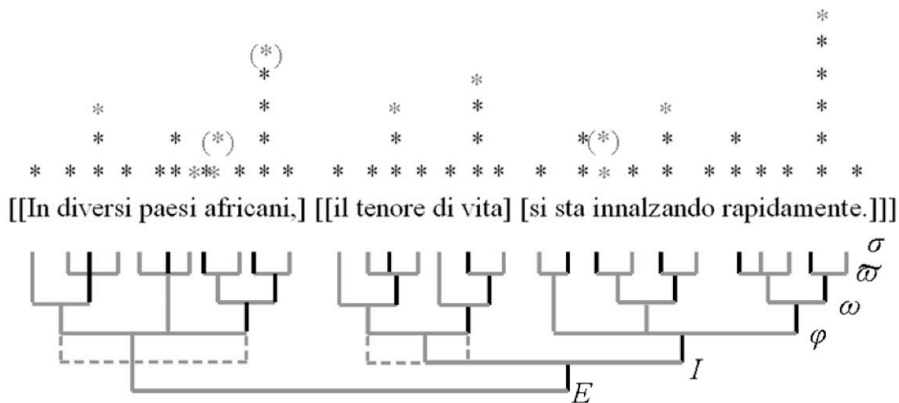


Fig. 3. Ejemplo de cuadrícula métrica y árbol de dependencias para una afirmación sometida a verificación perceptiva. Distinguimos, con asteriscos negros, los que resultan de la comparación entre las cuadrículas propuestas por dos operadores independientes y, con asteriscos más claros, los que resultan de la evaluación de los pesos asociados (omitiendo ciertos fenómenos fonosintácticos previsibles). Las líneas discontinuas indican otras posibles modalidades de escaneo (por ejemplo: escaneo de las dos primeras unidades como intonemas independientes en lugar de como «sintagmas fonológicos»).

Mairano y presentada en Romano & Mairano (2011).

El ejemplo es útil para ilustrar, por un lado, la correlación entre las prominencias efectivas, como resultado de las jerarquías de acento y la aparición de patrones métricos en el habla conectada, y, por otro, la correspondencia no siempre perfecta con formulaciones teóricas elaboradas de manera demasiado distante del habla real.

He escogido aquí (figura 3) sólo un ejemplo que compara el modelo de análisis de fonología prosódica de Nespor & Vogel (1986) con el modelo de cuadrícula métrica de Liberman & Prince (1977), evaluados por fonólogos con distinta experiencia que, por tan-

to, discrepan en ciertas decisiones a la hora de verificarlas en términos fonéticos, incluso partiendo de las realizaciones de un hablante profesional. Nuestro locutor en esta frase, por ejemplo, realizó tres unidades de entonación (Podría haberlo hecho de otra manera, pero lo hizo así, respetando una regularidad que quizá otro hablante habría ejecutado de forma diferente).

### 2.11. Elasticidad de los intervalos interacentuales

Una forma de evaluar las posibilidades de que en una lengua pueda variar la distancia entre las posiciones prominentes es medir la variabilidad de un intervalo que definimos preci-

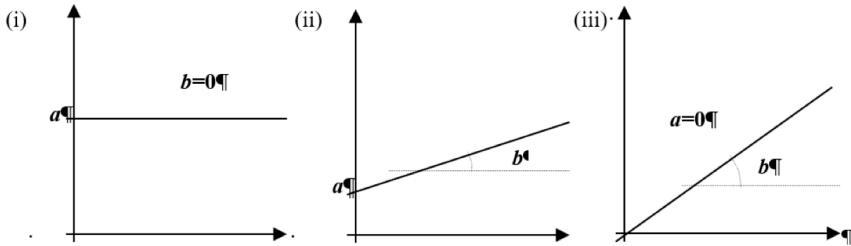


Fig. 4. Línea de crecimiento ideal para la duración de los intervalos interacentuales (en función del número de sílabas) para (i) lenguas estrictamente isoacentuales (a la izquierda), (iii) lenguas estrictamente isosilábicas (a la derecha) y (ii) lenguas con isocronía mixta (en el centro).

samente como intervalo interacentual, *Inter-stress Interval*, y que se define en función del número de sílabas incluidas entre una sílaba prominente y otra. Ahora bien, dependiendo del tipo rítmico, encontraremos lenguas que tienden a mantener constante este intervalo. Y así, si la fórmula es ésta

$$I(n) = a + b \cdot n$$

y  $n$  es el número de sílabas, tener un intervalo constante (como en el caso de las lenguas isoacentuales) significa que el valor del factor  $b$ , que afecta a la neutralización de este número, debe ser un valor muy bajo, idealmente cero, para que el intervalo entre acentos sea constante.

Lo contrario ocurre en el caso de una lengua en la que la sílaba domina debido a su rigidez, obligando a  $I$  a crecer en función de  $n$ .

En el caso en que el número de sílabas imponga una variación en la duración del intervalo acentual, tendremos un caso en que este coeficiente  $a$  es

nulo y en cambio la longitud del intervalo varía progresivamente, proporcional al número de sílabas, a través de este parámetro de rigidez que hemos llamado  $b$ .

Muchos estudios han intentado verificar este parámetro en los años 80, evaluando empíricamente el valor que adquieren en las muestras de las lenguas observadas (véanse las exhaustivas reseñas de Marotta 1985 para el italiano y de Pamies 1999 para el español y el francés). Lo que vemos surgir es generalmente una tendencia de tres tipos diferentes (v. figura 4).

Así pues, esto se aplica a las lenguas que consideramos aquí en figura 5, que suponemos isosilábicas. A pesar de esta regularidad, vemos que sigue habiendo una constante, un intercepto  $a$  distinto de cero. A la inversa, haciendo experimentos con el inglés y las lenguas isoacentuales supuestas, la pendiente cambia indudablemente, de modo que la curva se aplana, pero el valor de  $b$  no es 0. Así que incluso las duraciones de los

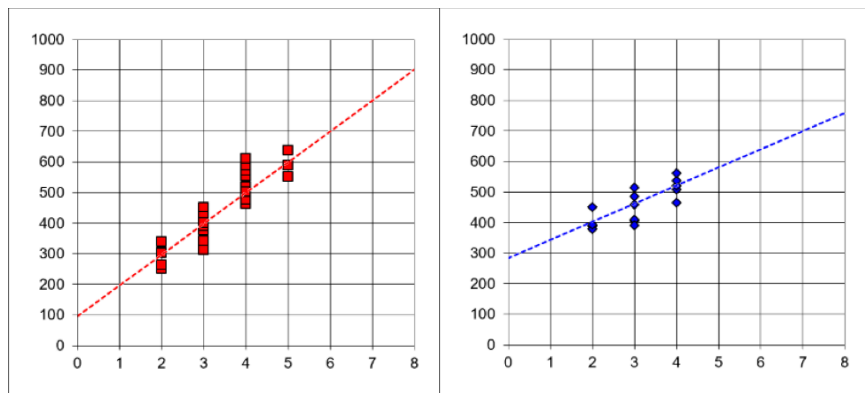


Fig. 5. Línea de crecimiento de  $I(n)$  en muestras de italiano e inglés (en función del número de sílabas,  $2 \div 5$  vs.  $2 \div 4$ , entre las posiciones prominentes).

intervalos interacentuales de las lenguas que suponemos isoacentuales son variables (cf. Romano & Mairano 2010)<sup>22</sup>.

### 3. Nuevos métodos de evaluación

#### 3.1. Métodos de evaluación basados en métricas

Teniendo en cuenta la estrecha relación que se establece entre la organización temporal y las manifestaciones de diversos niveles de prominencia, a partir de 1999, los estudios de Ramus *et al.* mostraron la posibilidad de obtener una clasificación indirecta de los tipos rítmicos en un espacio continuo. Se comenzó entonces a trabajar con

medidas de variación de las duraciones de secuencias de segmentos.

Las alternancias de sílabas fuertes y débiles y/o en los pies métricos definen la posición del acento en el pie métrico, pero en el plano segmental, a estas alternancias pueden corresponder secuencias constituidas por un número diferente de sonidos, definiendo intervalos vocálicos,  $I_v$ , y consonánticos,  $I_c$  (véase figura 6).

La originalidad del método propuesto por Ramus *et al.* (1999) radica en el hecho de que se toman las duraciones de estas secuencias ( $DI_v$  y  $DI_c$ ), en muestras de habla conectada suficien-

<sup>22</sup> En referencia a investigaciones anteriores, estos trabajos aportaron pruebas sobre la reducción de las propiedades de compresión observadas en lenguas como el español y el francés, y el papel discriminante de los parámetros de rigidez relacionados con la duración de las sílabas y los segmentos. Un mayor control de los intervalos entre sílabas

conduce a modelos de compresión regulares que afectan a todas las sílabas del intervalo o a patrones de compresión específicos que reducen (e incluso eliminan) solo las sílabas débiles. Además, estos patrones muestran una resistencia diferente a la variación del tiempo y al aumento de la duración de las expresiones verbales.

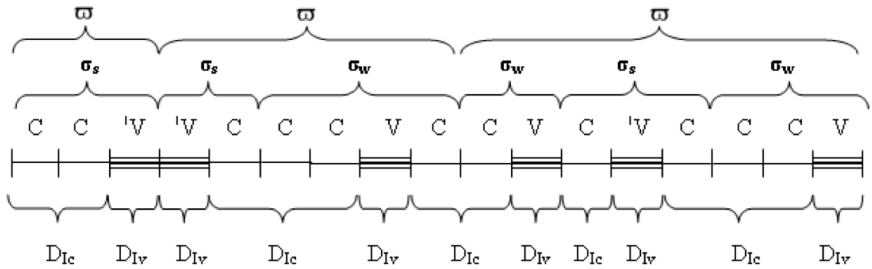


Fig. 6. Ejemplo ficticio de secuencias silábicas con diferente número de consonantes y diferentes posiciones de las sílabas prominentes, con pies métricos de dimensiones (y duración) variables.

temente largas, y se evalúa su variabilidad ( $\Delta$ ) como índice de compresión y dilatación inducida por las propiedades rítmicas: se trata del conocido «método de los deltas».

En un caso ideal, como el de la figura 6, deberíamos tener toda una secuencia de sonidos de igual duración que se organizan en función del hecho de que la vocal nuclear de la sílaba puede ser más fuerte o más débil. De hecho, sabemos que la duración depende más o menos de la prominencia según la lengua, pero, en resumen, este es el modelo ideal en el que definimos con precisión el constituyente de la sílaba, el constituyente del pie y, a continuación, el intervalo consonántico y el intervalo vocálico, que han sido objeto de numerosas mediciones y con evaluaciones estadísticas diferentes<sup>23</sup>.

Mientras Ramus *et al.* (1999) observaban muestras de diferentes idiomas

experimentando con estas medidas (incluida la no secundaria del porcentaje vocálico en la muestra de habla observada, %V), se han impuesto diferentes métodos alternativos para estimar los cambios medios, como el propuesto por Grabe & Low (2002), denominado PVI (*Pair-wise Variability Indexes*, índices de variabilidad por pares), o el de los denominados *VarCos* (*VarCoV* vs. *VarCoC*) de Dellwo & Wagner (2003).

Al mismo tiempo, otros autores como Gibbon & Gut (2001) o Barry & Russo (2003), aunque prestan mucha atención a las características del *timing*, no han renunciado a evaluar otras características rítmicas emergentes, basándose en la observación de diferentes tendencias en modos de organización más sutiles del habla (en términos de (co-)articulación, producción de ruidos como estallidos, fricciones etc., así como en variables como la inten-

<sup>23</sup> Existen varios métodos de medición. Aquí se ven las duraciones de los intervalos, pero podemos dar más importancia a lo que hemos observado, es decir, el hecho de que el *p-center*

se alinee con el ataque, el *onset* vocálico, y tomar como intervalo a evaluar estadísticamente el intervalo de una vocal a la siguiente, lo que se conoce comúnmente como *V-to-V* (cf. §3.4).

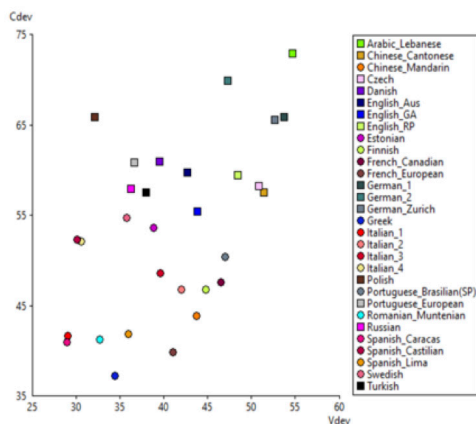


Fig. 7. Ejemplo de dispersión de diferentes muestras de lengua en un plano definido por  $Vdev$  ( $= \Delta(DI_v)$ ) y  $Cdev$  ( $= \Delta(DI_c)$ ).

sidad y la altura melódica; véase la reseña bibliográfica de Roach 2003 y más recientemente la de Payne 2021; cf. §3.6).

### 3.2. Las posibilidades que ofrece Correlatore

En el grupo de trabajo del que formé parte, junto con Paolo Mairano desde principios de los años 2000, empezamos aplicando las mismas técnicas propuestas por estos autores a nuestras muestras de habla, midiendo así el porcentaje vocálico ( $\%V$ ) en términos de variación de duración de los intervalos vocálicos ( $\Delta V = \Delta(DI_v)$ ) y de los intervalos consonánticos ( $\Delta C = \Delta(DI_c)$ ); véase figura 7.

Mediante distintos sistemas de visualización, con el software *Correlatore* (Mairano & Romano 2010) hemos estudiado la posición de los pares de valores *Deltas* en un espacio continuo en el que se agrupan las muestras de lenguas consideradas acentuales y silábicas o, precisamente por

ello, reconocidas como pertenecientes a uno u otro tipo<sup>24</sup>.

A primera observación, las muestras lingüísticas analizadas de esta manera tendían a dispersarse en gran medida según lo esperado (basándose en referencias bibliográficas y en la impresión auditiva que causan en los oyentes)<sup>25</sup>.

<sup>24</sup> La aplicación está disponible en el sitio web del LFSAG en la dirección <https://www.lfsag.unito.it/correlatore/index.html>.

<sup>25</sup> Una primera fuente de variabilidad puede estar en la forma en que se delimitan los segmentos y en las decisiones que se toman (a nivel fonético o fonológico) sobre el estatus de los sonidos como vocales o consonantes. Además, las interrupciones (como la oclusión glotal y otros fenómenos) y las epétesis vocálicas también influyen mucho en la percepción del ritmo y requieren evaluaciones precisas. En todos nuestros trabajos hemos procedido mediante segmentación humana (decidir simplemente qué es una vocal y qué es una consonante no es trivial en diferentes

Sin embargo, el espacio que se definía era continuo y la disposición de los pares de valores que deberían contraponer dos polos ( $\Delta V$  y  $\Delta C$  altos para las lenguas isoacentuales, *LA*, y  $\Delta V$  y  $\Delta C$  bajos para las lenguas isosilábicas, *IS*) mostraba una dispersión en áreas en las que se emparejaban valores diferentes (alto  $\Delta V$  con bajo  $\Delta C$  o bajo  $\Delta V$  con alto  $\Delta C$ , como en el ejemplo de la muestra portuguesa de la figura 7). Además, en algunas muestras se obtuvieron resultados inesperados, a veces contradictorios con lo observado por Ramus *et al.* (1999) (como en el caso del japonés, véase más abajo) y en trabajos que proceden con evaluaciones automáticas<sup>26</sup>.

Con la esperanza de superar estas condiciones de incertidumbre y dar mayor estabilidad a las medidas en muestras de lenguas diferentes, como se anticipó en el §3.1, se han

propuesto otras métricas rítmicas.

Sin embargo, no siempre un modelo propuesto posteriormente consigue mejorar al anterior: en algunos casos, dependiendo de las cualidades de la muestra registrada, el primero parece funcionar mejor. Algunos autores han mostrado entonces cierta desconfianza ante estos resultados: no hay respuesta a por qué en un caso funciona mejor uno y en otro el otro, y la capacidad discriminatoria de estos instrumentos queda limitada<sup>27</sup>.

En nuestro caso, para facilitar el trabajo de comparación entre los diferentes métodos, Paolo Mairano ha desarrollado, en el mismo software *Correlatore*, la posibilidad de comparar las métricas rítmicas, pasando fácilmente de una representación de un tipo a la de otro (así, desviación vocálica, desviación consonántica, *PVT* normalizado, *VarCo* etc.)<sup>28</sup>.

---

idiomas y ha sido relevante en nuestro caso que yo fuera alumno de Michel Contini, que fue alumno de Georges Straka, procediendo, por tanto, de una escuela en la que se evaluaba cuidadosamente la percepción de la frontera entre consonantes y vocales).

<sup>26</sup> Las condiciones reales en las que los hablantes oralizan su texto son muy variables y dificultan el entrenamiento de un sistema automático (al menos tal y como lo intentaron Rouas & Farinas en 2004). A estos fenómenos impredecibles se suman las alteraciones inducidas en las mediciones por la inclusión de segmentos hablados demasiado cortos para obtener valores estables de desviación de la duración de los intervalos y, a menudo, afectados por fenómenos lingüísticos específicos como el alargamiento

---

prepausal (*prepausal lengthening*). Además, el hecho de que haya una aceleración o una ralentización en una fase inicial o en una fase intermedia de la unidad interpausal, *IPU*, conduce a una mayor variación de la duración en la medición de los intervalos (véase notas siguientes).

<sup>27</sup> Por ello, varios investigadores han preferido seguir trabajando evaluando variables más finas, insistiendo en la importancia de la separación de la organización temporal, recuperando el interés por evaluar la calidad de la articulación o la presencia de ruidos, explosiones etc. (véase Arvaniti 2012).

<sup>28</sup> *Correlatore* incluye también la posibilidad de decidir cómo realizar la evaluación estadística basándose en decisiones tomadas por un operador humano (excluyendo o incluyendo las

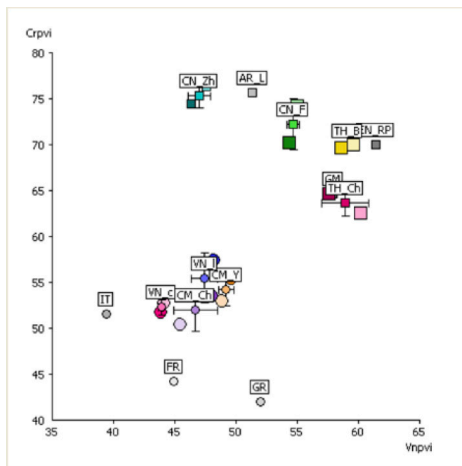


Fig. 8. Ejemplo de dispersión de diferentes muestras de lenguas asiáticas en un plano definido por  $VnPVl$  y  $CrPVl$  [CM = chino moderno estándar, CN = cantonés, TH = tailandés, VN = vietnamita; cf. AR = árabe, EN = inglés, GM = alemán, IT = italiano, FR = francés, GR = griego] (Romano *et al.* 2011).

A partir de la aplicación de estas diferentes medidas y del uso de las distintas funciones de *Correlatore*, pudimos evaluar cómo variaban estas diferentes medidas en la misma muestra y en muestras diferentes por el mismo idioma (como en caso de algunas lenguas asiáticas, véase figura 8; cf. Romano *et al.* 2011).

Aunque no se vea claramente aquí, hemos destacado cómo en gráficos de

---

unidades interpausales, más cortas y con métodos de normalización locales o globales sobre las secuencias de *IPU* o sobre toda la grabación). Sin embargo, aunque ninguna aplicación accesible al público tenga el efecto de mostrar lo que nos interesa, una evaluación automática parece prometedora hoy en día porque las técnicas de inteligencia artificial revelan más fácilmente las cualidades de los datos que se le escapan al mismo fonetista experto (cf. ya el *GVI - Generalized Variability Index* de Tepperman & Nava 2011).

este tipo una lengua como las muestras del tailandés pueden ocupar el mismo lugar que en el trabajo de Ramus *et al.* (1999) se asigna al japonés. En nuestros datos, las muestras de japonés consideradas se sitúan a medio camino entre las lenguas isosilábicas (IS) y las isoaccentuales (IA), con ligeros desplazamientos según la velocidad del habla. Esto depende de la variación en la duración vocálica (sin que intervengan conceptos fonológicos como el de *mora*)<sup>29</sup>.

Si se exceptúan las difíciles decisiones sobre los límites entre los sonidos (y sobre la ausencia de sonidos

---

<sup>29</sup> Por otra parte, en la ubicación en el gráfico parecen influir poco también las distinciones cuantitativas en el sistema vocálico (como en el caso de las muestras estonia y finlandesa; véase la fig. 7).

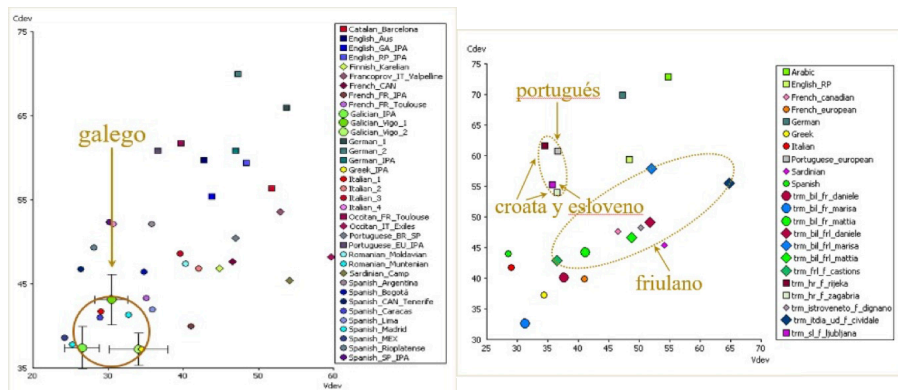


Fig. 9. Ejemplo de dispersión de diferentes muestras de lengua en planos definidos por  $V_{dev}$  ( $= \Delta(DI_V)$ ) y  $C_{dev}$  ( $= \Delta(DI_C)$ ). A la izquierda tres muestras de gallego, en el área más isosilábica (Romano *et al.* 2018); a la derecha muestras de friulano, con diferentes valores de  $V_{dev}$  (y  $C_{dev}$  medios) en comparación con croata y esloveno, en un área específica (Petris & Romano 2020).

esperados o la presencia de sonidos inesperados), se trata únicamente de observaciones de calidad fonética que permiten identificar tendencias relativamente recurrentes en lenguas de diferentes áreas geográficas. Las muestras de las lenguas analizadas ocupan espacios diferentes del plano solo en virtud de las desviaciones en los intervalos vocálicos y consonánticos estimados con las diferentes métricas. No obstante, las aplicaciones que hemos realizado con estos procedimientos y con esta herramienta en diferentes idiomas nos han proporcionado resultados interesantes, precisamente en términos de caracterización lingüística (v. figura 9).

Así, por ejemplo, identificamos entre las muestras más representativas de altos valores de desviación de las métricas las de algunas variedades del alemán y,

entre las de valores más bajos, las muestras del gallego (Romano *et al.* 2018; cf. Boula de Mareüil & Roseano 2025)<sup>30</sup>.

Las lenguas con altos valores de  $\Delta V$  y  $nPVI$ , asociados a valores contenidos de  $\Delta C$  y  $rPVI$ , son, por el contrario, algunas variedades románicas con cantidad vocálica potencialmente distintiva (francés de Canadá, hablas occitanas o francoprovenzales, cf. Romano 2015) o friulanas (Petris & Romano 2020). De estos mismos estudios también han surgido lenguas con

<sup>30</sup> Lamento no haber podido seguir en los últimos años los avances de las líneas de investigación emprendidas por Paolo Roseano y por Philippe Boula de Mareüil, con los que también mantengo estrechas relaciones científicas. Por el contrario, en este caso es una pena que nuestro trabajo anterior sobre el gallego y otros idiomas iberorrománicos no haya sido tenido en cuenta en las investigaciones más recientes (véase la nota siguiente).

valores bajos de  $\Delta V$  y  $nPVI$ , asociados a valores altos de  $\Delta C$  y  $rPVI$ : además del caso del portugués, cabe destacar la extrema regularidad de las vocales que caracteriza (también auditivamente, si prestamos atención) a lenguas como el croato o el esloveno (aunque en estrecho contacto geográfico con el friulano y en proximidad con otras lenguas retorrománicas; cf. Roseano 2020)<sup>31</sup>.

### 3.3. Diferencias rítmicas entre lenguas en contacto

Otras regularidades surgen de la aplicación de estos métodos a las producciones lingüísticas de hablantes bilingües o bidialectales: también en este campo varios estudiosos han estudiado diversos binomios de lenguas (Gamal 2007, White & Mattys 2007, Gut 2012, Romito & Tarasi 2012, Schmid & Dellwo 2013, Mairano *et al.* 2018).

Aunque se han obtenido sin utilizar modelos estadísticos especialmente sofisticados (que, por el contrario, han

sido objeto de investigaciones que en los últimos años han ampliado la lista de estudios en este ámbito), los datos presentados en la figura 9 permiten apreciar también las diferentes dispersiones de valores que hemos podido obtener en el caso de los hablantes friulanos y franceses, y que han mostrado precisamente cómo un mismo hablante es capaz de presentar ritmos diferentes según la lengua que habla (Petris & Romano 2020).

### 3.4. Métodos de evaluación *por niveles*

Como empezaba a precisar anteriormente, al evaluar métricas rítmicas en realidad solo estamos evaluando duraciones de secuencias de sonidos que se sitúan dentro de unidades distintas, en comparación con duraciones dentro de otras unidades. No estamos evaluando lo que ocurre dentro de un intervalo en el que puede haber un mayor control o compensación en la calidad de los segmentos que lo componen.

También por esta razón, se ha cuestionado la validez de estas métricas, ya que se ha sugerido que no describen el ritmo, sino quizá solo un efecto del ritmo, que reverbera en la organización temporal<sup>32</sup>.

Por otra parte, según Marotta (2016):

---

<sup>31</sup> La rapidez con la que en pocos años se ha extendido un interés general por el estudio del ritmo de las lenguas románicas ha llevado a algunas publicaciones a basarse en un estado del arte a menudo muy incompleto y orientado a valorizar solo las contribuciones de algunas escuelas (cf. Payne 2021). Esperamos que las colaboraciones en este ámbito lingüístico puedan reanudarse pronto, al menos entre estudiosos de materias afines y con metodologías compartidas.

---

<sup>32</sup> Son reflexiones de este tipo las que han llevado a Arvaniti (2012) a debatir sobre la «usefulness of metrics in the quantification of speech rhythm».

«Rhythm metrics only reflect the duration and clustering of segments, whereas the complexity of syllable structure and the reduction of unstressed vowels are fundamental to the perception of speech rhythm, in that more reduction and more consonant clusters indicates stress timing, while less reduction and shorter clusters means more syllable timing» (Marotta 2016, p. 491).

Al evaluar la posibilidad de que haya otro nivel de construcción en la definición de estos intervalos, otros autores han considerado entonces que un modelo de construcción del ritmo debería incluir reflexiones sobre un nivel fonotáctico. A un nivel en el que predominan las restricciones de una fonología de la que depende la formación de unidades mayores, el de la organización del enunciado, debe corresponder un nivel de organización más fino, en el que se despliegan las actividades articulatorias locales (cf. Bertinetto & Vékás 1991; Bertinetto & Bertini 2010; Prieto *et al.* 2012).

Según Bertinetto & Bertini (2010), el ritmo perceptivo se derivaría de las condiciones de organización temporal de un *Nivel I* (fonotáctico), basado en el acoplamiento de los osciladores vocálicos y consonánticos (siguiendo las líneas sugeridas por Goldstein *et al.* 2007), que se determinan conjuntamente con las de un *Nivel II* (frasológico), basado en el acoplamiento de los osciladores acentuales y silábicos

(adoptando las sugerencias de O'Dell & Nieminen 1999).

Situándonos en un primer nivel (*Nivel I*), investigamos un ámbito en el que han sido fundamentales las verificaciones de los estudios de Öhman (1966) y Fowler (1977) (que, a su vez, habían desarrollado las primeras observaciones de Lacerda y Menzerath de los años treinta) en términos de coarticulación y «acoustic/articulatory overlapping».

Sobre este tema llevan décadas trabajando investigadores de diferentes grupos, entre los que destaca en Italia el de Zmarich *et al.* (2007), cuyas reflexiones embrionarias se desarrollan ahora en Fontanella *et al.* (2024) mediante representaciones que discutimos sobre la base de esquemas alternativos (como los de la figura 10) referidos al *timing* impuesto por las condiciones de geminación de consonantes intervocálicas<sup>33</sup>.

---

<sup>33</sup> En estas investigaciones, cuyo objetivo es poner de relieve las distinciones articulatorias en el caso de oposiciones de longitud consonántica, se hace referencia a modelos alternativos de sincronización gestual, el «combined vowel and consonant timing» (*V-to-C*; cf. Browman & Goldstein 1989) y el «Vowel-to-Vowel timing» (*V-to-V*; cf. Öhman 1966). Sin embargo, si se demuestra una reorganización de la sílaba que precede a una geminada, mediante una reducción del espacio concedido al núcleo, las hipótesis avanzadas parecen sugerir también una erosión del núcleo vocálico siguiente, concluyendo a favor del establecimiento de condiciones de alargamiento en lugar de constatar una programación diferente del *timing*, estas condiciones derivando de una estructura que ya

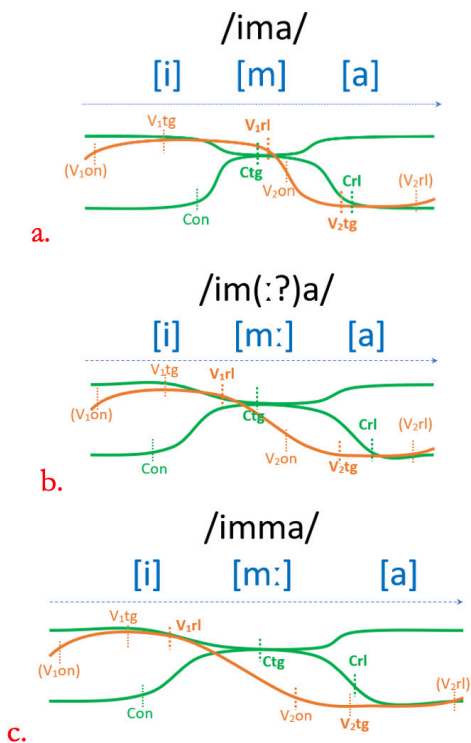


Fig. 10. Ejemplos de coordinación de gestos vocálicos (naranja) y consonánticos (verde) a niveles diferentes de organización fonotáctica.

Arriba, consonante intervocálica simple; abajo, consonante geminada; en el centro, alargamiento consonántico no recategorizado (que no implica reorganización silábica).

En esencia, dado que no se trata de condiciones que intervienen oponiendo consonantes sordas originadas por contracción, *change by shrinking*, sino de estructuras conservadas por rigidez, *by stiffness*, se trata de evaluar las condiciones específicas de determinadas hablas en las que la producción de los distintos (conjuntos) en los estados evolutivos anteriores de la lengua en los que se afirma la «geminación» preveían dos posiciones temporales.

tos de) fonemas está sujeta a formas de mayor control y compensación a diferentes niveles, precisamente<sup>34</sup>.

<sup>34</sup> A un experimento que intenta modelar una dimensión de articulación de mayor duración temporal debería corresponder la hipótesis de un salto categórico en términos fonotáticos: esta dimensión de evaluación, inspirada en lenguas en las que los alargamientos no son débil o dudosamente funcionales, falta en estudios recientes. De hecho, los resultados de algunos de ellos podrían verse invalidados por una visión

### 3.5. Osciladores acoplados

De este nivel se derivarían restricciones de organización temporal dentro de la sílaba y entre sílabas diferentes, según estrategias de control y compensación a varios niveles (quizás en el origen de las diferentes visiones inducidas por los modelos *V-to-C* vs. *V-to-V*) que no tienen en cuenta los estudios articulatorios más recientes citados<sup>35</sup>.

Situándonos en un segundo nivel (*Nivel II*), de hecho, ocurre lo mismo: las realizaciones acústicas/aerodinámicas en la respiración, *breath-grouping* y pausas (Libermann 1967, Grosjean *et al.* 1979), revelan diferentes posibilidades, dado que también en este nivel las unidades se estructuran con diferentes modalidades de transición de una posición prominente a otra (de una sílaba acentuada a otra en agrupaciones de duración variable).

Para intentar modelar las alternativas que permiten a los hablantes de una

---

de los fenómenos de distinción cuantitativa típica de estudiosos que, en su experiencia como hablantes, hacen un uso limitado de la geminación y que, sobre todo, basan sus observaciones en hablantes de zonas en las que esta muestra una funcionalidad limitada.

<sup>35</sup> Sin embargo, una visión de los fenómenos de erosión debidos al solapamiento de gestos o, por el contrario, al retraso de actividades articulatorias cuyo momento de ejecución difiere a la espera de que se complete el gesto anterior, sigue siendo decisiva en el estudio de las modalidades de organización temporal a este nivel.

misma lengua una escanciación en diferentes grupos acentuales, se han avanzado algunas posibilidades. A las observaciones experimentales de modelos como el «base and supplement» (Fujimura 1992) y de modelos superposicionales más probados (Fujisaki 1980-1999) se han añadido modelos que presuponen precisamente dos niveles de oscilación, intrasilábico y extrasilábico (véase el modelo de «coupled oscillators» de O'Dell & Nieminem 1999). En esta visión, las propiedades rítmicas en las unidades interpausales, los grupos de respiración, serían el resultado del acoplamiento de dos osciladores hipotéticos basados en parámetros como *r* (*coupling strength*, fuerza de acoplamiento) y  $\varphi$  (*phasing*, desfaseamiento).

Así pues, interactúan dos niveles de construcción diferentes, más o menos anclados entre sí por coeficientes de acoplamiento, o más o menos desfaseados. La evaluación de estos coeficientes de rigidez o de fase (que aquí todavía no tenemos en cuenta) puede seguir ofreciendo interesantes líneas de desarrollo.

La duración del *Interstress Interval* se describe entonces como la función del número de sílabas y de dos *clocks* cuyas contribuciones están reguladas por este parámetro *r*. De modo que *a*, *b* e *I* de la ecuación anterior (cf. §2.11) se redefinen como en la siguiente fórmula:

$$I(n) = \frac{r}{r\omega_1 + \omega_2} + \frac{1}{r\omega_1 + \omega_2} n$$

en la que  $\omega_1$  es la velocidad de oscilación del oscilador acentual,  $\omega_2$  es la velocidad del oscilador silábico y  $r$  es la fuerza de acoplamiento.

En esencia, con este modelo del oscilador acoplado estamos observando la superposición de eventos en el habla que pueden ser responsables precisamente de un tipo de control que se produce en un determinado nivel y, en cambio, de un control que se manifiesta en otro niveles de organización, el de los grupos respiratorios o de las unidades interpausales, y el de la escansión del enunciado en unidades más pequeñas dominadas por los acentos.

Cuando el valor de la fuerza de acoplamiento ( $r$ ) es 1, entonces la  $a$  de la ecuación anterior (§2.11) es igual a  $b$  y ambos osciladores tienen la misma influencia; pero cuando  $r$  es mayor que 1 ( $r > 1$ ), el oscilador acentual general es dominante, mientras que cuando  $r$  es menor que 1 ( $r < 1$ ), es el oscilador silábico el que es dominante.

Los estudios realizados en los años 80-90 sobre el sueco y el inglés (Fant, Eriksson y otros citados por Barbosa 2006) han evaluado  $r$  en diferentes corpus con varios tempos y han estimado valores en torno a 2 frente a los valores típicos obtenidos para el italiano o el griego ( $r \approx 0,9$ ).

Sin embargo, el modelo se basa en patrones acentuales esperados que también podrían manifestarse con diferentes formas de prominencia: de hecho,

en las aplicaciones probadas para el italiano, los valores de  $r$  observados no siempre se han mostrado estables en las dimensiones de variación esperadas (véase Romano & Mairano 2010).

En un sentido más amplio, además de las observaciones del §3.6, parecen más objetivas las perspectivas de aplicación que podrían surgir gracias a los recientes estudios de Deighton MacIntyre *et al.* (2022) y Aviad & Grice (en prensa) y a los prometedores métodos de los *Recurrence plots* (véase §3.9).

### 3.6. Métodos de evaluación de control y/o compensación

La insatisfacción con los resultados derivados de la aplicación de métricas que no permiten evaluar las distintas contribuciones en estos diferentes planos y la necesidad de definir procedimientos capaces de discernirlas en las evaluaciones de la calidad rítmica en muestras de habla ha estimulado nuevas reflexiones teóricas por parte de P.M. Bertinetto.

Retomando cierta literatura que había sido dejada de lado en el momento en que se impusieron las soluciones basadas en el uso de métricas rítmicas, se llegó al estudio de Bertinetto & Bertini (2010), en el que se consideraban ejemplos de lenguas que satisfacían diferentes combinaciones de estas disposiciones para controlar (*CTRL*) o compensar (*CMPS*) los componentes de los diferentes niveles (véase tabla I).

TYPE	LEVEL-I	LEVEL-II	EXAMPLE
1	CTRL	CTRL	<i>Italian:</i> relatively simple phonotactics, fairly rigid word stress pattern
2	CMPS	CMPS	<i>English:</i> fairly complex phonotactics, fairly mobile word stress pattern, density of secondary stresses yielding further prominence sites
3	CMPS	CTRL	<i>Polish:</i> very complex phonotactics, rigid word stress pattern
4	CTRL	CMPS	<i>Chinese:</i> very simple phonotactics, 'flexible' stress pattern, emerging prominences

Tabla I. Control y compensación a diferentes niveles

Los modelos más estándar de italiano serían un ejemplo de lengua que, como cierto español, conserva un buen control del primer nivel y del segundo («control» significa que intenta mantener segmentos de la misma duración); el inglés británico, por su parte, sería una lengua que compensa tanto en el *Nivel I*, donde vemos segmentos que se acortan y alargan, como en el *Nivel II*, donde vemos cadenas de sílabas que se alargan y acortan a través de mecanismos de reducción y alargamiento para regularizar las posiciones de los acentos (con efectos de reparación al *stress clash* por ej.).

Combinando los dos niveles tendríamos lenguas como el polaco, por ejemplo, o el chino. El polaco compensa en el primer nivel y controla mejor el segundo (el acento siempre está en la misma posición de palabra). El chino, en cambio, es una lengua que compensa en el segundo nivel mientras controla el primero.

En realidad, solo algunas variedades de lenguas responden perfectamente a esta ejemplificación, pero el mérito de Bertinetto y Bertini fue sin embargo el de proponer otro índice y otras métricas, los *CCI*, que intentan dar cuenta de cómo es que a partir de condiciones ideales de duración de los sonidos (véase esquema en figura 6) se pasa a tener métricas diferentes según que el modelo sea el de una lengua que ajusta los intervalos entre acentos (figura 11) frente a una lengua que compensa a nivel de duraciones silábicas o intrasilábicas (figura 12).

Operativamente, al aplicar estas métricas necesitamos entonces dar cuenta de estos aspectos y de que la evaluación, pasando por la fonología, debe hacer frente a fenómenos físicos (articulatorios y acústicos, de superposición y sincronización) ante los cuales el operador (o el instrumento de evaluación automática de los segmentos

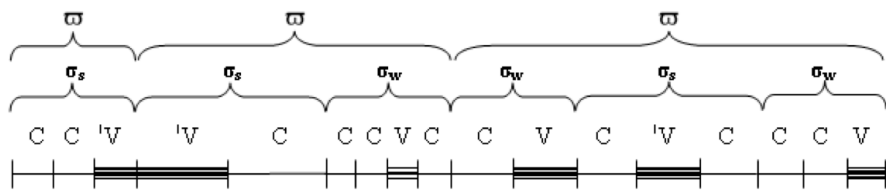


Fig. 11. Ejemplo de patrón ideal *Syllable-based* ( $\sigma_{w1} \approx \sigma_{wp}$ ;  $\sigma_{s1} \approx \sigma_{sp}$ , con adaptación de los constituyentes silábicos y pies métricos de duración variable en función de las agregaciones de sílabas).

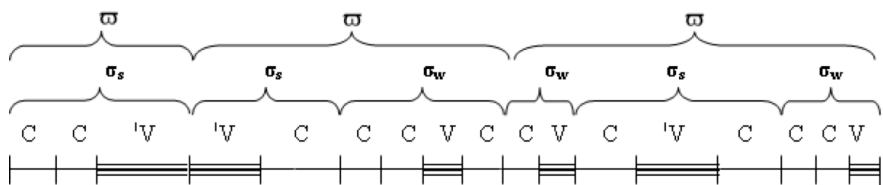


Fig. 12. Ejemplo de patrón ideal *Stress-based* ( $\sigma_1 \approx \sigma_p$ , con constituyentes silábicos variablemente dilatados en sílaba fuerte y comprimidos en sílaba débil).

que deben categorizarse como V o C) debe estar preparado para tomar decisiones clasificatorias determinantes.

Trabajando con dialectos italianos, nos dimos cuenta, por ejemplo, de que la evaluación de los índices de duración del intervalo consonántico o vocálico cambia drásticamente dependiendo de si una vocal dada en una sílaba débil se considera suprimida o no. Una lengua con compensación silábica y control de la duración interacentual tiende a reducir la sílaba y esta reducción puede reflejarse en una supresión (véase figura 13).

Esta supresión provoca una variación considerable de la duración, por lo que en este caso, si considero que aquí se ha suprimido la vocal de la sílaba débil, se crea una larga secuencia de consonantes que a veces tiene enormes conse-

cuencias en la métrica ( $\Delta C \uparrow \uparrow$ ). Si, por el contrario, considero que sigue habiendo una pequeña vocal, aunque reducida a unos pocos ciclos de vibración, la duración de los intervalos consonánticos se estabiliza ( $\Delta C \approx$ ) y, en cambio, lo que más varía es la duración de los intervalos vocálicos ( $\Delta V \uparrow$ ). Esta suposición depende de una observación fina que el operador hace al clasificar, segmentar y medir estos intervalos.

### 3.7. Aplicaciones de los nuevos índices CCI

Bertinetto & Bertini (2010) insertan otro factor de normalización en las métricas que se refiere al número de segmentos que se encuentran dentro del intervalo: con esta normalización, los índices de variación consonántica y vocálica se reorganizan en un gráfico en el

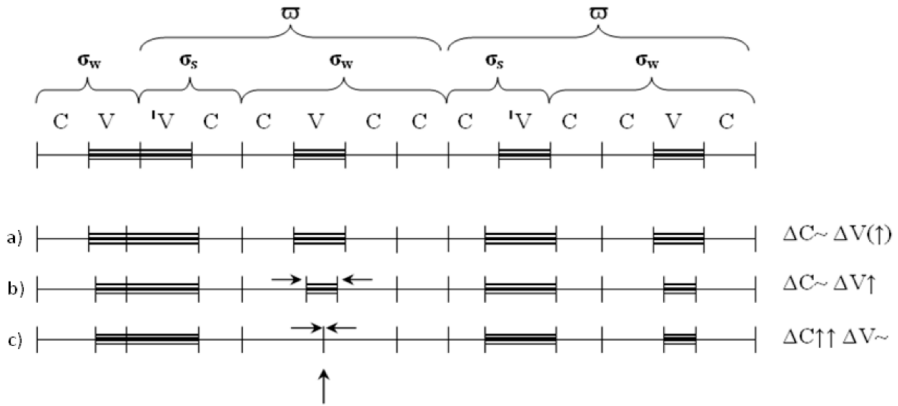


Fig. 13. Efecto diferente en las métricas al considerar una vocal eliminada o reducida (> necesidad de evaluar el número y las condiciones de los segmentos en un intervalo en términos objetivos).

que observan lo que idealmente debería ocurrir (véase ejemplo en figura 14).

Si en las ordenadas pongo el  $CCI(C)$  y en el eje  $x$  el  $CCI(V)$ , la lengua que controla la duración silábica concentra sus valores en torno a la diagonal, la bisectriz del gráfico, mientras que una lengua que compensa para controlar la duración interacentual lo hace comprimiendo y dilatando las sílabas y sus valores caen por tanto en esta zona.

Esto es exactamente lo que se observa trabajando sobre el ritmo de hablas del área apuliana (Romano 2020; cf. Schmid 2004), mostrando algunas muestras isoacentuales, con diferencias respecto a la cancelación de las vocales reducidas. Incluso los dialectos que generalmente se pueden atribuir a un modelo isoacentual pueden diferenciarse en función de las cualidades intrínsecas de las muestras analizadas.

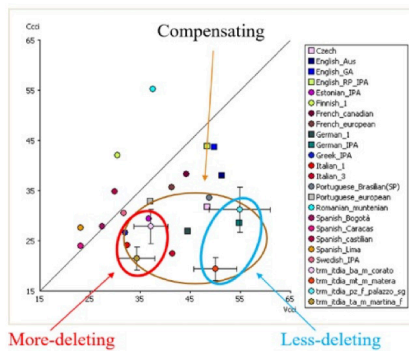


Fig. 14. Diferencias en las modalidades de compensación (con mayor tendencia a la cancelación o a la reducción/alargamiento).

Sin embargo, en la medida en que todos estos dialectos analizados tienden a ser dialectos que compensan, que controlan la duración interacentual, compensando a nivel de reducción silábica, se pueden hacer distinciones. Según la cantidad y la distribución de las vocales que se reducen, se conservan o se suprimen, los de la izquierda son dialectos en los que es más probable que se produzca la supresión y estos del otro lado son aquellos en los que generalmente se produce la conservación<sup>36</sup>.

En resumen, vemos que hay diferencias dentro de una misma zona dialectal, dentro de un mismo grupo de lenguas que tendemos a tratar de la misma manera, por lo que una vez más se confirma la idea de un continuo y de una emergencia del principio de oposición rítmica en términos de percepción subjetiva de otros niveles de organización del habla.

---

<sup>36</sup> No obstante, si atendemos a la opinión de algunos nativos, muchos jurarían que las dos muestras a la derecha son las en que los hablantes más suprimen: los autores locales se apresuran a borrar estas vocales en la escritura o a poner apóstrofes en lugar de las vocales supuestamente suprimidas, incluso en los casos en los que existe claramente una vocal *schwa* reducida residual que el hablante alfabetizado según el modelo del italiano escrito no logra percibir (y aquí vendrían muy bien las reflexiones de Coşeriu sobre los fonemas implícitos).

### 3.8. La regularidad rítmica no reside solo en la duración, sino también en el timing de los eventos

Por último, para tener en cuenta los estudios más recientes, terminar, sin embargo, con una referencia a las observaciones que hemos hecho desde 2009, con Paolo Mairano, que también ha realizado pruebas de percepción, pruebas de clasificación perceptiva de muestras de lenguas muy diferentes, en un espacio mucho más amplio. En el caso que presento aquí más sencillamente, he trabajado sobre lenguas de las que tenía muestras sobre las que habíamos probado un buen funcionamiento de las métricas, lenguas para las que había surgido con certeza una cierta regularidad sobre la organización isoacentual o isosilábica.

En la figura 15 propongo la representación de algunas curvas relativas a una unidad interpausal que hemos analizado al estudiar el ritmo del inglés en el marco de los diferentes trabajos publicados a lo largo de los años. Se trata de un segmento, un enunciado del cuento *El viento norte y el sol*, tomado del punto en el que el hablante dice «and, at last, the northwind gave up the attempt».

Utilicé herramientas que hemos definido en el proyecto *AMPER* para extraer y estilizar  $f_0$  (Romano *et al.* 2014).

A partir de las informaciones sobre

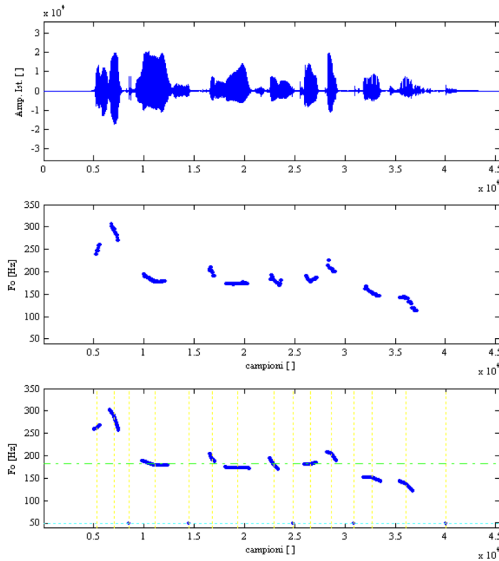


Fig. 15. Arriba, oscilograma y curva de  $f_0$  del pasaje «and, at last, the northwind gave up the attempt». Abajo, curva de  $f_0$  estilizada con puntos azules sobre la línea del eje de x correspondientes a las posiciones en las que se produce un evento ruidoso (explosión de las consonantes oclusivas).

la duración, la intensidad y  $f_0$ , es posible producir una versión sintetizada del fragmento analizado y escuchar las partes sonoras del pasaje (sin información léxica), con y sin los eventos ruidosos correspondientes a los segmentos consonánticos.

Antes de todo, en este caso vimos que si se suprimen las explosiones de las oclusivas, la versión sintetizada no conserva las mismas cualidades rítmicas que la versión original. Y esto hace pensar que incluso el *burst*, donde no hay vocal de acompañamiento, constituye un punto de anclaje en la percepción del ritmo.

Además, gracias al archivo *AMPER* se pueden sintetizar otras versiones,

manipulando algunas variables y probando la eficacia de cada correlato acústico de la prosodia.

Con esta representación de los datos es posible realizar fácilmente diferentes pruebas experimentales, ya que con estos se puede normalizar la duración, regularizar la intensidad o fijar el valor de  $f_0$  definiendo un valor único para cada núcleo vocálico (o segmento consonántico). Véase el ejemplo de la figura 16, donde la duración de los segmentos vocálicos se ha fijado en 100 ms.

El resultado es una percepción que no varía mucho; aunque no se haya realizado un experimento perceptivo estructurado, las pruebas de

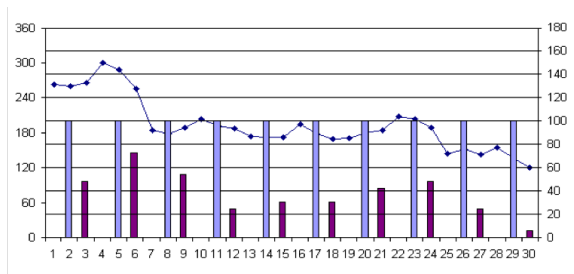


Fig. 16. Valores de  $f_0$  e intensidad variables y duración fijada por el pasaje de figura 15. Al escuchar una versión sintetizada con estos valores, se percibe un ritmo residual distinto al de pasajes similares en lenguas de otro tipo rítmico sometidas al mismo tratamiento (Romano 2010).

percepción informales (propuestas también al público presente en la presentación oral de estos materiales) y una simple prueba de escucha permiten observar con precisión cómo las cualidades rítmicas residuales persisten incluso en el caso en que se normalizan las duraciones (cf. Romano 2010).

Esto significa que en la percepción del ritmo no se pueden descuidar la intensidad y la  $f_0$ .

También los hemos comprobado (Romano 2010) comparando dos muestras: el inglés y el francés, este último claramente isosilábico. Normalizando las duraciones, la diferencia entre las muestras sigue siendo perceptible porque, precisamente si la duración se vuelve irrelevante, son  $f_0$  y la intensidad las que garantizan la percepción del ritmo diferente que caracteriza a los originales.

La conclusión es aquella hacia la que nos conducen estudios más recientes como los de Pellegrino, He & Dellwo (2017), Ravignani (2019) o los de Polianskaya, Busà & Ordin (2020).

Por tanto, la percepción del ritmo del habla no depende solo de variables temporales de características fonéticas de duración de los sonidos y de las secuencias, sino que depende de condiciones de producción de nivel superior que se manifiestan a través del manejo de diferentes variables, ligadas al *timing*, al tiempo y al modo con que se producen las impulsiones vocálicas y los índices de inicio y fin de los sonidos consonánticos<sup>37</sup>.

### 3.9. Recurrence plots

Entonces: ya sea que sean imputables al primer o al segundo nivel en el que

<sup>37</sup> Y esto lo encontramos también en la definición de «eurythmie» de Philippe Martin y, de hecho, en esta literatura complementaria (aunque lo que me desconcierta un poco es que en muchos estudios no haya ninguna referencia a los otros estudios y que cada uno parezca proceder a su manera). Lo que me parece interesante, sin embargo, es que con criterios diferentes, con metodologías definidas dentro de escuelas diferentes, en campos disciplinarios diferentes, todos estamos llegando a conclusiones parecidas.

operamos el control y la compensación, los reflejos de nuestras actividades de coordinación se reflejan, como en el caso de la entonación, en una dimensión multiparamétrica que, aunque no sea inmediatamente visible/deducible a partir de la observación de datos medidos de forma aislada, sin duda deja huellas perceptibles que pueden recuperarse quizá con instrumentos más sofisticados de extracción de datos.

El tema se ha planteado en varias ocasiones como un reto interdisciplinario en reuniones con estudiosos de la comunicación animal, como Andrea Ravignani o Marco Gamba, en el marco de congresos o jornadas de trabajo conjuntas.

De manera totalmente experimental, hemos aplicado una herramienta sugerida por estos estudiosos a una selección de datos ya explorados en profundidad con otras metodologías (véanse los §§ anteriores).

Con todas las limitaciones de un experimento aún incompleto, parece prometedor aprovechar la técnica de los *Recurrence plots*, gráficos de recurrencia (Marwan *et al.* 2007), aplicándola a algunos conjuntos de valores extraídos de muestras de lenguas ya bien analizadas en este sentido.

Un nuevo análisis de las métricas y los valores de  $f_0$  e intensidad para las muestras de gallego, francés, inglés y árabe (en los polos opuestos y a lo largo de un continuo de valores de *Deltas*

y *CCIs*; véanse las figuras 9 y 14) ha dado lugar a los gráficos que se muestran en las figuras 17-19.

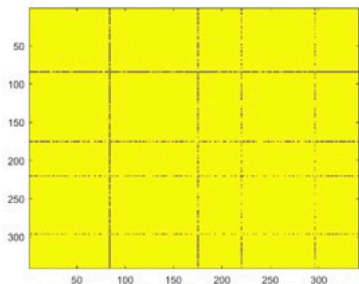
A una selección de estos datos se le ha aplicado también la técnica *RQA* (*Recurrence Quantification Analysis*), útil para traducir en un conjunto de valores comparables las características de variabilidad de los *plots*.

Remitiéndonos a Marwan *et al.* (2007) para una explicación detallada del funcionamiento de la técnica y del valor asumido por los trece parámetros de la *RQA*, nos limitaremos aquí a precisar que las diferencias entre los gráficos que se pueden obtener con esta técnica aplicada a los mismos datos se deben a un valor umbral ( $\epsilon$ ) que contribuye a variar la densidad de las líneas en el *plot*.

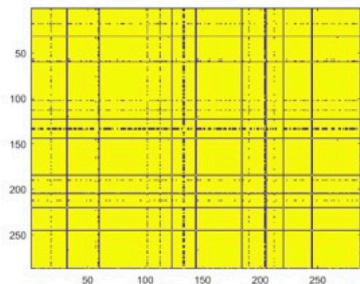
Sin embargo, fue necesario realizar una gran cantidad de trabajo preliminar para obtener estos resultados: al tiempo de procesamiento requerido por estos tratamientos, aquí propuestos utilizando las rutinas *MATLAB* distribuidas por los mismos autores (Marwan *et al.* 2007), se debe añadir el tiempo relacionado con el etiquetado manual, la extracción de datos y su pretratamiento con *PRAAT* y *Correlatore*<sup>38</sup>.

---

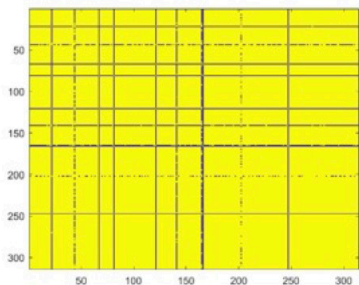
<sup>38</sup> La validez de las pocas consideraciones propuestas anteriormente sobre la base de este tipo de resultados no justifica la complejidad del proceso de tratamientos que se limitan, en definitiva, a una verificación de las evaluaciones obtenidas a través de las métricas rítmicas ya discutidas.



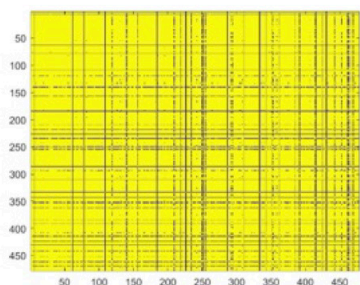
**Gallego con  $\epsilon = 200$**



**Inglés con  $\epsilon = 200$**

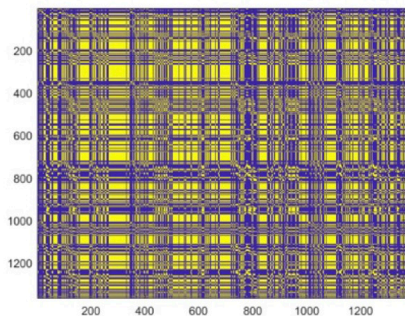


**Francés con  $\epsilon = 200$**

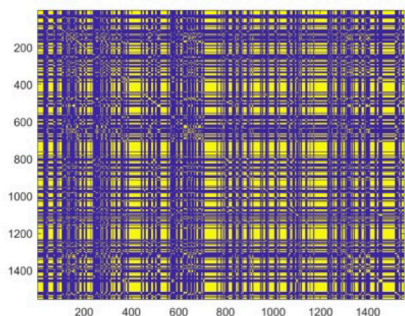


**Árabe con  $\epsilon = 200$**

Fig. 17. *Recurrence Plots* para muestras de ritmo diferente: procesamiento aplicado a las métricas extraídas por *Correlatore* en toda la secuencia de *El viento norte y el sol* de cuatro idiomas.



**Gallego  $\epsilon = 10$**

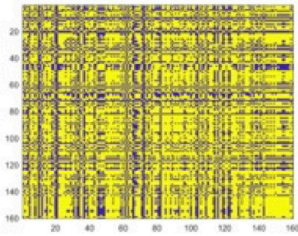


**Árabe  $\epsilon = 10$**

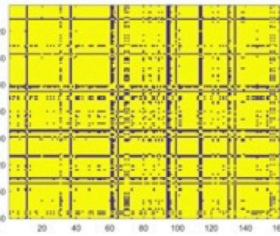
Fig. 18. *Recurrence Plots* para muestras de ritmo diferente: procesamiento aplicado a los valores de  $f_0$  en los primeros 5 s de las grabaciones de *El viento norte y el sol* en dos idiomas.

**Gallego con  $\epsilon = 50$**

$\Delta C$  LMAX 22 TT 4.3510 Vmax 37.0000 RTmax 36.0000

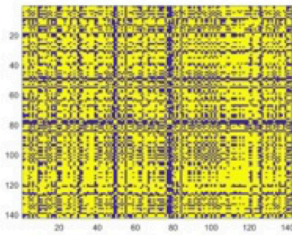


$\Delta V$  LMAX 34 TT 9.5778 Vmax 34.0000 RTmax 28.0000

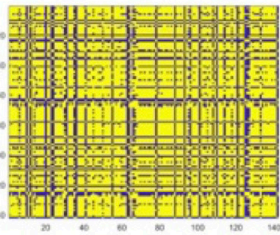


**Francés con  $\epsilon = 50$**

$\Delta C$  LMAX 17 TT 4.5497 Vmax 41.0000 RTmax 36.0000

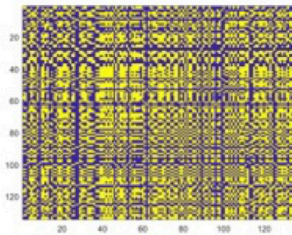


$\Delta V$  LMAX 18 TT 5.0332 Vmax 18.0000 RTmax 29.0000

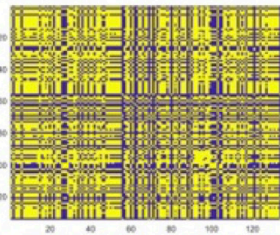


**Inglés con  $\epsilon = 50$**

$\Delta C$  LMAX 15 TT 2.8710 Vmax 10.0000 RTmax 35.0000

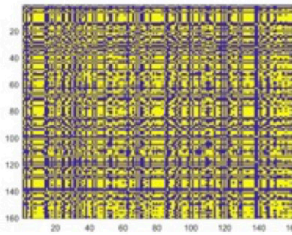


$\Delta V$  LMAX 13 TT 3.8505 Vmax 27.0000 RTmax 40.0000



**Árabe con  $\epsilon = 50$**

$\Delta C$  LMAX 12 TT 3.4278 Vmax 12.0000 RTmax 35.0000



$\Delta V$  LMAX 13 TT 3.8438 Vmax 17.0000 RTmax 51.0000

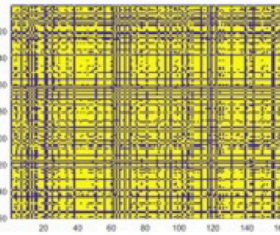


Fig. 19. *Recurrence Plots* para muestras de cuatro idiomas de ritmo diferente: procesamiento aplicado por separado a las métricas  $\Delta C$  y  $\Delta V$  extraídas por *Correlatore* en toda la grabación de las diferentes versiones de *El viento norte y el sol*.

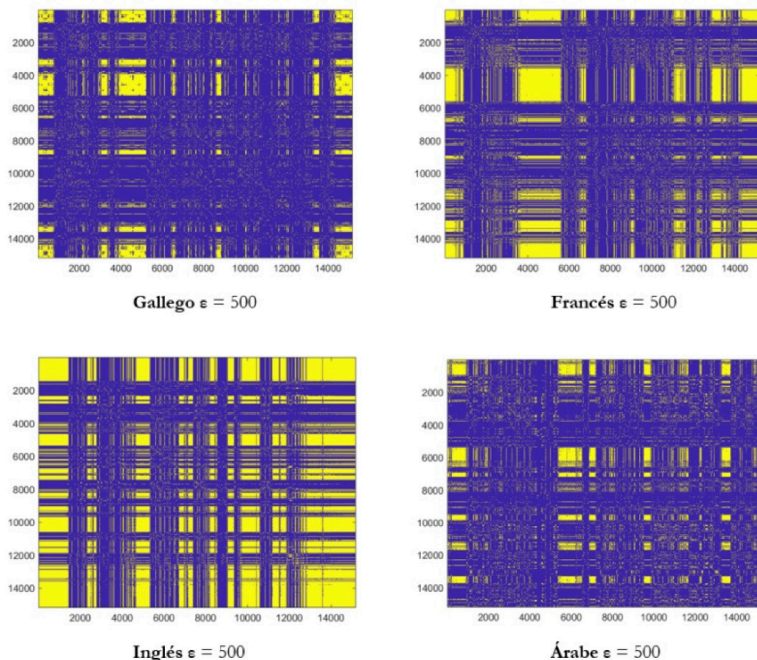


Fig. 20. *Recurrence Plots* para muestras de cuatro idiomas de ritmo diferente: procesamiento aplicado a los datos brutos (con baja frecuencia de muestreo) de las grabaciones de 4 versiones de *El viento norte y el sol*

Por el contrario, se obtiene una clara mejora con la aplicación directa de esta técnica a los datos brutos del habla (véase la figura 20). Los *Recurrence plots* obtenidos de esta manera (sobre los mismos pasajes objeto de manipulaciones y evaluaciones perceptivas de Romano 2010) presentan una densidad decreciente al pasar del gallego al francés y al inglés, en consonancia con las consideraciones perceptivas sobre su clasificación rítmica<sup>39</sup>.

<sup>39</sup> En este caso, la única operación preliminar fue submuestrear los fragmentos que se

La novedad es que la misma muestra de árabe que, evaluada a través de las métricas, parecía situarse cerca del polo isoacentual, con esta técnica muestra valores que lo relacionan con modelos de recurrencia similares a los de muestras isosilábicas (y esto es compatible con las características de primer nivel de esta lengua, con una fonotaxis relativamente simple).

---

iban a clasificar para limitar los tiempos de procesamiento, dada la considerable cantidad de datos analizados.

Es como si las métricas, al situar el gallego y el árabe en polos opuestos, reflejaran mejor la organización del segundo nivel.

Y es quizás esta, en conclusión, la dimensión que aún habría que evaluar bien, con esta prometedora técnica u otras que aún esperamos optimizar, probándolas en amplias muestras de lenguas.

### Referencias bibliográficas

Abercrombie D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Allen G.D. (1972). The Location of Rhythmic Stress Beats in English: An Experimental Study I. *Language and Speech*, 15(1), 72-100.

Allen G.D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics*, 3, 75-86.

Arvaniti A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66, 46-63.

Arvaniti A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3), 351-373.

Arvaniti A. (2021). Measuring Speech Rhythm, In R.A. Knight & J. Setter (eds.), *The Cambridge Handbook of Phonetics*, Cambridge University Press, 312-335.

Auer P., Couper-Kuhlen E. & Müller F. (1999). *Language in Time: The Rhythm and Tempo of Spoken Interaction*. Oxford: O.U.P.

Aviad A. & Grice M. (in press). Rhythm is a timescale. In L. Meyer & A. Strauss (eds.), *Rhythms of Speech and Language:*

*Culture, Cognition, and the Brain*. Cambridge: Cambridge University Press [https://www.researchgate.net/publication/379359800\_Rhythm\_is\_a\_timescale (accessed Jan 26 2025)].

Barbosa P.A. (2006). *Incurões em torno do ritmo da fala*. Campinas: Pontes.

Bertinetto P.M. (1981). *Strutture prosodiche dell'italiano*. Firenze: Accademia della Crusca.

Bertinetto P.M. & Bertini C. (2010). Towards a unified predictive model of Natural Language Rhythm. In M. Russo (ed.), 43-77.

Bertinetto P.M. & Vékás D. (1991). Controllo vs. compensazione sui due tipi di isocronia. In E. Magno Caldognetto & P. Benincà (a cura di), *L'interfaccia tra fonologia e fonetica*, Padova: Unipress, 155-162.

Browman C.P. & Goldstein L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 201-251.

Byrd D. & Krivokapić J. (2021). Cracking Prosody in Articulatory Phonology. *Annual Review of Linguistics*, 7, 31-53.

Boula de Mareüil Ph. & Roseano P. (2025). A speaking atlas of the languages of the iberian peninsula: focus on rhythm and varieties in contact. *Dialectologia*, 35, 27-54 (doi:10.1344/DIALECTOLOGIA.35.2).

Cloarec-Heiss F. (1999). From Natural Language to Drum Language An Economical Encoding Procedure in Banda-Linda (Central African Republic). In C. Fuchs & S. Robert (eds.), *Language Diversity and Cognitive Representations* [Human Cognitive Processing 3], 145-158 [https://doi.org/10.1075/hcp.3.13clo (accessed Jan 26 2025)].

Colonna V. (2022). *Voices of Italian Poets. Storia e analisi fonetica della lettura della poesia italiana del Novecento*, Alessandria: Dell'Orso.

Dauer R.M. (1983). Stress-timing and syllable-timing reanalyzed, *Journal of Phonetics*, 11, 51-62.

Dellwo V. & Wagner P. (2003). Relations between language rhythm and speech rate. *Proc. of the 15th International Congress of Phonetics Sciences* (Barcelona, Spain), 471-474.

Deighton MacIntyre A., Qing Cai C., Scott S.K. (2022). Pushing the envelope: Evaluating speech rhythm with different envelope extraction techniques. *J. Acoust. Soc. Am.*, 151, 2002-2026.

Demuth K. (1996). The prosodic structure of early words. In J. Morgan & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, Mahwah (New Jersey), Lawrence Erlbaum Associates, 171-184.

Fant G. & Kruckenberg A. (1996). On the quantal nature of speech timing. *Proceedings of ICSLP 1996*, 2044-2047.

Fontanella F. Balsemin T., Gili Fivela B. et al. (2024). L'organizzazione temporale delle geminate dell'italiano: uno studio di modellizzazione tramite la Fonologia Articolatoria. In *Studi AISV*, 12, 259-281 (DOI: 10.17469/O2112AISV000015).

Fougeron C. & Pillot-Loiseau C. (2015). Characterizing rhythmic alterations in the speech of French dysarthric patients. *Com. pres. a ISMBS 2015* (<http://ismbs.eu/data/documents/linked-abstracts-.pdf>).

Gamal D. (2007). Sul ritmo in italiano L2: l'eventuale isocronia. In: V. Giordani, V. Bruseghini & P. Cosi (a cura di), *Scienze Vocali e del Linguaggio - Metodologie di Valutazione e Risorse Linguistiche (Atti del III*

*Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce*, ITC-IRST Povo - Trento, 29 Nov. - 1 Dic. 2006), Torriana (RN): EDK, 101-118.

Gibbon D. & Gut U. (2001). Measuring Speech Rhythm. *Proceedings of Eurospeech* (Aalborg, Denmark), 91-94.

Ginepro (2024). L'impiego dei fonetogrammi nell'analisi di voci cantate in diverse lingue nella tradizione musicale eurocolta. *Bollettino del LFSAG*, 14, 8-56 [[www.lfsag.unito.it/ricerca/phonews/14/14\\_2.pdf](http://www.lfsag.unito.it/ricerca/phonews/14/14_2.pdf) (accessed Jan 26 2025)].

Grabe E. & Low E.L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (eds.), *Papers in Laboratory Phonology 7*, Berlin: Mouton de Gruyter, 515-546.

Grammont M. (1913). *Le vers français, ses moyens d'expression, son harmonie*. Paris: Champion.

Grammont M. (1946). *Traité de Phonétique*. Paris: Delagrave (1<sup>e</sup> éd. 1933).

Gut U. (2012). Rhythm in L2 speech. *Speech and Language Technology*, 14/15, 83-94.

Hernandez A., Yeo E., Kim S. & Chung M. (2020). Dysarthria Detection and Severity Assessment using Rhythm-Based Metrics. *Proc. of Interspeech* (Shanghai, 25-29 Oct. 2020), 2897-2901.

Jakobson R. (1941-42). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Universitets Årsskrift (trad. it. *Il farsi e il disfarsi del linguaggio: linguaggio infantile e afasia*, Torino: Einaudi, 1971).

Janker P.M. (1995). Evidence for the P-Center Syllable-Nucleus-Onset Correspondence Hypothesis. *ZAS Papers in Linguistics*, 7, 94-124.

- Kruckenberg A. & Fant G. (1993). Iambic versus trochaic patterns in poetry reading. *Nordic Prosody*, VI, 123-135.
- Keller E. (2007). Waves, Beats and Expectancy. *Proceedings of the 16th International Congress of Phonetic Sciences* (Saarbrücken, Germany), 355-360.
- Kubozono H. & Mizoguchi A. (2023). Text-To-Tune Alignment in Japanese “Happy Birthday To You”. In R. Skarnitzl & J. Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (Praga, 2023), 580-584.
- Lehiste I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5, 253-263.
- Leroi-Gourhan A. (1964-65). *Le Geste et la Parole: I. Technique et Langage, II. La Mémoire et les rythmes*. Parigi: A. Michel (trad. it. *Il gesto e la parola. Tecnica e linguaggio. La memoria e i ritmi*, Torino: Einaudi, 1977).
- Liberman M. & Prince A. (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry*, 8(2), 249-336.
- Liss J.M., White L., Mattys S.L. et al. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *J Speech Lang Hear Res.*, 52(5), 1334-1352.
- Loukina A. Kochanski G., Rosner B. et al. (2011). Rhythm measures and dimensions of durational variation in speech. *J. Acoust. Soc. Am.*, 129(5), 3258-3270.
- MacNeilage P.F. & Davis B.L. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (ed.), *Attention and Performance XIII - Motor Representation and Control*, Hillsdale (N.J.): LEA, 452-468.
- MacNeilage P.F. (1998). The Frame/Content theory of evolution of speech production, *Behavioral and Brain Sciences*, 21(4), 499-511.
- Maffia M., Schettino L., De Micco R. & Tessitore A. (2023). Disfluencies in Parkinson’s Disease. A study on Italian early-stage patients. In B. Gili Fivela et al. (eds.), *Il parlato in ambito medico: analisi linguistica, applicazioni tecnologiche e strumenti clinici*. Milano: Officinaventuno, 117-130.
- Mairano P. & Romano A. (2010). Un confronto tra diverse metriche ritmiche usando Correlatore. In S. Schmid, M. Schwarzenbach & D. Studer (a cura di), *La dimensione temporale del parlato - AISV2009* (Zurigo, Switzerland), Torriana (RN): EDK, 79-100.
- Mairano P. & Romano A. (2012). Testing the perception of speech rhythm on natural and artificial stimuli. *Proceedings of the 6th International Conference Speech Prosody 2012* (Shanghai, Cina), vol. II, Shanghai: Tongji University Press, 482-485.
- Mairano P., Mois M., De Iacovo V., Romano A. (2018). Acquisizione di fenomeni temporali e ritmici dell’italiano: Analisi di apprendenti anglofoni di italiano L2. *Ri-Cognizioni*, 5(10), 121-136.
- Marwan R., Romano M.C., Thiel M. & Kuths J. (2007). Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5-6), 237-329.
- Marotta G. (1985). *Modelli e misure ritmiche: la durata vocalica in italiano*. Bologna: Zanichelli.
- Marotta G. (2016). Prosodic structure. In A. Ledgeway & M. Maiden (eds.), *The Oxford Guide to the Romance Languages*, Oxford: Oxford University Press, 484-494.
- Marrero V. (2014). Linguistic Rhythm: Perception And Production In Normal And Disordered Speech. In A. Romano et al. (eds.) (2015), *Aspetti prosodici e testuali del raccontare: dalla letteratura orale al parlato dei*

media, Alessandria: Dell'Orso, abstract 4 pp. (58 pp. in CD-ROM).

Nespor M. & Vogel I. (1986). *Prosodic phonology*. Dordrecht: Foris.

O'Dell M. & Nieminen T. (1999). Coupled oscillator model of speech rhythm. *Proceedings of the 14th International Congress of Phonetic Sciences* (San Francisco, USA). 1075-1078.

Pamies Bertrán A. (1999). Prosodic Typology: On the Dichotomy between Stress-Timed and Syllable-Timed Languages. *Language Design*, 2, 103-130.

Pamies Bertrán A. (2010). Quelques malentendus à propos du concept de rythme en linguistique. In M. Russo (ed.), 227-263.

Patel A.D. (2008). *Music, language, and the brain*. Oxford: Oxford University Press.

Payne E. (2021). Comparing and deconstructing speech rhythm across Romance languages. In C. Gabriel, R. Gess & T. Meisenburg (eds.), *Manual of Romance Phonetics and Phonology*, Berlin: De Gruyter, 264-298.

Pellegrino, E., He, L., Dellwo, V. (2017). Computation of L2 speech rhythm based on duration and fundamental frequency. In: J. Trouvain, I. Steiner, B. Mobius (eds). *Elektronische Sprachsignalverarbeitung 2017*. Dresden: TUD press, 246-253.

Petris M. & Romano A. (2020). Correlati metrico-ritmici in campioni di parlato di friulani residenti in Belgio. *Bollettino del LFSAG*, 6, 27-44 [www.lfsag.unito.it/ricerca/phonews/06/6\_3.pdf (accessed Jan 26 2025)].

Pike Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Polyanskaya L., Busà M.G. & Ordín M. (2020). Capturing Cross-linguistic Differences in Macro-rhythm: The Case of Italian and English. *Language and Speech*, 63(2), 242-263 (doi: 10.1177/0023830919835849).

Prieto P., Vanrell M., Astruc L. et al. (2012). Phonotactic and phrasal properties of speech rhythm: Evidence from Catalan, English, and Spanish. *Speech Communication*, 54(6), 681-702.

Proto T. & Dell, F. (2013). The structure of metrical patterns in tunes and literary verse. Evidence from discrepancies between musical and linguistic rhythm in Italian songs. *PROBUS*, 25(1), 105-138.

Ramus F., Nespor M. & Mehler J. (1999). Correlates of Linguistic Rhythm in the Speech Signal. *Cognition*, 73(3), 265-292.

Ravignani A., Dalla Bella S., Falk S. et al. (2019). Rhythm in speech and animal vocalizations: a cross-species perspective. *Annals of the New York Academy of Sciences*, 1453(1), 79-98 [doi: 10.1111/nyas.14166 (accessed Jan 26 2025)].

Roach P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (ed.), *Linguistic controversies*, London: Edward Arnold, 73-79.

Roach P. (ed.) (2003). *A Bibliography of Timing and Rhythm in Speech* (University of Reading, Last updated 2nd April 2003), [http://www.personal.rdg.ac.uk/~llsroach/timing.pdf (accessed Jan 26 2025)].

Rodríguez Vázquez R. (2010). *The Rhythm of Speech, Verse and Vocal Music: A New Theory*. Bern: Peter Lang.

Romano A. (2010). Speech Rhythm and Timing: Structural Properties and Acoustic Correlates. In S. Schmid, M. Schwar-

zenbach & D. Studer (a cura di), *La dimensione temporale del parlato*, Torriana (RN): EDK, 45-75.

Romano A. (2020). Vowel reduction and deletion in Apulian and Lucanian dialects with reference to speech rhythm. In C. Anderson & N. Kuznetsova (a cura di), *Vowel reduction*, No. monografico di *Italian Journal of Linguistics*, 32.1 (2020), 85-102 (DOI: 10.26346/1120-2726-149).

Romano A. (2024). Grafici di densità di probabilità per voci di cantanti. *Bollettino del LFSAG*, 14, 55-66 [www.lfsag.unito.it/ricerca/phonews/14/14\_3.pdf (accessed Jan 26 2025)].

Romano A. & Mairano P. (2010). Speech rhythm measuring and modelling: pointing out multi-layer and multi-parameter assessments. In M. Russo (ed.), 79-116.

Romano A., Mairano P. & Calabrò L. (2011). Measures of Speech Rhythm in East-Asian Tonal Languages. *Proc. of the 17th International Congress of Phonetic Sciences* (Hong Kong, 17-21 Agosto 2011), 1714-1717.

Romano A., Contini M. & Lai J.-P. (2014). L'Atlas Multimédia Prosodique de l'Espace Roman: uno strumento per lo studio della variazione geoprosodica. In: F. Tosques (a cura di), *20 Jahre digitale Sprachgeographie* (Atti del Convegno Internazionale di Berlino, 2-3 nov. 2012), Berlin: Humboldt-Universität - Institut für Romanistik, 27-51.

Romano A., Mairano P. & De Iacovo V. (2018). Classificazione ritmica di alcuni campioni di galiziano parlato rispetto a quelli di altre parlate romanze. In: X.A. Álvarez Pérez & M. Brea (a cura di), *Obreiro da lingua, amigo da xente: Estudos de*

*xeografía lingüística en homenaxe a Manuel González González*, Universidade de Santiago de Compostela, 189-202.

Romito L. & Tarasi A. (2012). A Rhythmic-Prosodic Analysis of Italian L1 and L2. In M. Pettorino & A. De Meo (eds.), *Prosodic and Rhythmic Aspects of L2 Acquisition: The Case of Italian*, Cambridge: Cambridge -Scholars Publishing, 137-152.

Roseano P. (2020). Il ritmo linguistico del ladino dolomitico: studio acustico del badiotto. *Ladinia*, 44, 141-162.

Russo M. (ed.) (2010). *Prosodic Universals: comparative studies in rhythmic modeling and rhythm typology*, Roma: Aracne.

Russo M. & Barry W.J. (2008). Measuring rhythm. A quantified analysis of Southern Italian Dialects Stress Time Parameters. In A. Pamies, M.C. Amorós & J.M. Pazos (eds.), *Experimental Prosody - Proc. of the 4th Congreso Int. de Fonética Experimental* (Granada, Spain), *Language Design*, special issue 2, 315-322.

Schmid S. (2004). Une approche phonétique de l'isochronie dans quelques dialectes italo-romans. In T. Meisenburg & M.T. Selig (eds.), *Nouveaux départs en phonologie*, Tübingen: Narr, 109-124.

Schmid S. & Dellwo V. (2013). Sprachrhythmus bei bilingualen Sprechern. *Travaux Neuchâtelois de Linguistique*, 59, 109-126 [doi.org/10.26034/tranel.2013.2948 (accessed Jan 26 2025)].

Scott S.K. (1993). Perceptual centres in speech: an acoustic analysis. *PhD Thesis*, University College London.

Tepperman J. & Nava E. (2011). Long-distance rhythmic dependencies and their application to automatic language identification. *Proceedings of Interspeech 2011*, 1061-1064.

Wenk B. & Wioland F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10, 193-216.

White L. & Mattys S.L. (2007). *Rhythmic typology and variation in first and second languages*. In P. Prieto, J. Mascaró & M.J. Solé (eds.), *Segmental and prosodic issues in romance phonology*, Amsterdam: Benjamins, 237-257.

Yeo E., Kim S. & Chung M. (2021). Automatic Severity Classification of Korean Dysarthric Speech Using Phoneme-Level

Pronunciation Features. *Proc. of Interspeech* (Brno, 30 Aug.-3 Sept. 2021), 4838-4842.

Zmarich C., Gili Fivela B., Perrier P. *et al.* (2007). Consonanti scempie e geminate in Italiano: studio acustico e cinematografico dell'articolazione linguale e bilabiale. In V. Giordani, V. Bruseghini & P. Cosi (eds.), *Scienze Vocali e del linguaggio – Metodologie di valutazione e risorse linguistiche*, Torriana (RN): EDK, 151-163.