

# Le plurilinguisme et les technologies Internet



LexiPraxi 2000

Paris, 8 décembre 2000

François YERGEAU

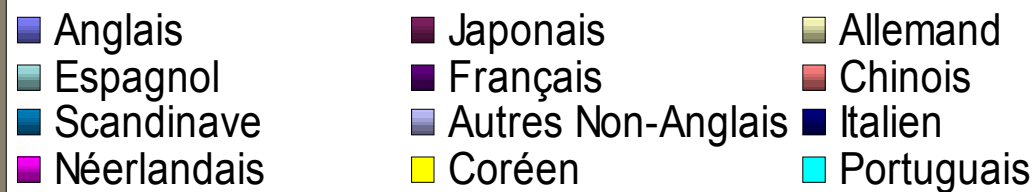
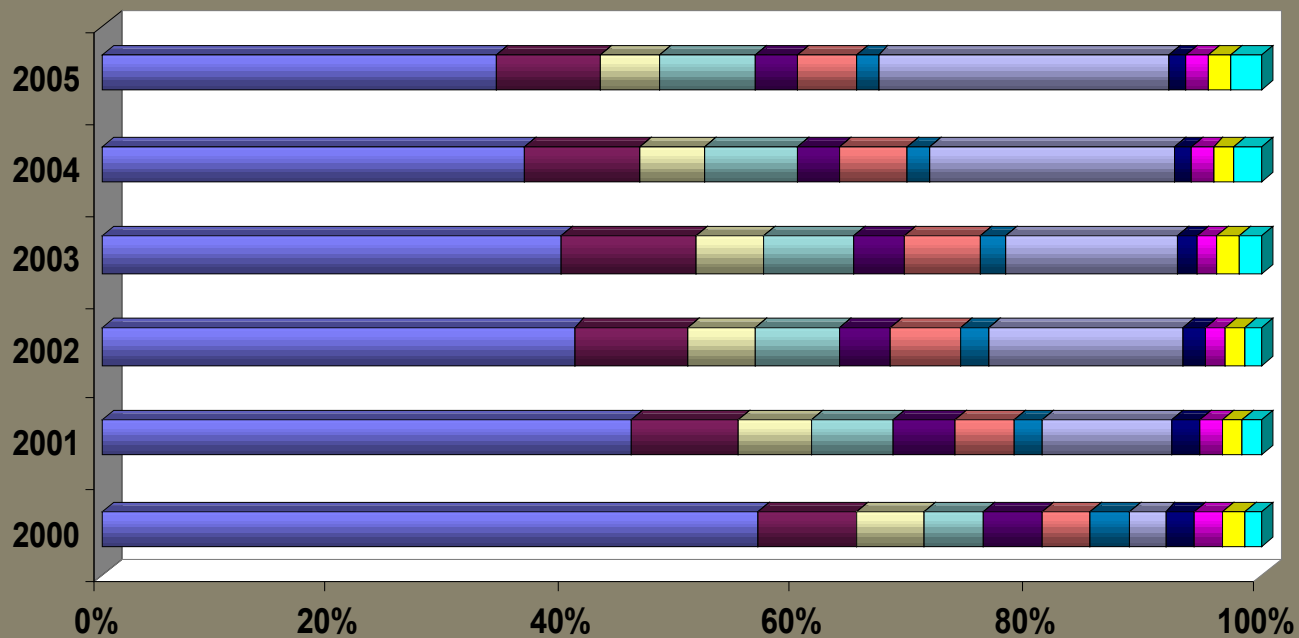
# Aperçu

- ◆ Introduction
- ◆ Les protocoles de base
  - ◆ Les noms de domaine
- ◆ Les messageries
- ◆ Le Web et les URL
- ◆ Conclusion

# Introduction

- ◆ Internet créé aux Etats-Unis
- ◆ Manque flagrant d'attention aux langues
- ◆ Domination de l'anglais (~70 %) renforcée par les difficultés technique
- ◆ Depuis 1994, explosion démographique
- ◆ Influence sur la technique

# Usagers Internet en ligne 2000-2005



# Les protocoles de base

- ◆ IP, TCP, UDP, ICMP, SNMP, etc.
- ◆ Numérique
- ◆ Transparence
- ◆ Adresses IP par ex. *199.84.165.125*
- ◆ Rien à dire !

# Les noms de domaine

- ◆ Notion de texte apparaît :

`ampère.edf.fr` ↔ `199.84.165.125`

- ◆ Mais non ! è n'est pas ASCII !
- ◆ Quid du chinois ? Désolé...
- ◆ L'Asie se mobilise
- ◆ Création de multiples systèmes incompatibles, fragmentation

# Noms de domaine (suite)

- ◆ À l'IETF, réaction tardive : IDN
  - ◆ Co-présidents Québécois, Chinois
- ◆ Extension du DNS fondée sur Unicode
- ◆ Base solide pour un DNS mondial
- ◆ Pilote : *Verisign Global Registry* en CJC
- ◆ Jalousie, soupçons d'impérialisme, etc.
- ◆ À suivre...

# Messageries

- ◆ Courriel : RFC 822, SMTP
- ◆ Inforums (Usenet) : NNTP
- ◆ Mess. Instantanées : protocoles privés
- ◆ Courriel internationalisé par MIME, par extension Usenet aussi



# Quelques zones d'ombre

- ◆ *Petites* langues : absence ou immaturité de jeux de caractères adéquats
  - ◆ Effet aggravé par Internet
- ◆ Réalisations MIME déficientes : =E9, &8129;
- ◆ Fragmentation linguistique
  - ◆ At 00/11/17 12:27 -0500, Fran 苺 is Yergeau wrote:

# Zones d'ombre (suite)

- ◆ Les adresses !
  - ◆ À droite du @ : nom de domaine, IDN
  - ◆ À gauche du @ : rien
- ◆ Unicode est la solution à long terme
- ◆ Prend du temps
- ◆ Résistance en Asie : impérialisme

# Messageries instantanées

- ◆ Caractéristique dominante : immaturité
- ◆ Fonctionne par hasard, qui assez souvent fait bien les choses
- ◆ Effort à l'IETF
  - ◆ Principes d'i18n (notamment Unicode)
  - ◆ Querelles de clocher

# Le web

- ◆ Bon bulletin : HTTP non contraignant, il n'y a pas de HTML déjà ancienne :
  - ◆ RFC 2070 (IETF), HTML 4 (W3C)
- ◆ Implantations modernes (IE5, NS6) très impressionnantes (Unicode)
- ◆ XML et C<sup>ie</sup> fondés sur Unicode

# Point noir : les URL

- ◆ URL = nom de domaine + chemin
- ◆ RFC 2306 : ASCII
- ◆ Pas clair comment les NDI seront appliqués
- ◆ Unicode est la seule solution praticable
- ◆ Inertie : « Il nous faut un vrai jeu de caractères universel : ASCII »

# Point noir : formulaires

- ◆ Mécanisme construit sur les URL : ASCII
- ◆ Méthodes de contournement : résultat délicat, mal documenté et peu robuste
- ◆ Seul l'anglais a un laisser-passer valide
- ◆ URL *unicodisés* ? Pas tout de suite
- ◆ Nouveau protocole XForms, XML, donc Unicode

# Conclusion

- ◆ Protocoles de base : OK
- ◆ Messageries, Web : pas mal
- ◆ Adresses & formulaires : pauvre
- ◆ Évolution  $\pm$  rapide, inertie grandissante
- ◆ Unicode de plus en plus répandu

# La première « autoroute » numérique !



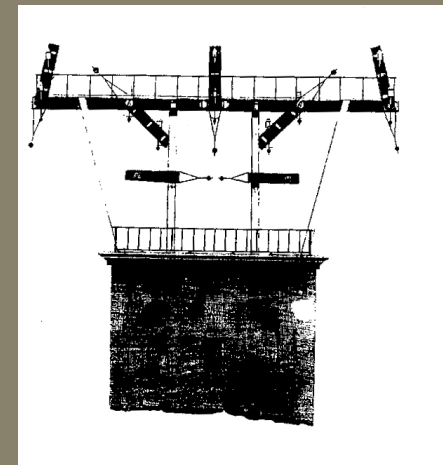
## Claude Chappe

- Né a Brûlon en 1763
- Créateur du *premier* réseau télégraphique optique international

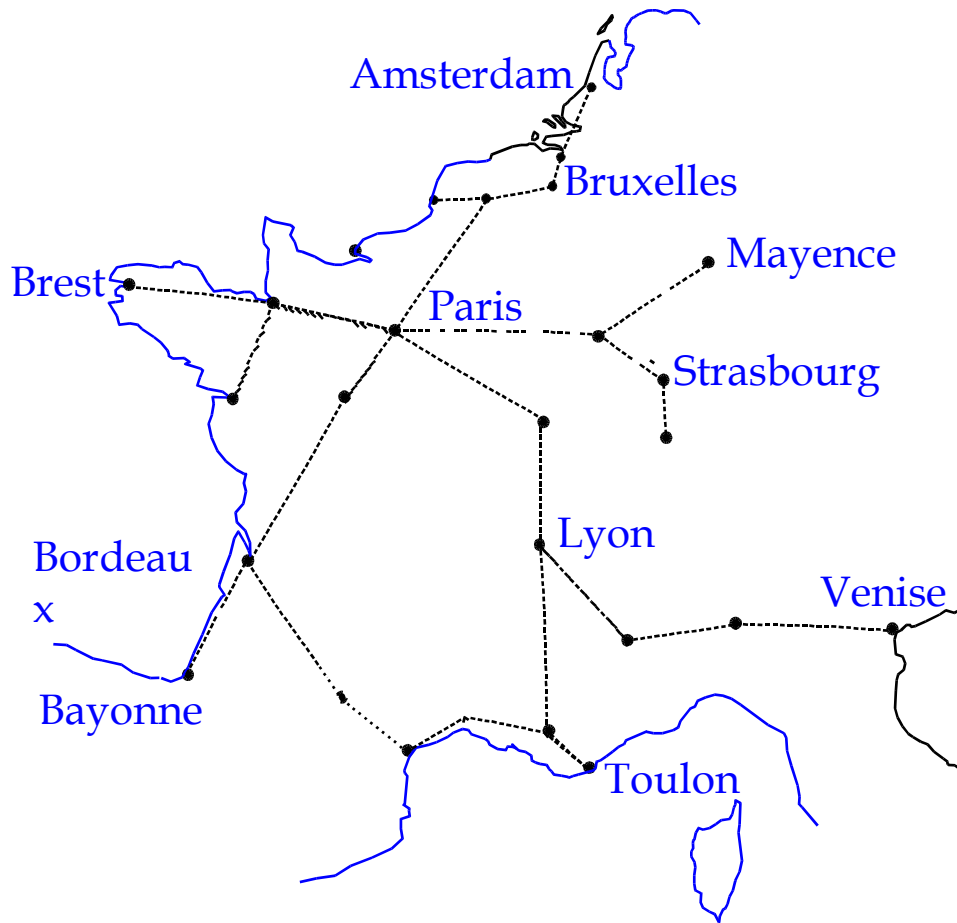
Message transmis le 2 mars 1791 sur 16 km:

« L'Assemblée Nationale récompensera les expériences utiles au public ».

Durée de transmission: 6 minutes 20 secondes







## En 1852...

- 556 stations
- 4800 km
- 27 grandes villes reliées